



HAL
open science

Évaluation du contenu d'une image couleur par mesure basée pixel et classification par la théorie des fonctions de croyance

Nadjib Guettari

► **To cite this version:**

Nadjib Guettari. Évaluation du contenu d'une image couleur par mesure basée pixel et classification par la théorie des fonctions de croyance. Sciences de l'ingénieur [physics]. Université de poitiers, 2017. Français. NNT: . tel-02427016

HAL Id: tel-02427016

<https://hal.science/tel-02427016>

Submitted on 3 Jan 2020

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

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

THÈSE
POUR L'OBTENTION DU GRADE DE
DOCTEUR DE L'UNIVERSITÉ DE POITIERS
FACULTÉ DES SCIENCES FONDAMENTALES ET APPLIQUÉES
DIPLOME NATIONAL - ARRÊTÉ DU 25 MAI 2016

Ecole Doctorale : Sciences et Ingénierie pour l'Information, Mathématiques - S2IM
Secteur de Recherche : Traitement du Signal et des Images

Présentée par :

Nadjib GUETTARI

**Évaluation du contenu d'une image couleur par mesure basée pixel et
classification par la théorie des fonctions de croyance**

Directeurs de Thèse :

Philippe CARRÉ

Anne-Sophie CAPELLE-LAIZÉ

Soutenue le 10 Juillet 2017

Devant la Commission d'Examen

JURY

Christophe CHARRIER, Maître de Conférences HDR, Université de Caen Normandie .. Rapporteur
William PUECH, Professeur des universités, Université de Montpellier Rapporteur
Olivier COLOT, Professeur des universités, Université Lille 1 Examineur
Philippe CARRÉ, Professeur des universités, Université de Poitiers Directeur de Thèse
Anne-Sophie CAPELLE-LAIZÉ, Maître de Conférences, Université de Poitiers Co-directeur

À Nassima, Anes & Aymen
À mes parents

Remerciements

Cette thèse est l'un des produits finis de la recherche que j'ai menée au laboratoire XLIM de l'université de Poitiers. L'ensemble des membres de XLIM m'ont réservé un accueil chaleureux et ont maintenu un environnement de travail des plus agréables, propice au bon déroulement et à l'aboutissement de ce travail. Ce travail a été financé par la région Poitou-Charentes dont je tiens à remercier toutes les personnes derrière le financement de cette recherche.

En rédigeant cette dernière page du manuscrit, je suis obligé de reconnaître que ce travail de thèse n'aurait pas été aussi fructueux sans l'aide de plusieurs personnes. A travers ces quelques paragraphes, je tiens à remercier tous ceux qui ont contribué, de près ou de loin, à sa réalisation.

Je tiens tout d'abord à remercier William Puech, Professeur des universités à l'Université de Montpellier et Christophe Charrier, Maître de Conférences HDR à l'Université de Caen Normandie, d'avoir accepté de rapporter ce manuscrit. Leur rapports témoignent de l'effort et du temps qu'ils y ont consacré, leurs commentaires m'ont été très utiles pour la préparation de la soutenance, je suis très reconnaissant.

Je remercie également Olivier Colot, Professeur des universités à l'Université Lille 1 pour avoir accepté d'examiner cette thèse et assuré le rôle de président du jury.

Un très grand merci à mon directeur de thèse Philippe Carré, Professeur des universités à l'institut XLIM de l'université de Poitiers, d'avoir accepté de diriger ma thèse. Je le remercie pour sa rigueur scientifique, ses nombreux conseils et ses réunions de travail.

Je tiens à remercier particulièrement Annes-Sophie Capelle-Laizé, Maître de conférences à l'Université de Poitiers, qui m'a permis de travailler sur un sujet aussi intéressant et qui m'a guidé pendant ces années de thèse, ses conseils avisés et nos discussions m'ont été extrêmement bénéfiques. Je voudrais la remercier également pour sa patience, la confiance qui m'a toujours octroyé, le soutien et la disponibilité dont elle a fait preuve.

Je remercie toutes les personnes avec qui j'ai eu la chance de partager mon bureau, Mohamed, Maxime et Ton pour les heures passées ensemble. Un grand merci aussi à mes compagnons de thèse, et en particulier à Abdeslam, Badereddine, Naty et Riadh pour nos nombreuses discussions passionnantes et tous les bons moments passés ensemble. Je remercie aussi tous mes camarades thésards que j'ai eu la chance de connaître : Abou, Jack, Taha, Ahmed . . .

Je tiens à adresser un remerciement particulier à ma femme Nassima la personne, qui a su me supporter pendant ces 3 années de thèse, et qui m'a vraiment soutenu pour achever ce travail. Je tiens également à m'excuser auprès d'elle car je sais que je n'étais pas toujours facile à vivre surtout durant les derniers mois de thèse.

Finalement, un très grand merci à ma famille pour son soutien, ses encouragements et d'avoir toujours été là pour moi, dans les bons comme dans les mauvais moments. Je ne remercierais jamais assez mes parents ; si j'en suis là aujourd'hui c'est grâce à eux.

Table des matières

Table des matières	iii
Table des tableaux	v
Table des figures	vii
Abréviations et notations	viii
Introduction générale	1
1. Contexte et objectifs	1
2. Principales contributions	3
3. Organisation du manuscrit	4
1 Apprentissage automatique et fonctions de croyance	7
1.1 Introduction	7
1.2 Qu'est ce que l'apprentissage automatique?	9
1.2.1 Les types d'apprentissage	10
1.2.2 Classification et régression : deux problèmes d'apprentissage su- pervisé	11
1.2.3 Construction d'un modèle pour un problème d'apprentissage	12
1.2.4 Apprentissage automatique et reconnaissance de forme	15
1.3 Théorie des fonctions de croyance	19
1.3.1 Fonctions de masse	20
1.3.2 Révision des croyances	23
1.3.3 Obtention d'une fonction de croyance	31
1.3.4 Niveau pignistique : Prise de décision	33

1.4	Conclusion	34
2	Évaluation de la qualité d'image basée sur la théorie des fonctions de croyance	37
2.1	Introduction	38
2.2	Évaluation de la qualité d'image	39
2.2.1	Évaluation subjective de la qualité d'image	40
2.2.2	Évaluation objective de la qualité d'image	44
2.2.3	Apprentissage automatique pour l'évaluation de la qualité d'image	48
2.3	Mesure sans référence d'évaluation de la qualité d'image basée sur la théorie des fonctions de croyance	51
2.3.1	Représentation floue pour la mesure de qualité	51
2.3.2	Algorithme <i>EVreg</i> pour l'évaluation sans référence de la qualité d'image	52
2.3.3	Définition des descripteurs	55
2.3.4	Un algorithme de régression évidentiel multi-sources	56
2.4	Évaluation des algorithmes <i>EVreg</i> et <i>ms-EVreg</i>	58
2.4.1	Protocole expérimental	58
2.4.2	Résultats de l' <i>EVreg</i> avec des caractéristiques prises séparément	59
2.4.3	Évaluation de <i>ms-EVreg</i>	62
2.5	Les coefficients d'affaiblissement optimisés pour <i>ms-EVreg</i>	64
2.5.1	Le principe	64
2.5.2	Estimation des coefficients d'affaiblissement	65
2.5.3	Évaluation de <i>wms-EVreg</i>	66
2.5.4	Comparaison avec d'autres mesures d'évaluation	69
2.5.5	La robustesse et l'indépendance de la base	74
2.5.6	Évaluation de <i>wms-EVreg2</i> sur toute la base TID2013	75
2.6	Conclusion	77
3	Détection d'une information cachée dans une image	79
3.1	Introduction	80
3.2	Le contexte de la stéganographie	81
3.2.1	La stéganographie au cours des siècles	82
3.2.2	La stéganographie de nos jours	83
3.3	Stéganographie dans les images	84
3.3.1	Propriétés d'un algorithme de stéganographie	86

3.3.2	Stéganographie dans le domaine spatial	88
3.3.3	Stéganographie dans un domaine transformé	91
3.4	La stéganalyse, détection de stéganographie	94
3.4.1	Les différentes classes de stéganalyse	95
3.4.2	Différents schémas de la stéganalyse	96
3.5	Un schéma de stéganalyse basé sur la théorie des fonctions de croyance	102
3.5.1	Positionnement général	102
3.5.2	Espace de caractéristiques pour la stéganalyse	104
3.5.3	Ev-Knn pour la stéganalyse	105
3.5.4	Discussion sur les espaces de caractéristiques	110
3.5.5	<i>EVEC-Knn pour la stéganalyse</i>	111
3.6	Évaluation de la méthode proposée	113
3.6.1	Paramétrage de la méthode : \mathcal{D}_{red} et R	113
3.6.2	Analyse sur nsF5	115
3.6.3	Introduction d'une classe de rejet	117
3.6.4	Analyse sur <i>J-UNIWARD</i>	118
3.6.5	Capacité de généralisation	120
3.6.6	Évaluation dans le domaine spatial sur HUGO et <i>S-UNIWARD</i>	121
3.7	Conclusion	123
	Conclusion & perspectives	125
	Annexe A	129
	Bibliographie	129
	Bibliographie de l'auteur	141

Liste des tableaux

1.1	Illustration de la règle de combinaison moyenne	27
1.2	Exemple de combinaison de Dempster appliquée sur les moyennes des masses	28
1.3	Illustration d'une matrice de confusion.	29
1.4	Exemple d'une matrice de confusion.	30
2.1	Les différentes caractéristiques utilisées	57
2.2	Les résultats des coefficients de corrélation de Spearman pour chaque groupe de caractéristiques	60
2.3	Les résultats des coefficients de corrélation de Pearson pour chaque groupe de caractéristiques	61
2.4	La médiane des coefficients de corrélation de Spearman sur 1000 ensembles apprentissage-test sur la base LIVE	63
2.5	La médiane des coefficients de corrélation de Pearson sur 1000 ensembles apprentissage-test sur la base LIVE	63
2.6	Première proposition des coefficients d'affaiblissement	66
2.7	Exemple du calcul du coefficient de fiabilité du vecteur x^ϕ	67
2.8	Deuxième proposition des coefficients de fiabilité issus de la matrice de confusion	67
2.9	La médiane des coefficients de corrélation de Spearman sur 1000 répétitions ensembles apprentissage-test sur la base LIVE	68
2.10	La médiane des coefficients de corrélation de Pearson sur 1000 répétitions ensembles apprentissage-test sur la base LIVE	69
2.11	La médiane des coefficients de corrélation de Spearman sur 1000 ensembles apprentissage-test sur la base LIVE	70
2.12	La médiane des coefficients de corrélation de Pearson sur 1000 ensembles apprentissage-test sur la base LIVE	70
2.13	Échelle d'interprétation des résultats	71

2.14	Les coefficients de corrélation de Spearman sur une partie de la base d'image TID2013 et apprentissage sur LIVE	75
2.15	Les coefficients de corrélation de Spearman sur la base d'image TID2013 et apprentissage sur LIVE	76
2.16	Les coefficients de corrélation de Spearman sur une partie de la base d'image CSIQ	77
3.1	Résultats des probabilités d'erreur P_E sur J -UNIWARD avec les caractéristiques $CCPEV$ et CF	120
3.2	Résultats des probabilités d'erreur P_E sur nsF5 et J -UNIWARD ensemble	120
3.3	Variation de P_E en fonction de \mathcal{D}_{red} sur les algorithmes HUGO et S -UNIWARD	122
3.4	Résultats des probabilités d'erreur P_E $EVEC$ - Knn sur HUGO et S -UNIWARD	123

Table des figures

1.1	Illustration graphique de la fonction de crédibilité <i>bel</i>	22
1.2	Illustration graphique de la fonction de plausibilité <i>pl</i>	23
2.1	Échantillon d'images de la base IVC.	41
2.2	Échantillon d'images de la base LIVE.	42
2.3	Échantillon d'images de la base Toyama.	43
2.4	Échantillon d'images de la base CSIQ.	43
2.5	Échantillon d'images de la base TID2013.	44
2.6	Différentes approches des mesures objectives	45
2.7	L'algorithme <i>EVreg</i> appliqué pour une mesure d'évaluation de la qualité d'image	56
2.8	Exemples de résultats obtenus sur une image originale ("churchandapitol") et ses cinq versions dégradées de la base d'image LIVE.	72
2.9	Exemples de résultats obtenus sur une image originale ("rapids") et ses cinq versions dégradées de la base d'image LIVE.	73
3.1	La page de garde de Steganographia de Johannes Trithemius, l'inventeur du mot "stéganographie"	81
3.2	Exemple de dissimulation d'un message secret par l'algorithme de substitution LSB	82
3.3	Scénario de Stéganographie / Stéganalyse présenté par le problème des prisonniers	85
3.4	Illustration des différents plans de bit d'une image en niveau de gris	90
3.5	Le processus classique de la stéganalyse	94
3.6	Schéma de l'algorithme Ev-Knn	106
3.7	Le choix du paramètre <i>k</i>	108
3.8	Influence de la mesure de distance sur la probabilité d'erreur	109
3.9	Schéma de principe l'algorithme EVEC-Knn	111

3.10	Influence de la dimension du sous-espace des caractéristiques et du nombre de classifieurs Ev-Knn sur le taux d'erreur de classification. Le vecteur de caractéristiques utilisé est CCPEV de dimension $\mathcal{D} = 548$	114
3.11	Comparaison des résultats obtenus par <i>EVEC-Knn</i> et Ensemble Classifier avec différents taux d'insertion (payload), $R = 40$, $\mathcal{D}_{red} = 350$	116
3.12	Variation de la probabilité d'erreur par rapport au seuil/taux de rejet	117
3.13	Variation de l'erreur de classification en fonction de la dimension du sous-espace de caractéristiques \mathcal{D}_{red}	119

Abréviations et notations

Apprentissage automatique

\mathcal{X} :	Matrice des données d'entrée.
N :	Nombre d'échantillons dans la matrice des données d'entrée
M :	Dimension du vecteur de caractéristiques extrait de chaque échantillon
\mathcal{Y} :	Matrice des données de sortie.
\mathcal{M} :	Modèle prédit de l'apprentissage automatique.
Θ :	Ensemble des hyper-paramètres de l'apprentissage.
ϵ_r :	Erreur de régression.
ϵ_c :	Erreur de classification.
f_R :	Fonction de risque.
\hat{y} :	Estimation du modèle.
EQM :	Erreur quadratique moyenne.
\mathcal{X}_{app} :	Ensemble de données d'apprentissage.
ϵ_{app} :	Erreur d'apprentissage.
\mathcal{X}_{val} :	Ensemble de données de validation du modèle.
$P(\omega_i)$:	Probabilités <i>a priori</i> associées aux classes ω_i .
$f(x \omega_i)$:	Densités des probabilités conditionnelles des vecteurs x aux classes ω_i .
x :	Vecteur de caractéristiques.
σ :	Fonction d'activation du réseau de neurones.
b :	Biais du neurone.

Théorie des fonctions de croyance

Ω :	Cadre de discernement.
m :	Fonction de masse.
$F(m)$:	Ensemble des éléments focaux de m .
bel :	Fonction de crédibilité.
pl :	Fonction de plausibilité.

$BetP$:	Probabilité pignistique.
\oplus :	Règle de combinaison de Dempster.
\odot :	Règle de combinaison conjonctive.
\cup :	Règle de combinaison disjonctive.
α :	Coefficient d'affaiblissement.
T_f :	Taux de fiabilité.
S_j :	Source d'information j .
$R_j p(x_j \omega_k)$:	Vraisemblance de ω_k et de toutes les décisions qui lui sont proches.
R_j :	Coefficient de pondération des probabilités lié à la source S_j .
α_{jk} :	Coefficient d'affaiblissement de la source S_j pour la classe ω_k .

Évaluation de la qualité d'image

MOS :	Score moyen de qualité attribué à chaque image (Mean Opinion Score).
DMOS :	Score de différence entre l'image originale et sa version dégradée.
EQM :	Erreur Quadratique Moyenne.
SNR :	Rapport signal sur bruit (Signal Noise Ratio).
PSNR :	Pic du signal sur bruit (Peak Signal Noise Ratio).
SVH :	Système visuel humain.
UQI :	Indice universel de qualité (Universal Quality Index).
SSIM :	Indice de similarité structurelle.
MS-SSIM :	Indice de similarité structurelle multi-échelle.
VIF :	Mesure de fidélité visuelle (Visual Information Fidelity).
BLIINDS-II :	Mesure sans référence de qualité d'image dans le domaine de la transformée en cosinus discrète (DCT)
NIQE :	Mesure de qualité des images naturelles (Natural Image Quality Evaluator).
EV-KNN :	K-plus proches voisins évidentiel (Evidential K-nearest neighbors).
EV_{reg} :	Algorithme de régression basé sur EV-KNN.
μ :	Fonction d'appartenance.
\hat{m} :	Fonction de croyance floue.
α^f :	Coefficient d'affaiblissement d'un vecteur de caractéristiques f .

ms-EVreg : Version multi-sources de *EVreg*.

wms-EVreg : Version pondérée de *ms-EVreg*.

Détection d'une information cachée dans une image

cover : Objet de couverture.

stégo : Objet stéganographié.

\mathcal{C} : Ensemble des objets *cover*.

\mathcal{M} : Ensemble des tous les messages qui peuvent être communiqués.

\mathcal{K} : Ensemble des clés possibles.

Emb : Fonction de dissimulation de messages (Embedding).

Ext : Fonction d'extraction de message (Extraction).

I_c : Image *cover*.

I_s : Image stégo.

LSB : Bits de poids faibles (Least Significant Bit).

DCT : Transformée en cosinus discrète.

\mathcal{D} : Dimension du vecteur de caractéristiques.

\mathcal{D}_{red} : Dimension réduite de vecteurs de caractéristiques.

\mathcal{F} : Sous-espace de caractéristiques.

FLD : Discrimination linéaire de Fisher.

P_E : Probabilité d'erreur.

P_{FA} : Probabilité de fausses alarmes.

P_{FD} : Probabilité de fausses détections.

\mathcal{X}^{app} : Ensemble d'apprentissage.

Introduction générale

Contexte et objectifs

De nos jours, il est devenu de plus en plus simple pour qui que ce soit de prendre des photos avec des appareils photo numériques, de télécharger des images sur l'ordinateur et d'utiliser différents logiciels de traitement d'images pour appliquer des modifications sur ces images (compression, débruitage, transmission, etc.). Cependant, ces traitements entraînent des dégradations qui influent sur la qualité visuelle de l'image. À cet effet, il est nécessaire de disposer d'outils efficaces capables de mesurer l'impact de ces distorsions sur la qualité de l'image. De plus, ces dernières années est apparue une perturbation particulière, à savoir l'insertion de messages "invisibles" à des fins légitimes ou malveillantes pour des communications confidentielles ou secrètes. De nos jours, avec la généralisation de l'internet, la stéganographie devient une pratique populaire et facilement accessible à toute personne désirant dissimuler un message ou communiquer de façon secrète. Par conséquent, la nécessité et le besoin de détecter des objets stéganographiés donnent lieu à la stéganalyse, le processus dual de la stéganographie.

Dans le cadre de cette thèse, nous nous sommes donc intéressés à la détection et la caractérisation des perturbations d'une image numérique. L'objectif dans un premier temps est de développer une mesure de qualité sans référence permettant d'évaluer de manière automatique la qualité d'une image en corrélation avec l'appréciation visuelle humaine. Nous avons ensuite abordé le problème de la stéganalyse permettant de détecter, avec la meilleure fiabilité possible, la présence d'informations cachées dans des images numériques.

Plusieurs travaux menés par le passé ont montré l'efficacité des méthodes d'apprentissage automatique dans des applications liées à l'évaluation objective de la qua-

lité d'image et la stéganalyse. L'apprentissage joue un rôle important dans de nombreux domaines tels que les statistiques, l'intelligence artificielle et la reconnaissance de formes. En particulier, l'apprentissage supervisé qui consiste à prédire une réponse (variable de sortie) y , en fonction d'un vecteur $x = (x_1, \dots, x_p)$ où x_i est une variable d'entrée appelée aussi prédicteur à partir d'un ensemble d'apprentissage au préalable labélisé. Il est nécessaire de trouver une relation fonctionnelle entre les variables d'entrée et les variables de sortie. Ce problème est également appelé classification lorsque la sortie y est qualitative ou bien régression quand il s'agit d'une mesure quantitative.

De nombreuses techniques ont été proposées dans la littérature pour estimer cette relation fonctionnelle, nous pouvons citer par exemple les méthodes à noyaux, les plus proches voisins, les réseaux de neurones sous toutes ses variantes etc.

L'objectif commun de ces techniques est de reconnaître et classer les formes en satisfaisant un critère, en général celui d'approcher le partage subjectif de ces formes réalisé par un être humain. Pour cela, un ensemble de caractéristiques est extrait d'une forme inconnue, cet ensemble est exploité pour produire des hypothèses sur sa classe d'appartenance. D'un point de vue historique, les méthodes utilisées pour estimer les relations de dépendance fonctionnelle entre les variables sont généralement déterminées à partir de modèles probabilistes. Bien que la théorie des probabilités repose sur des bases mathématiques solides, elle ne permet pas de prendre en compte facilement les incertitudes et les imprécisions des données. En effet, ces caractéristiques peuvent être imprécises ou incertaines et issues de sources hétérogènes avec différents degrés de précision. L'objectif est donc est de fusionner de manière plus souple que le formalisme probabiliste, les différentes sources en prenant en compte les imperfections dont elles sont entachées. En outre, cette fusion permet, grâce à des mécanismes de combinaison, d'obtenir des informations de synthèse plus fiables afin d'améliorer la prise de décision.

Pour aborder les questions ci-dessus, nous proposons l'utilisation du cadre théorique de la théorie des fonctions de croyance appelée aussi théorie de l'évidence, qui permet de combiner un ensemble de connaissances ou de croyances des différentes sources d'information d'une manière relativement naturelle facilitant ainsi la représentation des imperfections de ces informations. De plus, cette théorie permet de prendre en compte la fiabilité des sources à fournir des informations sûres en introduisant des coefficients d'affaiblissement dans le but de réduire les masses de croyance propres aux sources non fiables.

Principales contributions

Dans les travaux de cette thèse, nous avons tout d'abord proposé une nouvelle mesure sans référence d'évaluation de la qualité d'image capable d'estimer la qualité des images dégradées avec de multiple types de distorsion. L'objectif est d'être le plus universel ou adaptatif possible concernant les dégradations prises en compte. Plusieurs caractéristiques (proposées dans la littérature) extraites de ces images et issues de différents espaces de représentation ont été utilisées. Dans un premier temps, nous appliquons la méthode de régression appelée *EV-reg*, basée sur la théorie de l'évidence, sur chacune de ces caractéristiques pour l'évaluation de la qualité d'image. À partir des premiers résultats expérimentaux, nous avons constaté que la performance diffère selon le vecteur de caractéristiques utilisé. Dans le but d'améliorer les performances de la méthode et tirer profit des outils de fusion que propose la théorie de l'évidence, nous introduisons une combinaison des différentes masses de croyances issues de chaque vecteur de caractéristiques. Dans cette approche nous tentons alors de prendre en compte la fiabilité de chaque ensemble de caractéristiques en introduisant un coefficient d'affaiblissement pour chaque vecteur de caractéristiques avant la combinaison et la prise de décision. Pour cela, nous discuterons de trois stratégies permettant d'estimer ce coefficient d'affaiblissement : un coefficient d'affaiblissement équivalent pour tout les vecteurs de caractéristiques, un coefficient d'affaiblissement paramétré en regard du pouvoir de la caractéristique à évaluer la qualité d'image à travers un coefficient de corrélation et enfin un coefficient d'affaiblissement paramétré en regard du pouvoir de la caractéristique à identifier le type de la dégradation à travers une matrice de confusion. À partir des différentes expérimentations, nous constatons que l'approche proposée basée sur les coefficients d'affaiblissement s'adaptant à la robustesse de la mesure présente une bonne corrélation avec les scores de qualité subjectifs et fournit des performances de prédiction de qualité compétitives par rapport aux mesures avec référence.

Le deuxième contexte, à savoir la stéganalyse, est différent car les modifications et les dégradations sont plus faibles, non visibles, en revanche la décision est simplement binaire. Nous allons alors utiliser un très grand nombre de mesures, chacune très élémentaires, tentant de détecter de faible variation. Pour cela, nous avons proposé un schéma de stéganalyse basé sur la théorie des fonctions de croyance construit sur des sous-espaces aléatoires des caractéristiques. L'approche générale consiste tout d'abord à sélectionner un ensemble de caractéristiques pertinentes qui peuvent aider à révéler

la présence de données cachées dans une image. Ensuite, nous formons un classifieur à l'aide d'un apprentissage automatique supervisé pour classer les images *cover* et *stégo*. Nous avons choisi d'utiliser l'algorithme des k plus proches voisins évidentiel (Ev-Knn) et réalisé un paramétrage de celui-ci à travers une étude empirique. Dans le but d'améliorer les performances, nous avons proposé d'implémenter la stratégie d'ensemble en construisant Ev-Knn à partir des sous-espaces aléatoires de caractéristiques. Les masses de croyances issues de chaque classifieur seront ensuite fusionnées avant de prendre la décision finale de la classe d'appartenance des images. Dans ce contexte, deux opérateurs de combinaison ont été utilisés : Dempster et combinaison moyenne. Le soucis est toujours de prendre en compte la fiabilité du sous-groupe des caractéristiques et de privilégier les plus robustes. Enfin, nous avons proposé d'introduire la classe de rejet qui reçoit les images qui ne peuvent pas être affectées à une classe. Dans un contexte de sécurité, il nous semble que la non-décision peut être acceptable. La performance de la méthode proposée a été évaluée sur différents algorithmes de dissimulation dans le domaine de transformé *JPEG* ainsi que dans le domaine spatial. Ces tests expérimentaux ont montré l'efficacité de la méthode proposée dans certains cadres d'applications.

Durant tout le manuscrit, nous mènerons de nombreuses expérimentations afin d'analyser, de valider et de critiquer les propositions et les méthodes introduites dans ce travail.

Organisation du manuscrit

Ce présent document est scindé en deux parties. La première est dédiée à l'évaluation de la qualité d'image. La seconde partie est consacrée à la stéganalyse. Ainsi, le manuscrit est organisé en trois chapitres :

- Dans le premier chapitre, nous faisons un bref état de l'art des méthodes d'apprentissage automatique à partir d'un ensemble de données. Nous présentons quelques principes clés de l'apprentissage automatique. Ensuite, nous présentons la théorie des fonctions de croyance qui est le cadre mathématique d'apprentissage automatique que nous avons adopté pour résoudre les problématiques abordées dans les travaux de cette thèse.
- Le deuxième chapitre s'intéresse à l'évaluation de la qualité d'image. Il permet de présenter quelques notions et définitions populaires ainsi qu'un état de l'art

de l'évaluation de la qualité d'image. Nous décrivons notre proposition d'évaluation automatique de la qualité d'image à l'aide des fonctions de croyance floues, ensuite nous présentons et discutons les résultats obtenus.

- Le troisième chapitre est dédié quant à lui à la stéganalyse. Il propose une présentation des principales méthodes d'insertion de messages cachés dans des médias numériques. Il dresse ensuite un bref état de l'art des méthodes de stéganalyse proposées dans la littérature. Ensuite, nous proposons un schéma de stéganalyse universel basé sur la théorie des fonctions de croyance. Enfin, différents tests et résultats expérimentaux de ce schéma seront présentés et commentés.
- Enfin, nous clôturons ce manuscrit en présentant une conclusion générale à ces travaux, ainsi que les différentes perspectives envisagées.

Chapitre 1

Apprentissage automatique et fonctions de croyance

Sommaire

1.1	Introduction	7
1.2	Qu'est ce que l'apprentissage automatique?	9
1.2.1	Les types d'apprentissage	10
1.2.2	Classification et régression : deux problèmes d'apprentissage supervisé	11
1.2.3	Construction d'un modèle pour un problème d'apprentissage	12
1.2.4	Apprentissage automatique et reconnaissance de forme	15
1.3	Théorie des fonctions de croyance	19
1.3.1	Fonctions de masse	20
1.3.2	Révision des croyances	23
1.3.3	Obtention d'une fonction de croyance	31
1.3.4	Niveau pignistique : Prise de décision	33
1.4	Conclusion	34

1.1 Introduction

Les méthodes d'apprentissage automatique sont des outils mathématiques puissants pour résoudre les problèmes d'approximation ou de prédiction. En particulier,

elles peuvent être utilisées en reconnaissance de forme, ou discrimination, dont l'objectif est de construire une bonne approximation de la relation fonctionnelle entre les ensembles des données d'entrée, qui sont connues, et des données de sortie. Pour cela, un ensemble de caractéristiques, extraites d'une forme inconnue, est exploité de manière à produire une hypothèse sur sa classe d'appartenance. Un système de reconnaissance de forme consiste donc à développer un ensemble de règles permettant une interprétation automatique des caractéristiques extraites de la forme inconnue pour reconnaître et estimer sa classe d'appartenance.

En reconnaissance de forme, deux approches peuvent être distinguées : les méthodes supervisées pour lesquelles les informations de sortie (étiquettes des classes) sont *a priori* connues et les méthodes non supervisées qui visent à faire d'elles-même l'apprentissage de ces informations.

De nombreuses applications ont été développées dans ce cadre touchant différents domaines : imagerie médicale, imagerie satellitaire ou radar, prédiction de trajectoire, prévision climatique, etc. Pour ces applications, la quantité d'informations disponibles, est de nos jours, de plus en plus importante. Ces informations peuvent être imprécises, incertaines, ou incomplètes. Dans certains cas, les données sont fournies par plusieurs capteurs ou experts avec différents degrés de précision ou de fiabilité. Pour traiter correctement un problème posé avec des connaissances imparfaites, hétérogènes et potentiellement contradictoires il est essentiel d'utiliser des formalismes adaptés. Par ailleurs, afin de profiter pleinement de l'ensemble des informations, il est nécessaire que ces formalismes puissent proposer des mécanismes de fusion (combinaison ou agrégation) permettant d'obtenir une information synthétique, qui résume l'ensemble des données connues afin d'aider à la prise de décision.

Au fil des années, plusieurs formalismes ont été proposés. D'un point de vue historique, la théorie des probabilités est la plus souvent utilisée. Malheureusement, cet outil ne permet pas de bien modéliser l'imperfection, notamment la modélisation de l'ignorance en l'absence d'information suffisante. Les limitations de la théorie des probabilités ont poussé certains chercheurs vers le développement d'autres formalismes tels que la théorie des fonctions de croyance [1, 2] appelée aussi théorie de l'évidence.

La théorie de l'évidence permet de représenter les connaissances de façon relativement naturelle ; elle n'impose aucune relation entre un évènement et son contraire, et l'ignorance est facilement modélisable.

Dans ce chapitre, nous présentons quelques principes clés de l'apprentissage au-

tomatique. Nous décrivons ensuite les bases théoriques de la théorie des fonctions de croyance, les différentes méthodes de révision de croyances et les méthodes de prise de décision correspondant au niveau *pignistique*.

1.2 Qu'est ce que l'apprentissage automatique ?

Pour un nouveau né, il faut en moyenne entre 6 et 24 mois pour prononcer ses premiers mots intelligibles. L'ensemble du processus de l'apprentissage de la prononciation, de la structure de la langue, à la construction de mots prend encore plus longtemps que cela ; il s'agit potentiellement d'un apprentissage et d'une évolution à vie. Si nous suivons cet exemple, la quantité d'information que le cerveau du bébé doit traiter est énorme. Si nous nous limitons seulement à l'exemple de l'apprentissage de la langue, le cerveau doit traiter les mouvements musculaires de la bouche (imitation de l'entourage), en se souvenant des séquences de phonèmes pour créer des mots, et associer ces mots avec le contexte et le contenu.

Au moment d'écrire ces lignes, hormis le cerveau humain, il n'existe pas de machine capable de faire un tel apprentissage automatiquement. Bien que les progrès de la synthèse vocale et la modélisation de la langue aient permis la création de systèmes qui semblent être humains, nous sommes encore très loin de créer une structure réelle capable d'apprendre tout un langage de type humain par elle-même.

L'idée d'imiter les capacités humaines telles que l'apprentissage de la langue par des machines est généralement appelée *Intelligence Artificielle*, dont l'apprentissage automatique est une sous-classe. L'apprentissage automatique vise à observer des exemples d'un phénomène spécifique et à apprendre afin de modéliser le processus sous-jacent.

Présentons un autre exemple intéressant donné dans la partie introductive de [3], qui couvre la plupart des aspects du problème de l'apprentissage automatique. Considérons une chaîne de supermarchés importante, ayant beaucoup de magasins dans de nombreux pays. Pour commercialiser le bon produit au bon moment et au bon client, le marchand a besoin de savoir ce que les clients achètent, quand et peut-être pourquoi. Les caissiers de chaque terminal enregistrent toutes les informations sur ce qui a été acheté, quand et par quel client, ce qui va engendrer des grandes masses d'informations sachant que des millions de clients passent par jour.

Dans toutes ces données, seule une faible quantité d'informations est en fait perti-

nente et utile, à des fins publicitaires et d'optimisation du chiffre d'affaires. Bien qu'il soit très difficile de modéliser complètement le comportement d'un client sur la base de ces données, une assez bonne approximation du comportement est déjà suffisante. Et une fois qu'il a été identifié qu'un tel client achète des produits à base de chocolat pendant les jours sombres de l'hiver et de la bière et des saucisses quand les premiers jours de l'été s'approchent, il devient plus facile de cibler la publicité et les offres pour ce client spécifique.

Dans le cadre de cette thèse, nous étudions l'application de l'apprentissage automatique à deux problématiques liées à l'analyse des images. La première concerne l'évaluation de la qualité d'image. Il s'agit de mesurer l'impact des différentes dégradations que subit l'image tout au long la chaîne de vie (compression, traitement et transmission). Pour cela, nous cherchons à construire un modèle de régression capable de prédire un score de qualité des images numériques à partir d'un espace des caractéristiques extraites de ces images. La seconde problématique concerne la stéganalyse qui consiste à détecter la présence d'un message caché dans une image. L'objectif ici est de proposer un schéma à partir des caractéristiques extraites des images numériques. Ce schéma doit être capable d'apprendre la différence entre une image de couverture et une image stéganographiée porteuse d'un message caché (*stégo*) afin de pouvoir classer correctement ces images.

1.2.1 Les types d'apprentissage

Dans cette thèse, nous considérons que le problème est décrit par un ensemble de données, qui prend la forme d'une matrice \mathcal{X} de dimension $N \times M$, appelés entrées (ou des données d'entrée), où N représente le nombre des échantillons (exemples du phénomène observé), et M la dimension du vecteur de caractéristiques extrait de chaque échantillon.

Nous pouvons distinguer deux principales classes de problèmes d'apprentissage : apprentissage non supervisé et apprentissage supervisé [4].

Apprentissage non supervisé

Cette classe spécifique de l'apprentissage fait l'hypothèse que seules les données \mathcal{X} sont disponibles et tente d'en déduire une structure ou un comportement sous-jacent

dans ces données.

Un exemple classique de l'apprentissage non supervisé est le problème de clustering qui consiste à diviser l'ensemble de données de manière à créer des clusters d'individus similaires. Dans l'exemple précédent du supermarché vendant des marchandises à travers de nombreux pays, il pourrait être intéressant de trouver des groupes de clients ayant des attitudes spécifiques à analyser : des groupes de clients ayant des comportements similaires identifiés seront plus faciles à cibler en termes de stratégies de vente, tandis que d'autres groupes formés dans le processus de regroupement peuvent révéler un comportement différent d'une partie de la population, ce qui conduit à une nouvelle stratégie de vente, plus adaptée et spécifique.

Apprentissage supervisé

L'apprentissage supervisé se distingue par la disponibilité de données d'apprentissage étiquetées. Le nom invoque l'idée d'un «superviseur» qui instruit le système d'apprentissage sur les étiquettes, désignées par \mathcal{Y} , à associer à des exemples d'apprentissage. Les algorithmes d'apprentissage supervisés induisent des modèles à partir de ces données d'apprentissage et ces modèles peuvent être utilisés pour classer d'autres données non marquées (étiquetées).

Cette deuxième classe de problème d'apprentissage est celle considérée dans le reste de ce manuscrit.

Nous définissons dans la section qui suit les deux scénarios possibles dans un problème d'apprentissage automatique.

1.2.2 Classification et régression : deux problèmes d'apprentissage supervisé

Dans l'hypothèse d'un problème d'apprentissage supervisé, nous pouvons définir deux cas d'apprentissage supervisé, en fonction du type et la dimension de la matrice de sortie \mathcal{Y} .

Le premier cas est appelé *régression*. Dans ce contexte, la sortie \mathcal{Y} est généralement un vecteur de valeurs réelles. Cela signifie que le modèle \mathcal{M} vise alors à prédire des valeurs dans \mathbb{R} :

$$\mathcal{M} : \mathbb{R}^{N \times M} \rightarrow \mathbb{R} \tag{1.1}$$

avec N le nombre d'échantillons et M la dimension du vecteur des caractéristiques. Il s'agit de trouver la fonction qui minimise un critère d'erreur entre les valeurs de la fonction aux points d'apprentissage et les observations pour effectuer ensuite des prédictions.

Le second cas d'apprentissage supervisé est la *classification*. Cette fois ci, les sorties \mathcal{Y} sont qualitatives en attribuant une étiquette ou une classe à chaque entrée de l'ensemble d'apprentissage.

Si nous revenons maintenant à notre problématique de thèse, à savoir tout d'abord l'évaluation de la qualité d'image, les données de sortie représentent des scores appelés DMOS (Difference Mean Opinion Score)¹. Ces valeurs sont dérivées de la moyenne des notes subjectives attribuées par plusieurs observateurs humains, à qui il a été demandé de juger la qualité des images suivant une grille d'évaluation à plusieurs niveaux d'appréciations. Notre objectif est alors d'estimer la note de qualité d'une nouvelle image en corrélation avec l'évaluation subjective.

Quant à la stéganalyse qualitative (voir section 3.4.1), chaque image de l'ensemble d'apprentissage est associée à une étiquette indiquant la présence ou non d'un message caché dans cette image, autrement dit la classe d'appartenance de cette image (*cover* ou *stégo*). L'objectif est alors d'estimer la classe d'appartenance d'une nouvelle image en se basant sur les informations apportées par l'ensemble d'apprentissage.

Nous constatons que nous allons traiter les deux cas d'application : la régression et la classification. Avant de présenter le formalisme mathématique utilisé dans notre travail de thèse, nous présentons les différentes étapes nécessaires pour construire un modèle d'apprentissage automatique.

1.2.3 Construction d'un modèle pour un problème d'apprentissage

Dans un problème d'apprentissage, le processus de construction d'un modèle nécessite au moins trois étapes : le paramétrage du modèle, la construction du modèle et enfin la validation de ces choix. Chaque étape est expliquée dans ce qui suit.

1. Nous définirons cette notion plus précisément dans la suite

Paramétrage du modèle et critère d'erreur

Le choix de la structure du modèle constitue la première étape. Elle est intimement liée à l'étape de validation du modèle.

Un modèle a généralement un certain nombre d'hyper-paramètres, liés aux choix de la structure du modèle (par exemple l'ordre d'une courbe polynomiale pour l'ajuster aux données) et des paramètres qui doivent être déterminés en utilisant les données elles-mêmes (les coefficients de la courbe polynomiale d'ordre fixe).

Le but de cette étape est de trouver un éventuel ensemble optimal de l hyper-paramètres $\Theta = (\theta_1, \dots, \theta_l)$ de telle sorte que le modèle $\mathcal{M}(\mathcal{X}, \Theta)$ soit associé à la plus petite erreur sur les données \mathcal{X} , en ce qui concerne la sortie \mathcal{Y} . L'erreur ϵ_r peut être définie par une fonction de risque f_R qui quantifie (mesure) la différence entre l'estimation du modèle $\hat{y} = \mathcal{M}(\mathcal{X}, \Theta)$ et la valeur réelle de y :

$$\epsilon_r = f_R(\hat{y}, y) = f_R(\mathcal{M}(\mathcal{X}, \Theta), y) \quad (1.2)$$

Dans le cas d'un problème de régression, la fonction de risque f_R est représentée généralement par la fonction de l'erreur quadratique moyenne (EQM) définie par :

$$f_R(\hat{y}, y) = f_{EQM}(\hat{y}, y) = \frac{1}{N} \sum_{j=1}^N (\hat{y}_j - y_j)^2, \quad (1.3)$$

où \hat{y}_j est le $j^{\text{ème}}$ élément du vecteur $[\hat{y} = (\hat{y}_1, \dots, \hat{y}_N)^T]$ estimé par le modèle. y_j est le $j^{\text{ème}}$ élément du vecteur des observations $\mathcal{Y} = (y_1, \dots, y_N)^T$.

Dans le cas d'un problème de classification à deux classes, l'erreur est obtenue à partir du taux de bonne classification ($taux_{clas}$) donnée par le ratio :

$$taux_{clas} = \frac{T_P + T_N}{N}, \quad (1.4)$$

où T_P (true positive) et T_N (true negative) représentent respectivement le nombre d'éléments (prédits par le modèle \mathcal{M}) correctement classés et le nombre d'éléments mal classés. N représente le nombre total des observations. L'erreur de classification ϵ_c peut alors être définie par :

$$\epsilon_c = 1 - taux_{clas}, \quad (1.5)$$

Exprimé en pourcentage, il correspond au "pourcentage de classifications erronées". À noter que le problème de minimisation de l'erreur ϵ est différent pour les cas de la

régression et de la classification présentés ici. En effet, la plupart des modèles préfèrent minimiser l'EQM ϵ_r , même pour le cas de la classification, et éventuellement calculer l'erreur de classification ϵ_c obtenue en utilisant les mêmes hyper-paramètres $\Theta = \Theta_r$ qui minimisent l'EQM.

Dans le cas d'un problème de classification binaire, la minimisation de l'EQM ne pose pas un vrai problème. La minimisation de l'EQM conduit de toute façon à minimiser les écarts entre \hat{y} et y .

Construction du modèle

Une fois que les critères de la minimisation de l'erreur sont déterminés et l'ensemble d'hyper-paramètres est défini, le modèle peut être construit sur les données \mathcal{X} .

Il est important lors de la construction du modèle de diviser l'ensemble des données originales \mathcal{X} de sorte que la construction du modèle et la validation se produisent sur deux sous-ensembles différents des données. On peut dire que les hyper-paramètres ("externes" au modèle) sont optimisés pour minimiser l'erreur de validation, alors que les paramètres ("internes" au modèle) sont conçus pour minimiser l'erreur de l'apprentissage, calculée au cours de la construction du modèle.

Validation du modèle

Enfin une étape de validation intervient pour évaluer l'erreur ϵ .

Afin d'estimer le modèle \mathcal{M} pour un ensemble fixe d'hyper-paramètres Θ , nous avons besoin d'un ensemble de données d'apprentissage \mathcal{X}_{app} , qui sera un sous-ensemble de l'ensemble des données \mathcal{X} . Le modèle est donc construit sur \mathcal{X}_{app} et l'erreur d'apprentissage ϵ_{app} est calculée avec $\epsilon_{app} = \mathcal{M}(\mathcal{X}_{app}, \Theta)$. Pour la plupart des modèles de l'apprentissage automatique, il est possible d'obtenir une erreur $\epsilon_{app} > 0$ aussi petite que possible, en prenant suffisamment de temps pour trouver le bon jeu d'hyper-paramètres $\Theta = \Theta_{app}$ qui minimise ϵ_{app} . Le problème est que le modèle continue à améliorer ses performances sur les échantillons d'apprentissage donc il apprend également le bruit dans les données.

Ce problème peut être surmonté par l'utilisation d'un ensemble de données de validation.

Revenons à notre ensemble complet de données \mathcal{X} , nous proposons donc de le di-

viser au préalable en deux sous-ensembles différents : \mathcal{X}_{app} pour l'apprentissage du modèle, et \mathcal{X}_{val} pour valider le modèle. L'utilisation de ces nouvelles données (jamais exploitées par le modèle au cours de l'apprentissage) aide à correctement estimer l'erreur du modèle avec ces hyper-paramètres Θ : le sur-apprentissage du modèle sur les données d'apprentissage conduit inévitablement une grande erreur sur l'ensemble de validation. Par conséquent, l'ensemble des hyper-paramètres doit être modifié de sorte que le modèle se généralisera bien sur les nouvelles données.

Ce processus standard de validation peut être généralisé par une approche appelée validation croisée.

La validation croisée est un processus généralement effectué k fois (k validations croisées), où k détermine le nombre de sous-ensembles différents. Dans le cas d'un schéma de 3-validation croisées, les données de l'ensemble \mathcal{X} sont divisées en trois sous-ensembles égaux, dont le premier tiers sera utilisé comme ensemble de données de validation \mathcal{X}_{val} , et les deux tiers restants pour l'apprentissage \mathcal{X}_{app} . Le procédé est répété en faisant tourner les ensembles k fois, $k = 3$ pour cet exemple, et l'erreur moyenne sur les trois essais est rapportée. Il est évident que le nombre de fois (k) où le processus d'apprentissage et de la validation est répété peut être problématique pour les modèles qui nécessitent un grand temps de calcul.

Nous proposons maintenant un bref aperçu des stratégies de reconnaissance de forme.

1.2.4 Apprentissage automatique et reconnaissance de forme

Nous présentons dans cette section les différentes approches utilisées en reconnaissance de forme. Cette présentation est loin d'être exhaustive, mais nous permet d'avoir un aperçu des approches les plus étudiées et utilisées en reconnaissance de forme.

Approches probabilistes

La théorie des probabilités a pour objectif d'estimer les probabilités d'appartenance des observations aux classes. Elle repose sur des bases théoriques régies par le théorème de Bayes. Les informations sont modélisées par le biais d'une probabilité.

Dans le cas d'une classification, ces méthodes s'appuient sur la définition et l'estimation de densités de probabilités des vecteurs de caractéristiques associés aux don-

nées numériques. La tâche consiste à affecter un vecteur de caractéristique à l'une des classes.

Soient K classes connues, notées $\omega_1, \dots, \omega_K$, auxquelles on associe les probabilités *a priori* des classes $P(\omega_i)$, et $f(x|\omega_i)$ les densités des probabilités conditionnelles des vecteurs x aux classes (x représente le vecteur de caractéristiques). Ces probabilités sont issues soit d'un apprentissage paramétrique dans lequel il est supposé que $f(x|\omega_i)$ suit une loi connue, soit d'un apprentissage non paramétrique dans lequel aucune hypothèse de loi n'est faite.

Les probabilités *a posteriori* d'appartenance $p(\omega_i|x)$ sont ensuite estimées grâce à la théorie bayésienne. La décision revient à choisir la classe qui maximise ces probabilités.

Les réseaux de neurones

Les réseaux de neurones artificiels regroupent une grande variété de modèles qui utilisent les neurones comme unité de calcul élémentaire [5]. Bien que cette classe de modèles ait été historiquement inspirée des processus biologiques, elle fait maintenant partie intégrante du cadre mathématique de l'apprentissage automatique. Les réseaux neuronaux peuvent être utilisés dans le contexte de l'apprentissage supervisé pour la classification et la régression, ou bien dans des contextes non supervisés.

Un réseau de neurones peut être décrit comme un graphe dirigé dont les noeuds correspondent aux neurones et les arêtes correspondent à des liens entre eux. Chaque neurone reçoit en entrée une somme pondérée (poids) des sorties des neurones reliées à ses bords entrants.

Étant donné un ensemble d'entrées $x \in \mathbb{R}^n$ et une sortie $y \in \mathbb{R}$, le neurone associe un poids w_i à chaque entrée x_i et calcule la somme pondérée des entrées par leurs poids respectifs comme suit :

$$y = \sigma(b + \sum_i w_i x_i) \tag{1.6}$$

avec σ une fonction appelée fonction d'activation et b appelé biais du neurone. Plusieurs fonctions d'activation ont été utilisées en fonction du domaine d'application, telle que la fonction sigmoïde.

Plusieurs d'autres modèles de réseaux de neurones artificiels ont été développés. Ces modèles sont utilisés comme brique de base pour la conception des algorithmes de

reconnaissance de forme. Les plus simples sont des améliorations du modèle ci-dessus en remplaçant la fonction d'activation par une autre fonction non-linéaire. D'autres modèles plus complexes ont été proposés tel que les réseaux à base radial (Radial basic function) [5], etc.

Apprentissage profond (*Deep learning*)

Dernièrement, grâce aux récentes avancées dans le domaine des réseaux de neurones, et à la puissance des calculs fournie par les GPU, des approches d'apprentissage en profondeur (*Deep learning*) ont été proposées comme une extension des réseaux de neurones. Ils sont devenus de plus en plus populaires en raison de leur haute performance de classification. Les réseaux d'apprentissage en profondeur sont de "gros" réseaux de neurones qui peuvent prendre directement des données en entrée. Dans le traitement d'image, le réseau prend les valeurs des pixels en entrée [6]. Un réseau d'apprentissage en profondeur gère deux étapes à la fois (extraction des caractéristiques et classification). Depuis 2006, de nombreux ajustements ont été proposés pour améliorer la robustesse et de réduire les coûts de calcul.

L'un des modèles les plus répandus de l'apprentissage profond dans le domaine de la classification des images sont les réseaux de neurones convolutionnels (*Convolutional Neural Networks* (CNN)) [5]. La méthodologie d'apprentissage est similaire à la méthode classique. Une base de données d'images est nécessaire, avec pour chaque image, son label (c'est-à-dire sa classe). Chaque image est donnée en entrée du réseau ; dans ce cas, chaque valeur de pixel est prise comme entrée d'un ou de plusieurs neurones. Le réseau est constitué d'un nombre donné de couches. Une couche se compose de neurones qui prennent des valeurs d'entrée, effectue des calculs, puis retourne des valeurs qui sont fournies à la couche suivante. Plus précisément, au sein d'une couche, les calculs se font en trois étapes successives : une étape de convolution, l'application d'une fonction d'activation, puis une étape de sous-échantillonnage.

Notons que la mise en oeuvre de l'apprentissage profond nécessite un accès à une très grande base de données, ce que ne permet pas d'adopter cette approche, aussi efficace soit-elle, dans tous les problèmes d'apprentissage.

Il existe bien sûr d'autres stratégies en reconnaissance de formes, mais nous avons présenté ici les principales utilisées dans les applications étudiées durant cette thèse.

Limites des modèles probabilistes

Les méthodes de reconnaissance de forme utilisées pour estimer les relations de dépendance fonctionnelle entre les variables sont généralement déterminées à partir de modèles probabilistes. Bien que la théorie des probabilités repose sur des bases mathématiques solides elle présente plusieurs inconvénients. Parmi les inconvénients nous pouvons citer que la théorie des probabilités raisonne seulement sur les singletons, ce qui ne permet pas de résoudre des cas où un vecteur appartient à plusieurs classes à la fois. Ensuite, dans la théorie des probabilités, si nous connaissons la probabilité d'un événement A , le complément de cette probabilité est automatiquement attribué à \bar{A} , ce qui rend la représentation de l'ignorance dans ce cas-là difficilement modélisable.

Une observation peut être imprécise, incertaine, ou incomplète. Ces situations d'exploitation de connaissances imparfaites doivent être bien représentées dans un système d'apprentissage. Dans certains cas les données sont fournies par plusieurs capteurs ou experts avec différents degrés de précision, nous avons besoin alors de formalismes capables de traiter ces données imprécises et/ou partiellement contradictoires. En outre, afin de profiter au maximum de toutes les informations (vecteurs) de l'entrée, il est également nécessaire que ces formalismes fournissent des mécanismes de fusion (combinaison ou agrégation). Cette phase de fusion permet d'obtenir des informations de synthèse, qui résument l'ensemble de données initial, ce qui est essentiel afin de prédire la sortie y .

Des formalismes tels que la théorie des possibilités ou la théorie de l'évidence ont été développés pour répondre aux limitations de la théorie des probabilités.

Approches possibilistes

La théorie des possibilités [7, 8] permet de tenir compte de l'imprécision des données ainsi que de l'incertitude, en affectant à une donnée un degré d'appartenance relatif à une hypothèse. La théorie des possibilités repose sur les fonctions de possibilité et de nécessité obtenues à partir d'un ensemble d'hypothèses considérées $\Omega = \{H_1, H_2, \dots, H_K\}$.

La théorie des possibilité offre un important choix d'opérateurs de combinaison tels que le t-normes et t-conormes généralisant les opérateurs classiques d'intersection et d'union de la théorie des ensembles[9]. D'autres opérateurs de type moyenne, sommes symétriques, hybrides, etc... peuvent également être utilisés. Pour la prise de décision, le choix d'une classe se fait à partir du maximum de degrés d'appartenance.

Approches évidentielles

La théorie des fonctions de croyance ou théorie de l'évidence permet de représenter de façon relativement naturelle les connaissances. En reprenant les défauts recensés précédemment pour la théorie des probabilités, la théorie des fonctions de croyance n'impose aucune relation entre un événement et son contraire et elle permet de modéliser facilement l'ignorance. Les différents formalismes présentés sont capable représenter et gérer les imperfections ; ils ne sont ni contradictoires ni à mettre en concurrence. Chacun des formalismes possède sa spécificité, et aucun n'est supérieur à l'autre.

La théorie des fonctions de croyance est considérée comme unificatrice de l'ensemble des théories de l'incertain (les théories de probabilité et des possibilités) puisque l'on retrouve ces derniers comme des cas particuliers.

Pour ces raisons, nous proposons dans cette étude de nous placer dans le cadre théorique des Fonctions de Croyance afin de prendre en considération l'aspect imparfait (imprécis et incertain) des connaissances.

Dans ce qui suit, nous présentons les notions et les bases théoriques des fonctions de croyance. Nous introduisons ainsi les différentes méthodes de révision des croyances. Ensuite, un bref état de l'art de la modélisation des données dans le cadre des fonctions de croyance sera présenté, ainsi qu'une description des méthodes de décision, correspondant au niveau *pignistique*.

1.3 Théorie des fonctions de croyance

La théorie des fonctions de croyance, également appelée théorie de Dempster-Shafer ou théorie de l'évidence, introduite par Shafer [1], a pour origine les travaux de Dempster [2]. Inspirée de la notion des probabilités, qui permet de représenter les connaissances incertaines, la théorie des fonctions de croyance permet également la représentation de connaissances entachées d'imprécisions. La théorie est née des travaux de Dempster sur les bornes inférieures et supérieures, puis formalisée par Shafer et reprise par la suite par Kohlas et Monney [10]. Une autre extension de la théorie des fonctions de croyance a été proposée par Smets [11] appelée modèle de croyance transférable. Ce modèle fait la distinction entre la représentation des connaissances et la prise de décision.

Le modèle de croyance transférable repose sur deux niveaux de perception des in-

formations. D'un côté, le niveau *crédal* permet la mise en forme et la révision des informations. De l'autre, le niveau *pignistique* est entièrement dédié à la prise de décision et est clairement séparé de la modélisation des données.

Dans la suite de ce chapitre, nous présentons les différentes notions et les bases théoriques de la théorie des fonctions de croyance.

1.3.1 Fonctions de masse

Définition du cadre de discernement

Soit une question Q pour laquelle nous cherchons une réponse. Soit $\Omega = \{w_1, w_2, \dots, w_K\}$ un ensemble fini contenant K solutions possibles à cette question. Cet ensemble est appelé cadre de discernement. Les hypothèses, ou ensembles de solutions, considérées dans le cadre de la théorie des fonctions de croyance sont données par :

$$2^\Omega = \{A \mid A \subseteq \Omega\} = \{\emptyset, \{w_1\}, \dots, \{w_K\}, \{w_1 \cup w_2\}, \dots, \Omega\}. \quad (1.7)$$

A correspond à une hypothèse pouvant être simple appelée singleton, ou multiple (au moins deux solutions) appelée disjonction. Il est à noter que A peut aussi correspondre à l'ensemble vide (\emptyset), ce qui veut dire que la réponse à la question Q se trouve en dehors de Ω .

Fonction de masse de croyance

Une fonction m , appelée fonction de masse (en anglais basic belief assignment ou bba) représente une connaissance imparfaite sur Ω . Une masse allouée à une hypothèse A est appelée "masse de croyance" et est définie par :

$$\begin{aligned} m : 2^\Omega &\rightarrow [0, 1] \\ A &\mapsto m(A) \end{aligned} \quad (1.8)$$

telle que :

$$\sum_{A \subseteq \Omega} m(A) = 1 \quad (1.9)$$

La quantité $m(A)$ représente la croyance sur le fait que A contienne la réponse à la question Q . À la différence des théories de probabilités, il est possible d'allouer de la

croyance non seulement à des singletons mais aussi à des sous-ensembles de Ω , ce qui traduit l'ignorance que nous avons vis-à-vis des sous-hypothèses de A . Un ensemble A tel que $m(A) > 0$ est appelé élément focal. Il est à noter qu'à l'origine, Shafer [1] a imposé la contrainte supplémentaire $m(\emptyset) = 0$ faisant référence au *monde fermé* ce qui signifie que l'ensemble Ω est exhaustif. Une fonction de masse satisfaisant cette propriété est appelée *normalisée*. Smet [12] a relâché cette contrainte dans le cadre du modèle de croyance transférable et suppose que l'ensemble Ω peut être incomplet (*monde ouvert*).

Tous les modèles présentés dans ce travail sont basés sur l'hypothèse du monde fermé. Cela signifie que le cadre de discernement est supposée être exhaustive. En conséquence, les fonctions de croyance sont normalisées.

Transformations de la fonction de masse

Nous noterons $F(m)$ l'ensemble des éléments focaux de m . Les informations fournies par une fonction de masse peuvent être représentées par une fonction de crédibilité $bel(A)$ (pour *belief*) ou par une fonction de plausibilité $pl(A)$.

La fonction de crédibilité est définie par :

$$bel(A) = \sum_{\emptyset \neq B \subseteq A} m(B) \quad \forall A \subseteq \Omega, A \neq \emptyset \quad (1.10)$$

La quantité $bel(A)$ représente la quantité totale de croyance strictement allouée à A . La figure 1.1 illustre graphiquement la définition ci-dessus appliquée pour le calcul de la croyance de l'ensemble A5.

Tout comme il est possible d'obtenir une fonction de crédibilité bel d'une fonction de masse m en utilisant l'équation (1.10), il est également possible de calculer une fonction de masse à partir d'une fonction de crédibilité bel . Cette conversion est basée sur la transformée de Möbius [13] définie par :

$$m(A) = \sum_{B \subseteq A} (-1)^{|A-B|} bel(B) \quad \forall A \subseteq \Omega \quad (1.11)$$

où $|A - B|$ est le nombre d'éléments de A qui n'appartiennent pas à B .

La fonction de plausibilité est quant à elle la quantité de croyance qui pourrait potentiellement être allouée à A . Elle correspond à l'ensemble des croyances qui ont

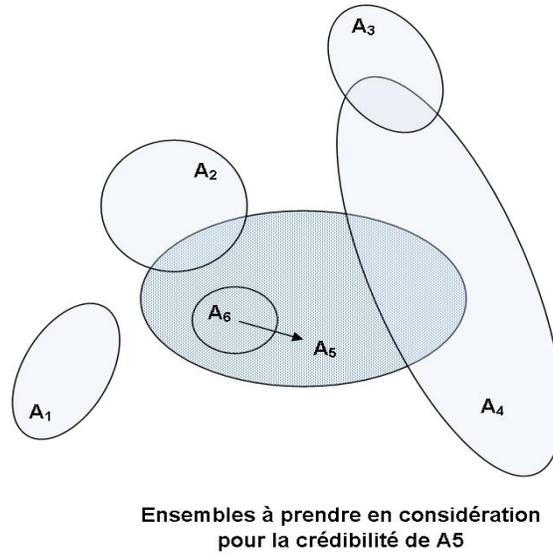


FIGURE 1.1 – Illustration graphique de la fonction de crédibilité bel .

une intersection avec l'élément considéré. Cette fonction est définie par :

$$pl(A) = \sum_{A \cap B \neq \emptyset} m(B) \quad \forall A \subseteq \Omega \quad (1.12)$$

Ceci est illustré par un exemple dans la figure 1.2.

Une fonction de plausibilité peut être déterminée à partir de la fonction de crédibilité bel grâce à la relation :

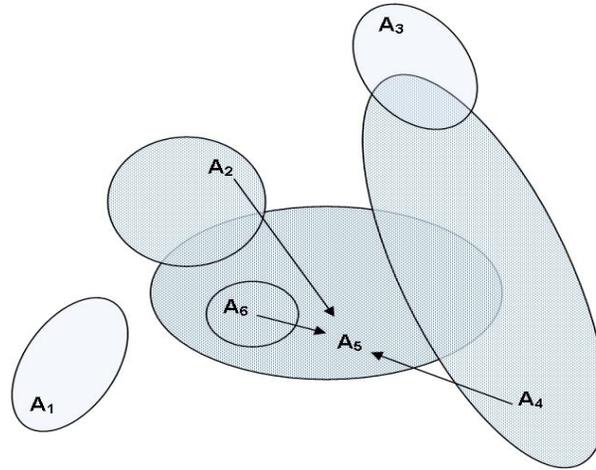
$$pl(A) = bel(\Omega) - bel(\bar{A}), \quad \forall A \subseteq \Omega. \quad (1.13)$$

où \bar{A} est le complément de A .

Pour le cas particulier des fonctions de plausibilité normalisées où $bel(\Omega) = 1$, cette définition se réduit à :

$$pl(A) = 1 - bel(\bar{A}), \quad \forall A \subseteq \Omega. \quad (1.14)$$

La différence $pl(A) - bel(A)$ représente l'ignorance sur une hypothèse précise A . Par exemple, si une source d'information nous permet d'écrire $pl(A) = 1$ et $bel(A) = 0$ pour un ensemble A , cela signifie que la source ne fournit aucune information supplémentaire sur A . Cela peut être interprété comme l'ignorance totale de la source concernant l'hypothèse A . Lorsqu'une source ne peut pas affecter ses connaissances à un ensemble



Ensembles à prendre en considération
pour la plausibilité de A5

FIGURE 1.2 – Illustration graphique de la fonction de plausibilité pl .

plus petit que Ω , il assigne : $m(\Omega) = 1$, ce qui implique $m(A) = 0, \forall A \in \Omega, A \neq \Omega$ et $bel(A) = 0, pl(A) = 1$. En d'autres termes, la source sait que la solution (l'hypothèse) se trouve dans l'ensemble de discernement, mais ne peut rien dire de plus. Cela se réfère également à l'ignorance totale de la source. À l'opposé, $m(A) = 1$ indique la certitude totale de la source sur l'hypothèse A .

1.3.2 Révision des croyances

Il existe différents mécanismes de révision des croyances. Nous présentons ici deux méthodes : la fusion et l'affaiblissement.

Afin de regrouper les informations provenant de différentes sources, les fonctions de croyance issues des différentes sources d'informations doivent être combinées. Généralement, chaque source d'information est représentée par une fonction de croyance distincte. Les règles de combinaison sont ensuite utilisées pour fusionner successivement toutes ces fonctions de croyance afin d'obtenir une fonction de croyance globale représentant toutes les connaissances disponibles.

Dans les sections suivantes nous présentons quelques règles de combinaison dans le cadre des fonctions de croyance. Nous commençons par la règle de Dempster.

La règle de combinaison de Dempster

La règle de combinaison de Dempster a été introduite dans [2], puis réinterprétée par Shafer comme base de la théorie des fonctions de croyance. Elle permet de combiner les fonctions de croyance normalisées qui sont définies dans le même cadre de discernement.

Soient deux fonctions de masses m_1 et m_2 issues de deux sources fiables et distinctes, S1 et S2, et ayant le même cadre de discernement Ω . La fonction de masse $m_{1\oplus 2} = m_1 \oplus m_2$ combinée selon la règle de Dempster \oplus est définie par :

$$m_{1\oplus 2}(A) = \frac{1}{1-k} \sum_{B \cap C = A} m_1(B)m_2(C) \quad \forall A \subseteq \Omega, A \neq \emptyset \quad (1.15)$$

où

$$k = m_{1\oplus 2}(\emptyset) = \sum_{B \cap C = \emptyset} m_1(B)m_2(C) \quad (1.16)$$

Ici, le terme k représente le conflit entre les sources d'information. Il représente le produit des valeurs de masse correspondant à toutes les intersections vides des éléments focaux. La règle de Dempster est commutative, associative, et possède un élément neutre ($m(\Omega)$) qui représente l'ignorance totale. Dans le cas où il y a plus de deux sources d'information, les fonctions de croyance correspondantes sont simplement combinées successivement. L'ordre de combinaison est sans importance.

Combinaison conjonctive

Dans le cadre du modèles des croyances transférables, la combinaison conjonctive a été proposée. Elle est identique à la règle de Dempster. La différence réside en l'étape de normalisation qui est omise.

Soient m_1 et m_2 deux fonctions de masses issues de deux sources fiables et distinctes, S1 et S2, et ayant le même cadre de discernement Ω . La fonction de masse $m_{1\odot 2} = m_1 \odot m_2$ résultante de la combinaison conjonctive \odot est définie par :

$$m_{1\odot 2}(A) = \sum_{B \cap C = A} m_1(B)m_2(C) \quad (1.17)$$

La combinaison conjonctive est commutative et associative, ce qui permet de généraliser la combinaison à p sources. Il est à rappeler que l'ordre n'est pas important pour

réaliser la combinaison. Nous avons ainsi :

$$m_{\odot}(A) = \sum_{(A_1 \cap A_2 \cap \dots \cap A_p)} \left(\prod_{i=1}^p m_i(A_i) \right) \quad (1.18)$$

Par ailleurs il est important de noter que cette règle de combinaison produit une quantité de croyance sur l'ensemble vide ($m_{\odot}(\emptyset)$) qui s'interprète comme le conflit. Cette situation n'est pas possible dans le cas du monde fermé, en utilisant la combinaison orthogonale de Dempster, visant à normaliser les masses générées par le conflit.

Combinaison disjonctive

La règle de combinaison disjonctive est appliquée lorsque l'on sait qu'au moins l'une des sources d'informations est fiable, mais sans savoir laquelle. Pour obtenir une fusion disjonctive, Smets [14] a proposé de prendre l'union d'hypothèses plutôt que leur intersection. Cette règle permet de transférer les masses sur les plus grands sous-ensembles. Par conséquent l'incertitude et l'imprécision augmentent. La fonction de masse $m_{1 \odot 2} = m_{1 \odot} m_2$ résultante de la combinaison conjonctive \odot est définie par :

$$m_{1 \odot 2}(A) = \sum_{B \cup C = A} m_1(B) m_2(C) \quad (1.19)$$

Puisque que l'union $B \cup C$ n'est jamais vide, à moins que les deux éléments focaux soient vides, il n'y a pas conflit généré par la règle de combinaison disjonctive. Il n'y a donc pas besoin d'une étape de normalisation.

La règle de Yager

Yager propose une règle de combinaison qui affecte la masse associée à des ensembles focaux contradictoires au cadre de discernement Ω (au lieu d'effectuer une normalisation ou de l'attribuer à \emptyset) [15]. Soit m_Y le résultat de la combinaison de deux fonctions de masse m_1, m_2 provenant de deux sources distinctes. La règle de Yager exprimée au moyen de la règle conjonctive est définie par :

$$\begin{cases} m_Y(A) = m_{1 \odot 2}(A) & \forall A \subset \Omega, A \neq \emptyset, \\ m_Y(\Omega) = m_{1 \odot 2}(\Omega) + m_{1 \odot 2}(\emptyset) & \text{si } A = \Omega, \\ m_Y(A) = 0 & \text{si } A = \emptyset. \end{cases} \quad (1.20)$$

Cela signifie que des sources très contradictoires conduisent à un état de grande ignorance en attribuant le conflit lié à la masse $m_{1 \odot 2}(\emptyset)$ à l'ensemble Ω .

La règle de Dubois et Prade

La règle de combinaison proposée par Dubois et Prade est proche de la règle de Yager. Cependant, au lieu d'affecter la masse des éléments focaux contradictoires au cadre du discernement, il est affecté à l'union des éléments focaux correspondants. Soit m_{DP} le résultat de la combinaison de deux fonctions de masse m_1, m_2 . La règle de Dubois et Prade exprimée en utilisant la combinaison conjonctive est définie par :

$$m_{DP}(A) = \begin{cases} m_{1 \odot 2}(A) + \sum_{B \cap C = \emptyset, B \cup C = A} m_1(B).m_2(C) & \forall A \subseteq \Omega, A \neq \emptyset, \\ 0 & \text{si } A = \emptyset. \end{cases} \quad (1.21)$$

Pour le cas où $|\Omega| = 2$, la règle de Yager et la règle de Dubois et Prade sont identiques.

Notons que la plupart des règles de combinaison même si calculatoirement ressemblent à la règle de Dempster, diffèrent par la façon dont elles gèrent les situations de conflits (informations contradictoires).

L'une des raisons pour lesquelles des nouvelles règles de combinaison continuent à être proposées au fil du temps peut s'illustrer par la critique de Zadeh de la règle de Dempster lorsqu'elle est confrontée à des situations très contradictoires. Dans l'exemple de Zadeh [15], deux médecins diagnostiquent indépendamment le même patient, permettant d'écrire deux fonctions de croyance m_1 et m_2 où le cadre de discernement $\{a, b, c\}$ est composé de trois maladies possibles.

$$\begin{aligned} m_1(a) &= 0.99 & m_1(b) &= 0.01 \\ m_2(c) &= 0.99 & m_2(b) &= 0.01 \end{aligned}$$

Le résultat de la combinaison de ces deux fonctions de croyance en utilisant la règle de Dempster est $m(b) = 1$, ce qui signifie qu'il y a une certitude absolue que le patient souffre de la maladie b . Ceci a été interprété comme étant une contre intuition parce que les deux médecins croient que b est hautement improbable. Les règles de combinaison permettant de traiter ce conflit d'une manière différente sont la règle conjonctive proposée par Smets, la règle de Yager, et la règle proposée par Dubois et Prade. Il est à noter que la critique de Zadeh n'est pas seulement limitée aux fonctions de croyance, mais peut être utilisée aussi bien contre la théorie des probabilités.

La règle de combinaison moyenne

Comme mentionné précédemment dans la critique de Zadeh, lorsque le conflit entre p sources (contradictoires) est présent, la règle de combinaison de Dempster produit des résultats qui ne reflètent pas la distribution réelle des croyances. Une autre règle alternative à la normalisation de Dempster a été proposée. CK. Murphy [16] propose de calculer la moyenne des fonctions de masses des différentes sources :

$$m_{moy}(A) = \frac{1}{p} \sum_{i=1}^p m_i(A_i) \quad (1.22)$$

Si nous reprenons l'exemple des deux médecins qui examinent indépendamment le même patient, la combinaison des deux fonctions de croyance en utilisant la moyenne est présentée dans la table 1.1.

	a	b	c
m_1	0.99	0.01	0
m_2	0	0.01	0.99
Dempster	0	1	0
Moyenne	0.495	0.0001	0.495

TABLE 1.1 – Illustration de la règle de combinaison moyenne

Comme nous pouvons le constater, la combinaison moyenne permet mieux de répartir la croyance contrairement à la règle de Dempster mais ne permet pas de converger vers une hypothèse. Pour assurer la convergence avec la méthode du calcul de la moyenne, l'auteur propose d'insérer les valeurs moyennes des masses dans la règle de combinaison de Dempster. Supposons que nous avons n sources d'information, nous utilisons l'équation 1.15 pour combiner les moyennes pondérées des masses $n - 1$ fois :

$$m_{moyD} = \oplus_{i=1}^{n-1} m_{moy} \quad (1.23)$$

Donnons un autre exemple de combinaison de deux masses m_1 et m_2 présenté dans la table 1.2.

La répartition égale de la croyance fournit par la combinaison de Dempster est contre-intuitive, car les deux hypothèse $\{A\}$ et $\{C\}$ avait des masses individuelles ainsi qu'une partie partagée avec $\{B\}$, par contre $\{B\}$ avait seulement deux masses partagées.

	A	B	C	{A,B}	{A,C}	{B,C}	Ω	\emptyset
m_1	0.5	0	0	0	0	0.5	0	0
m_2	0	0	0.5	0.5	0	0	0	0
Dempster	1/3	1/3	1/3	0	0	0	0	0
Moyenne	0.25	0	0.25	0.25	0	0.25	0	0
Moyenne+Dempster	0.3	0.2	0.3	0.1	0	0.1	0	0

TABLE 1.2 – Exemple de combinaison de Dempster appliquée sur les moyennes des masses

Étant donné que le problème se produit dans l'opération d'intersection, en omettant l'étape de normalisation, les affectations relatives restent inchangées. Cependant, avec la combinaison moyenne des masses, on obtient : $m(A) = m(C) = m(\{A, B\}) = m(\{B, C\}) = 0.25$.

L'opération de combinaison moyenne suivie par la combinaison de Dempster donne : $m(A) = m(C) = 0.3, m(B) = 0.2, m(\{A, B\}) = m(\{B, C\}) = 0.1$.

Ce résultat attribue une masse à $\{A\}$ et $\{C\}$ supérieure par rapport à $\{B\}$ bien qu'il attribue la même plausibilité de 0.4 à $\{A\}$, $\{B\}$ et $\{C\}$.

Cette opérateur de fusion permet de mieux répartir la croyance et converger vers une hypothèse intuitive.

Affaiblissement

Les opérateurs de fusion ne sont pas les seuls permettant la révision des masses de croyance. Il est parfois utile de pondérer l'influence d'une source, avant combinaison, en fonction de la fiabilité qui lui est attribuée. Ceci est réalisée une opération d'affaiblissement. Cette opération permet de déplacer une partie de la croyance de A pour tout $A \subseteq \Omega$ vers un sous-ensemble $B \subseteq \Omega$, tel que $A \subseteq B$. Cela revient à diminuer la crédibilité de chaque sous-ensemble et à augmenter leur plausibilité.

Shafer [1] a proposé un affaiblissement qui consiste à transférer la part de croyance allouée aux éléments focaux vers l'ensemble Ω représentant l'incertitude. Soit $\alpha \in [0, 1]$, un coefficient d'affaiblissement qui représente le degré de confiance accordée à une

source d'information S , la fonction de masse affaiblie est obtenue par :

$$\begin{cases} m_\alpha(A) = \alpha.m(A), & \forall A \subset \Omega \\ m_\alpha(\Omega) = (1 - \alpha) + \alpha.m(\Omega), \end{cases} \quad (1.24)$$

Dans le cas où $\alpha = 0$, la source S n'est pas du tout fiable et la totalité de la masse est transférée sur Ω . On obtient alors une fonction de masse vide (ignorance totale). Au contraire, lorsque la source est parfaitement fiable $\alpha = 1$, les masses restent inchangées.

La fiabilité d'une source d'information n'est généralement pas connue à l'avance. Si nous souhaitons avoir des informations concernant la fiabilité des sources en jeu, il est alors nécessaire de les estimer. Plusieurs travaux ont été réalisés pour estimer les coefficients d'affaiblissement. Dans le cadre de la classification, la matrice de confusion a été utilisée pour estimer ces coefficients [17] [18]. L'utilisation de la matrice de confusion pour les mécanismes d'affaiblissement a été repris dans [19]. Les auteurs ont proposé de calculer le coefficient d'affaiblissement en se basant sur le taux de fiabilité ou taux de classification correcte issu de la matrice de confusion.

La matrice de confusion $M = (n_{kl})_{k \in \{1, \dots, K\} l \in \{1, \dots, K\}}$ (voir table 1.3) associée à une source décrit les performances de cette source sur un ensemble de test composé de n individus.

Décision	Vérité	w_1	w_K
w_1		n_{11}	...	n_{1l}	...	n_{1K}
\vdots		\vdots	...	n_{kk}	...	\vdots
w_K		n_{K1}	...	n_{Kl}	...	n_{KK}

TABLE 1.3 – Illustration d'une matrice de confusion.

Les colonnes l correspondent au cas où la vérité est w_l . Les lignes k correspondent à une décision en faveur de w_k . Le nombre d'éléments de classe w_l ayant été classés dans la classe w_k par la source est représenté par n_{kl} .

Le taux de classifications correctes qui correspond au taux de fiabilité T_f d'une matrice est alors définie par :

$$T_f = \frac{\sum_{k=1}^K n_{kk}}{n} \quad (1.25)$$

Il suffit maintenant d'utiliser le taux de fiabilité T_f d'une source d'information pour réaliser un affaiblissement de l'information m fournie par cette source en utilisant l'équation 1.24 avec $\alpha = T_f$.

Donnons un exemple d'une matrice de confusion issue d'une classification s'appuyant sur une source d'information S chargée de reconnaître 4 types d'objets a, b, c et d (voir table 1.4). Dans le tableau présenté dans cet exemple nous comptons le nombre

	Vérité	a	b	c	d
Décision					
a		8	1		
b			6		
c		2	6	1	
d		1	4	9	

TABLE 1.4 – Exemple d'une matrice de confusion.

d'éléments $n = 38$, ce qui nous donne le taux de fiabilité $T_f = \frac{8+6+6+9}{38} = 0.7631$.

Supposons maintenant que la source fournisse une information m sur des hypothèses tel que : $m(\{a, c\}) = 0.8$ et $m(\Omega) = 0.2$. Par affaiblissement on obtient :

$$m_\alpha = m \cup \Omega_{0.76},$$

$$m_\alpha(\{a, c\}) = 0.61, \quad m_\alpha(\Omega) = 0.39.$$

Une autre méthode a été proposée dans [20], basée sur la minimisation de la distance entre les probabilités pignistiques (calculées à partir des fonctions de masse affaiblies) et les valeurs réelles (connues) des classes.

Une mesure de conflit basée sur la distance entre les fonctions de croyance a été utilisée dans [21] pour une estimation a posteriori de la fiabilité relative entre les sources d'information. Cette estimation de la fiabilité ne nécessite aucun apprentissage ou une connaissance préalable et peut ensuite être utilisée pour affaiblir les sources non fiables avant l'étape de combinaison.

1.3.3 Obtention d'une fonction de croyance

Nous introduisons dans cette section les méthodes usuelles permettant de modéliser les croyances. Déjà difficile avec la théorie des probabilités, la modélisation de l'information pose aussi problème dans le cadre des fonctions de croyances. À l'exception d'un expert qui donne son opinion d'une manière directe sous la forme d'une fonction de croyance, il n'existe pas de méthode générique pour la modélisation des fonctions de croyance. Cela dépend généralement de l'application concernée. Des méthodes de modélisations ont été proposées dans la littérature, même si ces méthodes peuvent parfois être très spécifiques, on peut distinguer deux grandes familles :

- les masses de croyances estimées à partir de vraisemblances.

Pour ce modèle, nous distinguons deux approches, l'approche globale et l'approche séparable. L'approche globale, développée par Shafer [1] est ainsi nommée car la construction de la fonction de croyance nécessite l'ensemble des hypothèses. La seconde approche, dite séparable, construit des fonctions de masse pour chaque élément du cadre de discernement. Elle a été initialement proposée par Smets [22]. Elle fut ensuite reprise par Appriou [23] qui propose deux modèles utilisant les probabilités conditionnelles $p(x_j|\omega_k)$ où x_j est une observation selon la source j permettant d'affecter à une source une masse relative à une hypothèse donnée : *Modèle 1* :

$$\begin{cases} m_{jk}(\{\omega_k\}) = \alpha_{jk}K, \\ m_{jk}(\overline{\{\omega_k\}}) = \alpha_{jk}(1 - K), \\ m_{jk}(\Omega) = 1 - \alpha_{jk} \end{cases} \quad (1.26)$$

Modèle 2 :

$$\begin{cases} m_{jk}(\{\omega_k\}) = 0, \\ m_{jk}(\overline{\{\omega_k\}}) = \alpha_{jk}(1 - K), \\ m_{jk}(\Omega) = 1 - \alpha_{jk}(1 - K) \end{cases} \quad (1.27)$$

où $K = R_j p(x_j|\omega_k)$ représente la vraisemblance de ω_k et de toutes les décisions qui lui sont proches compte tenu de la source S_j , avec R_j un coefficient de pondération des probabilités lié à la source S_j . α_{jk} est le coefficient d'affaiblissement de la source S_j pour la classe ω_k . La difficulté de cette méthode de modélisation réside dans l'estimation de $p(x_j|\omega_k)$.

- les masses de croyances estimées à partir de distances.

Cette démarche a été proposée par Denœux [24], et améliorée par la suite par Denœux et Zouhal [25][26], puis étendue dans le cadre de la classification multi-labels [27].

Dans cette approche, les fonctions de masse sont construites à partir des distances calculées dans l'espace des caractéristiques (vecteurs d'apprentissage). Deux versions ont été proposées : la première basée sur l'algorithme de k -plus proches voisins, la deuxième est basée sur l'algorithme k -means.

Masses de croyance estimées selon les k -plus proches voisins

Soit un ensemble d'apprentissage $\mathcal{L} = \{x_i, \omega_j^i\}$ où x_i représente le vecteur de caractéristiques associé à une hypothèse ω_j^i . Supposons que nous avons une nouvelle observation x pour laquelle nous voulons obtenir des informations sur sa classe d'appartenance. L'idée est de considérer chaque élément de l'ensemble de ses k -plus proches voisins comme une source d'information pour ce vecteur x à classer. Les fonctions de masses sont alors obtenues en fonction de la distance entre x_i et x , et sont définies par :

$$\begin{cases} m_i(\omega_j^i) = \alpha\phi(d(x, x_i)) \\ m_i(\Omega) = 1 - \alpha\phi(d(x, x_i)) \end{cases} \quad (1.28)$$

où ϕ est une fonction décroissante de \mathfrak{R}^+ vers $[0, 1]$ tel que $\phi(0) = 1$ et $\lim_{d \rightarrow \infty} \phi(d) = 0$. α est le coefficient d'affaiblissement, d est une distance à déterminer en fonction des données. Pour la fonction ϕ , parmi le grand nombre de fonctions, Denœux a proposé la relation suivante, notamment dans le cas des distances euclidiennes :

$$\phi(d) = \gamma \exp(-d^2) \quad (1.29)$$

où $\gamma \in [0, 1]$ est un paramètre positif. La masse de croyance m associé à x est ensuite obtenue par la combinaison des masse m_i .

Une étude comparative entre les approches vraisemblance et distance a été réalisée [28]. D'après cette étude, il n'y a pas de méthode générique d'obtention de masses de croyance. Généralement, la modélisation des croyances dépend fortement de l'application.

Pour notre part, nous avons utilisé dans les travaux de ce manuscrit l'approche basée sur les k -plus proches voisins pour sa simplicité de mise en oeuvre et la qualité de ses résultats.

1.3.4 Niveau pignistique : Prise de décision

À ce stade, nous venons de présenter le niveau *crédal*, qui permet de modéliser et réviser les connaissances sous forme de fonction de masse m . Pour prendre la décision et choisir l'hypothèse qui minimise le risque de se tromper, Smets et Kennes [12] proposent de quitter le niveau *crédal* pour aller au niveau de prise de décision dite *pignistique*.

Rappelons que l'avantage de la théorie des fonctions de croyance est de considérer les singletons et les disjonctions. Néanmoins, dans un système de classification automatique la décision sur les hypothèses multiples est peu utilisée car l'objectif est souvent d'avoir une réponse sans imprécision, et par conséquent porter son choix sur un singleton. Pour cette raison Smets propose de se ramener à un cadre probabiliste au niveau de la décision afin de décider en faveur des singletons, en transformant la fonction m en une fonction de probabilité, appelée probabilité *pignistique*, notée $BetP$. La probabilité pignistique redistribue uniformément les masses vers les hypothèses ω_k de Ω selon l'expression suivante :

$$BetP(\omega_k) = \frac{1}{1 - m(\emptyset)} \sum_{\omega_k \in A} \frac{m(A)}{|A|} \quad (1.30)$$

où $|A|$ représente la cardinalité de $A \subseteq \Omega$. Avec la probabilité pignistique ainsi obtenue, nous nous trouvons dans le cadre classique de la décision Bayésienne qui préconise l'action pour laquelle l'espérance du coût est la plus faible. Ainsi, l'espérance pignistique est définie par :

$$f : \Omega \longrightarrow \mathbb{R} \\ E_{BetP}(f) = \sum_{\omega_k \in \Omega} f(\omega_k) BetP(\omega_k) \quad (1.31)$$

Si on considère le problème du choix d'une action parmi l'ensemble des N actions possibles $A = \{a_1, \dots, a_N\}$, l'action a_k correspond à l'action du choix de ω_k . Sélectionner une action a_i , alors que la réalité est l'hypothèse ω_k , génère un coût noté $\lambda(a_i|\omega_k)$, alors le risque conditionnel de décider a_i est défini par :

$$R_{BetP}(a_i|\omega_k) = \sum_{\omega_k \in \Omega} \lambda(a_i|\omega_k) Betp(\omega_k) \quad (1.32)$$

La décision dans le cas de coût $\{0, 1\}$ s'effectue alors par maximisation de la probabilité pignistique :

$$BetP(\omega_i) = \max_{\omega_k \in \Omega} BetP(\omega_k) \quad (1.33)$$

D'autres stratégies de prise de décision ont été proposées, le choix d'une règle de décision dépend donc du problème considéré. Parmi les solutions, Appriou [29] propose de se baser sur le maximum de plausibilité. Cette stratégie attribue l'hypothèse de plus haute plausibilité w_i à l'observation considérée :

$$pl(w_i) = \max_{\omega_k \in \Omega} pl(\omega_k) \quad (1.34)$$

Sur le même principe, la décision peut être prise sur l'hypothèse de plus haute crédibilité :

$$bel(w_i) = \max_{\omega_k \in \Omega} bel(\omega_k) \quad (1.35)$$

1.4 Conclusion

Dans ce chapitre nous avons introduit brièvement les méthodes d'apprentissage automatique à partir d'un ensemble de données. Dans cette thèse, nous considérons le cas spécifique de l'apprentissage supervisé, dans lequel le modèle est présenté avec les données d'entrée et une sortie - issue d'un phénomène spécifique - à partir duquel il faut construire une relation entre l'entrée et la sortie, afin de prédire le comportement futur du phénomène sous-jacent. Nous avons présenté les différentes étapes pour la construction d'un modèle pour un problème d'apprentissage, dont l'objectif est d'obtenir un ensemble d'hyper-paramètres optimale.

Nous avons ensuite présenté la théorie des fonctions de croyance qui un cadre mathématique particulier d'apprentissage automatique. Ce formalisme, adapté aux informations imparfaites (imprécises et incertaines), permet de représenter de façon relativement naturelle les connaissances. Ensuite le modèle de croyance transférable a été présenté à travers ses 2 niveaux.

Tout d'abord, le niveau *crédal* composé de deux parties : la partie statique qui consiste à modéliser les croyance et la partie dynamique qui correspond au transfert de croyance lors de l'arrivée d'une nouvelle information, avec les opérations de combinaison et d'affaiblissement.

Enfin, le deuxième niveau dit *pignistique* permettant la prise de décision a été illustré. Dans ce niveau un large choix de stratégies de décision est offert, parmi eux Smets recommande de convertir les fonctions de masse en fonctions de probabilité, appelées probabilité *pignistique*.

Nous proposons maintenant d'utiliser ces différents concepts dans deux problématiques : l'évaluation de la qualité d'image et la détection d'informations cachées dans une image (stéganalyse).

Chapitre 2

Évaluation de la qualité d'image basée sur la théorie des fonctions de croyance

Sommaire

2.1	Introduction	38
2.2	Évaluation de la qualité d'image	39
2.2.1	Évaluation subjective de la qualité d'image	40
2.2.2	Évaluation objective de la qualité d'image	44
2.2.3	Apprentissage automatique pour l'évaluation de la qualité d'image	48
2.3	Mesure sans référence d'évaluation de la qualité d'image basée sur la théorie des fonctions de croyance	51
2.3.1	Représentation floue pour la mesure de qualité	51
2.3.2	Algorithme <i>EVreg</i> pour l'évaluation sans référence de la qualité d'image	52
2.3.3	Définition des descripteurs	55
2.3.4	Un algorithme de régression évidentiel multi-sources	56
2.4	Évaluation des algorithmes <i>EVreg</i> et <i>ms-EVreg</i>	58
2.4.1	Protocole expérimental	58
2.4.2	Résultats de l' <i>EVreg</i> avec des caractéristiques prises séparément	59
2.4.3	Évaluation de <i>ms-EVreg</i>	62

2.5	Les coefficients d'affaiblissement optimisés pour <i>ms-EVreg</i>	64
2.5.1	Le principe	64
2.5.2	Estimation des coefficients d'affaiblissement	65
2.5.3	Évaluation de <i>wms-EVreg</i>	66
2.5.4	Comparaison avec d'autres mesures d'évaluation	69
2.5.5	La robustesse et l'indépendance de la base	74
2.5.6	Évaluation de <i>wms-EVreg2</i> sur toute la base TID2013	75
2.6	Conclusion	77

2.1 Introduction

Au cours d'une chaîne de vie, les images numériques subissent divers traitements, qui permettent par exemple la transmission et la diffusion de l'image. Cependant, ces traitements entraînent des dégradations qui influent sur la qualité visuelle des images. Il est nécessaire alors de disposer d'outils efficaces capables de mesurer l'impact de ces distorsions sur la qualité de l'image. Au fil des années, plusieurs études ont été menées dans ce sens, afin de trouver des solutions à ce problème [30].

La qualité d'image est une notion liée à la perception visuelle humaine. En effet, notre capacité à évaluer la qualité d'une image sans avoir recours à sa version originale vient du fait que notre cerveau a la capacité de stocker et d'apprendre une énorme quantité d'informations. Naturellement, des observateurs sont capables de juger la qualité d'image par rapport à une échelle de référence liée à ses sens [30]. L'étude de ces mécanismes - pour comprendre quelles sont les échelles internes utilisées par un être humain pour évaluer cette qualité - est devenue un domaine de recherche à part entière. En traitement d'images, il semble alors évident d'associer la mesure de la qualité d'une image à la perception de l'être humain. L'étude bibliographique des travaux portant sur la qualité des images distinguent deux approches d'évaluation de la qualité des images : les évaluations subjectives et les évaluations objectives.

Dans l'évaluation subjective de la qualité, les observateurs humains sont appelés à évaluer la qualité d'une image dans des conditions environnementales très spécifiques et identiques pour chaque observateur et suivant un protocole bien défini [31]. Cette évaluation est considérée comme un outil important pour comparer les différentes mesures proposées. Cette procédure est à la fois coûteuse en temps, en moyen et elle est difficile à mettre en oeuvre. Par conséquent des algorithmes objectifs d'évaluation

de la qualité de l'image ont été activement étudiés pour permettre d'évaluer de manière automatique la qualité d'une image en corrélation avec l'appréciation visuelle humaine. Trois principaux types de mesures objectives peuvent être distingués : (1) mesures avec référence (*Full-Reference image quality assessment*) (2) mesures avec référence réduite (*Reduced-Reference image quality assessment*) (3) mesures sans référence (*No-Reference image quality assessment*).

Dans ce chapitre, nous proposons un algorithme d'évaluation sans référence de la qualité d'image basée sur l'extension floue de la théorie des fonctions de croyance. Les fonctions de croyance permettent de représenter les connaissances d'une manière relativement naturelle sous la forme d'une structure de croyances. De plus, ce formalisme offre de nombreux outils pour prendre en compte la fiabilité de la source d'information et permet de fusionner différentes sources d'information. L'originalité de notre méthode est de prendre en considération et d'estimer la fiabilité des caractéristiques extraites de l'image par un processus d'affaiblissement, et ceci dans le cadre de l'évaluation de la qualité.

Dans la suite du chapitre, nous présentons en section 2.2 quelques notions et définitions classiques ainsi qu'un bref état de l'art de l'évaluation de la qualité de l'image. Dans la section 2.3, une description détaillée de notre proposition d'évaluation automatique de la qualité de l'image à l'aide de fonctions de croyance floues est présentée. Les résultats obtenus sont décrits et discutés dans la section 2.4. Des améliorations des performances de la mesure de qualité initiale proposée sont présentées dans la section 2.5.

2.2 Évaluation de la qualité d'image

Plusieurs études ont été effectuées afin de trouver des solutions permettant de mesurer l'impact de distorsions sur la qualité d'image. Avant de détailler les solutions proposées dans état de l'art, nous allons donner quelques définitions importantes.

Comme nous l'avons vu, il existe essentiellement deux types d'évaluation de la qualité d'image : évaluation subjective et évaluation objective. Dans ce qui suit, nous allons décrire les deux approches respectivement dans les sections 2.2.1 et 2.2.2. Ensuite, dans la section 2.2.3, nous allons nous concentrer sur les mesures d'évaluation objectives basées sur les méthodes d'apprentissage automatique.

2.2.1 Évaluation subjective de la qualité d'image

L'évaluation subjective fait appel à des observateurs humains pour évaluer la qualité d'une image selon un protocole bien défini [31]. Étant donné que les êtres humains sont très majoritairement les observateurs finaux des images et qu'ils prennent des décisions en fonction de ce qu'ils observent sur une image, ils restent encore actuellement les meilleurs juges de la qualité des images. Ainsi, les tests psychovisuels demeurent un outil prépondérant pour comparer les différentes métriques proposées. Il s'agit de demander à un observateur humain de juger la qualité des images qui lui sont présentées selon une grille d'évaluation à plusieurs niveaux d'appréciation. A l'issue de ces tests, une note subjective est obtenue pour chaque image. Ces valeurs subjectives deviendront ainsi les vérités que toute mesure de qualité algorithmique devra être en mesure de reproduire le plus rigoureusement possible. La réalisation de ces tests nécessite d'avoir une attention sur les conditions et facteurs pouvant altérer le jugement des observateurs, et diminuent la fiabilité de ces tests. Parmi les facteurs influant nous pouvons citer les plus pertinents :

1. La distance d'observation,
2. Les conditions de visualisation tel que l'éclairage et l'environnement,
3. La technologie et le calibrage de l'écran utilisé,
4. Le choix des images,
5. Les facteurs psychologiques de l'observateur.

Différentes méthodes d'évaluation subjective ont été définies par l'ITU [32]. Nous distinguons principalement trois grandes familles : les tests à simple stimulus permettant de juger la qualité d'image sans aucune information relative à l'image originale. Les tests à double stimuli où il est demandé à l'observateur de juger la similarité entre deux images et les tests comparatifs dont l'objectif principal est de comparer deux ou plusieurs stimuli.

A la fin de ces tests une note unique décrivant la qualité moyenne est attribuée à chaque image, cette note est appelée MOS (Mean Opinion Score) est donnée par la formule suivante :

$$MOS(i) = \frac{1}{P} \sum_{j=1}^P Note_i(j) \quad (2.1)$$

où P représente le nombre total des participants et $Note_i(j)$ la note donnée à l'image i par l'observateur j .

D'autres mesures peuvent être calculées après l'évaluation subjective, il s'agit des valeurs différentielles des scores d'opinion (*DMOS* : Differential Mean Opinion Score) obtenus après la conversion des scores bruts en scores de différence entre les images dégradées et leurs versions originales correspondantes :

$$DMOS(i) = \frac{1}{P} \sum_{j=1}^P (Note_i(j_0) - Note_i(j_k)) \quad (2.2)$$

où $Note_j(j_0)$ représente la note attribuée par l'observateur j à l'image originale i n'ayant pas subi de dégradation ($k = 0$).

Les bases d'images utilisées pour l'évaluation

Plusieurs bases d'images associées avec leurs tests subjectives ont été proposées depuis quelques années. Nous allons présenter et décrire les bases les plus utilisées, les types d'images qu'elles contiennent et les différentes dégradations appliquées.

IVC

Dans la base IVC (*Image and Video-Communication*) [33] 10 images originales ont été utilisées, 235 images ont été générées à partir de 4 différents processus de dégradations : JPEG, JPEG2000, LAR coding et le flou. L'évaluation subjective a été faite à une distance égale à 6 fois la hauteur de l'écran à l'aide d'un protocole à double stimuli, avec une échelle de 5 niveaux d'appréciation et 15 observateurs.



FIGURE 2.1 – Échantillon d'images de la base IVC.

LIVE

La base LIVE (Laboratory for Image & Video Engineering) [34] contient 29 images de référence. 5 types de distorsions ont été appliquées pour produire un total de 982 images dont 203 images qui ne présentent aucune dégradation. Ces distorsions sont : JPEG2000, JPEG, bruit blanc dans les composantes RGB, le flou gaussien dans les composantes RGB et les erreurs de bits dans le canal de transmission (JPEG2000 bitstream). Une vingtaine d'observateurs sont sollicités pour juger ces images sur une échelle de 5 niveaux. Les notes obtenues de chaque observateur sont converties en scores de différence (*DMOS* : Difference Mean Opinion Score).



FIGURE 2.2 – Échantillon d'images de la base LIVE.

Toyama

Deux dégradations sont considérées dans cette base : JPEG et JPEG2000 [35], le nombre d'images généré pour chaque dégradation est 98. Ces images ont été jugées par un panel de 16 observateurs qui donnent une note comprise entre 1 et 5 (respectivement la plus mauvaise et la meilleure qualité). La figure 2.3 présente un échantillon d'images de référence.

CSIQ

La base d'image CSIQ (Categorical Subjective Image Quality) [36] est composé de 30 images de référence obtenues à partir d'une source de domaine public. Les images ont été choisies pour couvrir cinq catégories : animaux, paysages, personnes, plantes et urbain. Les distorsions utilisées dans CSIQ sont : compression JPEG, compression



FIGURE 2.3 – Échantillon d'images de la base Toyama.

JPEG2000, décalage de contraste global, bruit Gaussien rose, bruit Gaussien blanc et le flou Gaussien. Au total, 866 images dégradées ont été générées.



FIGURE 2.4 – Échantillon d'images de la base CSIQ.

TID 2013

La base TID2013 contient 25 images de référence et 3000 images dégradées (25 images de référence x 24 types de distorsions x 5 niveaux de distorsions) [37]. La valeur MOS a été obtenue à partir des résultats de 971 expériences effectuées par des observateurs provenant de cinq pays différents : la Finlande, la France, l'Italie, l'Ukraine et les États-Unis. La valeur MOS obtenue est comprise entre 0 et 9, correspondant, respectivement, à la qualité la plus mauvaise et la meilleure qualité. Un échantillon d'images de référence est présenté par la figure 2.5

Tous ces tests nécessitent le déploiement des outils informatiques et un matériel spécifiques (salle dédiée, éclairage, écran, environnement, observateurs) ce qui rend



FIGURE 2.5 – Échantillon d'images de la base TID2013.

l'évaluation subjective complexe, lourde et difficilement reproductible. Pour ces raisons des mesures objectives permettant d'évaluer automatiquement la qualité des images sont élaborées. Il s'agit des solutions algorithmiques qui permettent d'estimer la qualité de l'image en corrélation avec les données subjectives.

Ces bases d'images et leurs scores de qualité associées sont les références que visent à atteindre les algorithmes d'évaluation objective de qualité.

2.2.2 Évaluation objective de la qualité d'image

L'évaluation objective de la qualité d'image est basée sur l'analyse et la mesure quantitative du niveau de dégradation au moyen de métriques directement liées au signal physique. Les métriques de qualité objectives sont évaluées à travers l'étude de la cohérence avec les notes subjectives. Plusieurs travaux ont été menés dans ce sens, et plusieurs mesures d'évaluation ont été proposées. Les mesures proposées sont classées en 3 catégories : mesures avec référence, mesures sans référence et mesures avec référence réduite.

Mesures avec référence

Ce type de mesure nécessite un accès complet au signal de l'image originale, il s'agit donc d'une mesure de similarité entre l'image originale et sa version dégradée. Ce type de mesures est très largement utilisé pour quantifier numériquement la qualité de la reconstruction d'une image. Parmi le panel de mesures exploitées, l'une des mesures couramment utilisée est l'Erreur Quadratique Moyenne (EQM). Elle représente

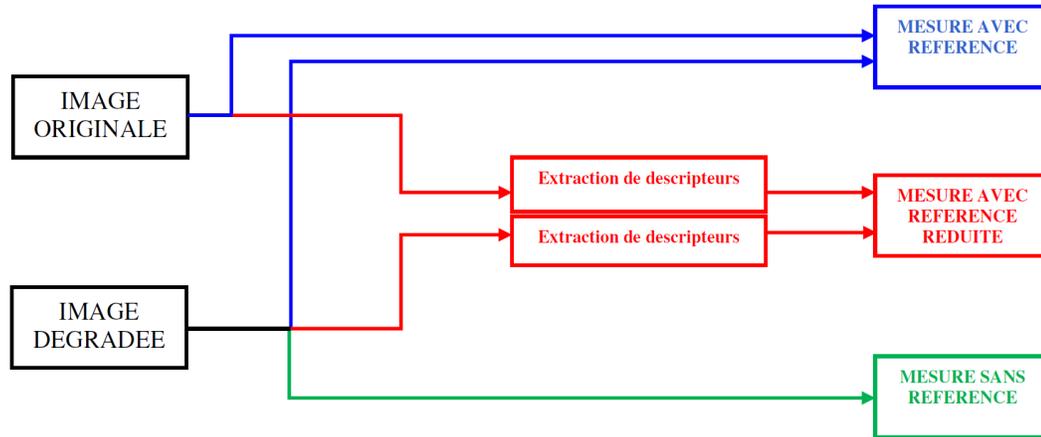


FIGURE 2.6 – Différentes approches des mesures objectives

la moyenne quadratique du signal erreur ou distorsion. Elle est donnée par :

$$EQM = \frac{1}{N.M} \sum_{i=1}^M \sum_{j=1}^N (I(i, j) - I_d(i, j))^2 \quad (2.3)$$

avec $I(i, j)$ et $I_d(i, j)$ le pixel de coordonnées (i, j) de l'image originale et de sa version dégradée, respectivement. N et M représentent respectivement le nombre de lignes et de colonnes de l'image.

Plusieurs variantes basées sur l'EQM ont été proposées. Parmi ces méthodes nous pouvons citer le rapport signal sur bruit (SNR : Signal Noise Ratio) et le pic du signal sur bruit (PSNR : Peak Signal Noise Ratio) qui restent les plus utilisées [38] :

$$SNR = 10 \log \left(\frac{\sum_{i=1}^M \sum_{j=1}^N (I(i, j))^2}{EQM} \right) \quad (2.4)$$

$$PSNR = 10 \log \left(\frac{\text{Valeur pic du signal original}}{EQM} \right) \quad (2.5)$$

où la valeur pic du signal original est fixée à 255 pour les images codées sur 8 bits par pixels.

Bien que ces métriques soient très fréquemment utilisées en traitement d'image pour quantifier la qualité, elles ne permettent pas de prendre en compte la sensibilité du système visuel humain (SVH), et présentent une faible corrélation avec l'appréciation subjective.

Plusieurs tentatives afin d'améliorer la performance des mesures basées sur EQM ont été conduites, notamment dans [39] et [40], les auteurs ont proposé d'intégrer au calcul de l'EQM une fonction linéaire et un filtrage par une fonction de sensibilité au contraste. D'autres auteurs ont proposé des modèles appliqués dans différent espaces de transformées tels que l'espace temps-fréquence de Wigner-Ville [41], le domaine des ondelettes [42] et le domaine de la transformée en cosinus discrète (TCD). Comparativement avec les approches classiques telles que le PSNR, ces méthodes donnent des résultats satisfaisant, mais elles restent toujours peu corrélées avec l'appréciation subjective.

Plusieurs recherches ont été orientées vers les métriques de qualité basées sur une analyse des structures locales du signal image pour faire face à la faiblesse des métriques basées EQM. Parmi les métriques les plus populaires et les plus utilisées, nous trouvons celle appelé UQI (Universal Quality Index) proposée dans [43] et SSIM (Structural Similarity Index) [44]. Ces mesures sont basées sur l'extraction d'attributs structurels locaux de l'image à partir desquels chaque bloc de l'image est décrit par sa luminance (l), son contraste (c) et sa structure (s). La mesure SSIM est une version améliorée de UQI. Une version multi-échelle de SSIM a été proposée dans [45] appelée MS-SSIM qui fournit plus de flexibilité en intégrant les variations de conditions de visualisation. Sheikh et Bovik proposent une mesure appelée VIF : Visual Information Fidelity [46]. Il s'agit d'une mesure de fidélité entre des modèles statistiques de scènes naturelles (NSS) générés à partir des images originales et leurs versions dégradées.

Toutes les mesures avec référence citées auparavant exigent un accès complet aux images de référence, ce qui amène de fortes restrictions quant à leur usage. De ce fait, le groupe des experts de la qualité des vidéos (VQEG : Video Quality Expert Group) [47] a intégré les mesures avec référence réduite comme l'un de ses axes des futures recherches.

Mesures avec référence réduite

Les mesures objectives de qualité avec référence réduite [48] diffèrent des mesures avec référence complète par le fait que l'image originale n'est pas disponible et/ou utilisée. La référence est alors remplacée par une représentation formée de descripteurs pertinents qui peuvent être comparés dans les deux versions de l'image originale et dégradée. Ceci est très utile lorsque les informations de référence doivent être transmises avec une faible bande passante. Liu et al. [49] proposent une mesure avec référé-

rence réduite qui mesure la différence de régularité de la congruence de phase (*Phase Congruency*) entre l'image de référence et l'image dégradée. Plusieurs autres mesures oeuvrant dans le domaine des ondelettes ont été proposées. Dans [50], les coefficients d'ondelettes sont modélisés par une densité gaussienne généralisée par image.

Dans le laboratoire, nous avons proposé une métrique à référence réduite basée sur la modélisation des coefficients issus de la transformée en ondelettes quaternioniques [51]. Pour obtenir les références réduites, nous modélisons la distribution des coefficients suivant différentes lois dont les paramètres constituent les références. La transformée quaternionique est une amélioration de la transformée en ondelettes classiques et fournit une représentation plus riche qui est invariante à la translation et avec une information de phase. La dégradation est alors caractérisée par la différence entre les paramètres des distributions de l'image d'origine et les paramètres de l'image dégradée. Les résultats obtenus ont montré une grande efficacité de la méthode.

Cependant, comme pour toute métrique à référence réduite, cela nécessite la transmission de ces références, c'est pourquoi nous nous tournons dans ce travail de thèse vers les métriques sans référence, basées sur l'idée que l'être humain est capable de juger la qualité d'une image sans avoir recours à la comparaison avec une image de référence.

Mesures sans référence

Les mesures sans référence sont les plus attrayantes car elles permettent d'estimer la qualité uniquement à partir de l'image dégradée. Plusieurs mesures sans référence orientées pour une dégradation spécifique ont été proposées. Nous pouvons citer par exemple le modèle des statistiques des scènes naturelles utilisé dans [52] pour définir une mesure de qualité pour les images compressées par JPEG2000. Dans [53] un algorithme ciblant aussi les images compressées par JPEG2000 est exploité, il est basé sur les distorsions des pixels et les informations de contour. Une mesure pour les images JPEG est considérée dans [54] pour les attaques de type flou et effets de blocs qui représentent les artefacts les plus rencontrés lors d'une compression JPEG.

Enfin, plusieurs chercheurs ont exploré la possibilité d'utiliser l'apprentissage automatique pour prédire le score de qualité d'une image, basé sur l'extraction des caractéristiques pertinentes des images, et les scores subjectifs attribués par les observateurs humains. Les travaux de recherche présentés dans ce chapitre se focalisent sur ces mesures d'évaluation de la qualité d'image basées sur l'apprentissage automatique.

2.2.3 Apprentissage automatique pour l'évaluation de la qualité d'image

Dans le contexte des mesures avec référence, Bouzerdoum et al. [55] proposent une mesure basée sur les réseaux de neurones. Le vecteur des caractéristiques utilisé comprend la moyenne et l'écart-type de l'image de référence et sa version dégradée, la covariance et l'EQM entre l'image originale et l'image dégradée. L'approche est testée sur une base contenant 352 images compressées JPEG et JPEG2000. Plus tard, Narwaria and Lin [56] proposent une mesure basée sur la régression des machines à vecteur de support (SVR : *Support Vector Regression*) avec un noyau type RBF (*Radial Basis Function*). Les caractéristiques utilisées sont extraites après l'application de la décomposition en valeurs singulières de la matrice d'image.

Dans le cas des mesures avec références réduites, Carrai et al. [57] proposent de nouveau un algorithme basé sur les réseaux de neurones. Les caractéristiques sont extraites de chaque bloc de l'image originale et de sa version dégradée, à partir de l'histogramme (moyenne, écart-type, kurtosis, ...) et de la matrice de co-occurrence (entropie, énergie,...). Le même principe a été suivi par Dongxue et al. [58], mais en changeant le domaine d'extraction des descripteurs, en décomposant les images au moyen de la transformé en ondelettes.

En ce qui concerne les mesures sans référence, Gastaldo et al. [59] proposent de prédire le score de la qualité des images compressées JPEG en utilisant les réseaux de neurones de type perceptron multicouche. Les caractéristiques utilisées décrivent le contenu de l'image en termes de distribution de luminance, orientation spatiale et distribution des fréquences. Une autre mesure spécifique aux images compressées JPEG a été proposée dans [60], avec un réseau de neurone à base radiale (RBF : *Radial Basis Function*). Les caractéristiques utilisées extraites de chaque bloc (8×8) de l'image représentent l'amplitude et la longueur des contours dans chaque bloc. Dans [61], un perceptron multi-couches a été utilisé pour combiner trois mesures sans référence chacune spécifique à une distorsion dans un seul dispositif, afin d'évaluer la distorsion de type floue, effet de bloc et le ringing.

Pour toutes les méthodes d'évaluation sans référence présentées ci-dessus, les algorithmes développés visent à évaluer un seul type de distorsion (par exemple JPEG, JPEG2000, flou). Par contre, souvent, plus d'un type de distorsion peuvent être présent dans une image. Ceci induit un choix délicat et arbitraire de la métrique nécessaire

pour évaluer ces distorsions.

Pour ces raisons, des nouvelles approches capables d'évaluer de multiples types de distorsions ont été proposées. Li et al. [62] ont développé une mesure basée sur la régression générale par les réseaux de neurone (GRNN : *General Regression Neural Network*), où ils ont utilisé un certain nombre de caractéristiques de l'image qui possèdent des informations pertinentes : phase de la congruence, l'entropie et le gradient.

M. Saad et al. [63] proposent d'extraire un ensemble de caractéristiques à partir d'un modèle statistique de scènes naturelles (NSS) des coefficients la transformée en cosinus discrète (DCT). L'approche proposée appelée *BLINDS-II*, repose sur le modèle simple d'inférence Bayésienne pour prédire les scores de qualité d'image en fonction des caractéristiques extraites.

Une autre mesure sans référence d'évaluation de la qualité d'image appelée BRISQUE a été présentée dans [64] basée sur la statistique de scène naturelle dans le domaine spatial. Cette mesure utilise les coefficients de luminance localement normalisés pour quantifier les pertes possibles de "naturalité" dans l'image en raison de la présence de distorsions. Ensuite, l'indice de qualité est estimé par le biais de la régression des vecteurs de support (SVR) en utilisant le package LIBSVM [65] avec la fonction à base radial (RBF) comme noyau.

Mittal et al. [66] proposent ce qu'ils appellent mesure de qualité des images naturelles (NIQE : Natural Image Quality Evaluator). Cet algorithme repose sur la construction d'une collection de caractéristiques similaires à ceux utilisées dans BRISQUE, mais qui est extraite d'un corpus d'images naturelles et non dégradées. La qualité de l'image dégradée est exprimée par une simple métrique de distance entre les statistiques du modèle et celles de l'image dégradée.

Récemment une étude a été réalisée sur l'efficacité de la transformée en curvelet comme descripteur de l'image pour une mesure d'évaluation de la qualité [67]. Pour cela, un ensemble de caractéristiques statistiques est extrait à partir des coefficients de la transformée en curvelet de l'image. Afin d'évaluer la qualité, deux niveaux de traitement sont utilisés. Le premier niveau est une classification : l'ensemble des caractéristiques extraites de l'image est utilisé pour identifier le type de distorsion de l'image dégradée. Une étape de régression interviendra ensuite où les mêmes caractéristiques sont utilisées pour évaluer la qualité de la distorsion spécifique de l'image classée. Les auteurs ont utilisé les machines à vecteurs de supports (SVM) du package LIBSVM, à la fois pour la classification et la régression.

L'étude de l'état de l'art montre que les mesures objectives de l'évaluation de la qualité basées sur l'apprentissage automatique sont très efficaces. Les techniques de régression, généralement SVM et réseaux de neurones, sont construits en utilisant les scores obtenus à partir des tests subjectifs sur des images numériques, et de plusieurs caractéristiques extraites de ces images afin de produire des scores objectifs de la qualité.

Toutefois, même si les résultats sont bien corrélés aux scores subjectifs, nous considérons un inconvénient majeur, à savoir que ces mesures ne prennent pas en compte l'aspect d'imperfection des données manipulées. En effet, les caractéristiques utilisées proviennent de différentes méthodes d'extraction proposées dans la littérature. Cette information est donc fournie par différentes sources d'information avec différents degrés de précision ou de fiabilité : chaque vecteur de caractéristiques est plus ou moins adapté à une catégorie d'images ou de distorsion. Afin de profiter au maximum de l'ensemble des informations d'entrée, nous avons besoin des formalismes capables de gérer ces données multi-source et fournir certains mécanismes de fusion (combinaison ou agrégation), pour obtenir des informations résumant l'ensemble initial de données, ce qui est essentiel pour améliorer la prise de décision et prédire le score de qualité.

Pour faire face à ces limitations, nous proposons un algorithme d'évaluation sans référence de la qualité d'image s'appuyant sur le cadre théorique des fonctions de croyance. Comme nous l'avons vu, cette théorie possède plusieurs avantages utiles pour notre contexte : (1) Elle est construite pour gérer les connaissances imparfaites, (2) Elle fournit certains outils qui aident à prendre en compte la confiance (fiabilité) dans les données d'apprentissage, (3) Elle offre des outils de combinaison permettant de fusionner plusieurs sources d'information afin d'améliorer la connaissance du monde observé.

Dans la section suivante, nous proposons d'utiliser le principe du k-plus proche voisins évidentiel (EV-KNN : Evidential K-nearest neighbors) dans le contexte de la prédiction et nous proposons un algorithme d'évaluation de qualité sans référence basé sur la théorie des fonctions de croyance.

2.3 Mesure sans référence d'évaluation de la qualité d'image basée sur la théorie des fonctions de croyance

Afin d'évaluer la qualité d'une image, nous avons besoin d'établir une correspondance (*mapping*) à partir de l'espace des caractéristiques vers la prédiction de la qualité. Pour y parvenir, nous proposons un modèle de régression par le biais de l'apprentissage utilisant l'extension floue de la théorie des fonctions de croyance. Le reste de cette section est dédiée à notre proposition.

2.3.1 Représentation floue pour la mesure de qualité

Comme expliqué dans la section 2.2.1, les valeurs subjectives *MOS* et *DMOS* sont des scores quantitatifs donnés par un ensemble d'observateurs humains sur la qualité d'une image. Les observateurs ont classé les images en fonction des degrés symboliques de qualité tels que *très mauvaise*, *mauvaise*, *moyenne*, *bonne* ou *très bonne* pour produire ces scores. Ils représentent l'évaluation humaine sur la qualité de l'image et sont encore considérés comme des références majeures dans l'évaluation de la qualité d'image. Les informations délivrées par les observateurs sont sous forme symboliques qui appartiennent au langage naturel ; ils sont difficilement transposables sous forme de valeurs réelles précises. De plus, ces informations dépendent de plusieurs facteurs qui peuvent influencer la fiabilité de l'observateur à bien évaluer la qualité de l'image, ce qui veut dire que ces scores sont difficilement reproductibles. Par conséquent, il semble tout à fait naturel d'introduire la représentation floue pour les valeurs *DMOS* dans notre travail.

En effet, la logique floue [68] est souvent utilisée pour des informations numériques ou symboliques. Représenter un concept ou une hypothèse H est effectuée au travers d'une fonction d'appartenance. La fonction d'appartenance peut prendre de nombreuses formes telles que Sigmoidale, Gaussienne, exponentielle ou trapézoïdale. Sa forme dépend de l'application. Par conséquent, il est intéressant d'utiliser la logique floue dans notre problématique d'estimation de scores de qualité.

De plus, notre processus doit pouvoir être efficace quelle que soit la dégradation (floue, compression ...). Cependant, étant placé dans un contexte aveugle, nous

sommes sans information *a priori* sur le vecteur de caractéristiques le plus adapté que nous devrions exploiter.

Afin de profiter au maximum des différents vecteurs de caractéristiques potentiels la combinaison semble être l'une des solutions. Cette fusion doit permettre d'obtenir une information synthétique qui nous aide à prédire un score de qualité. Cependant, puisque la fiabilité d'un vecteur de caractéristiques utilisé dépend de la dégradation présente sur l'image, il est également opportun de pouvoir introduire ce concept dans l'algorithme d'estimation.

Pour cela, nous proposons l'utilisation de la représentation floue conjointement à la théorie des fonctions de croyance. Une telle approche a été introduite par Petit-Renaud et al. dans [69] pour résoudre des problèmes de régression. Les auteurs ont proposé un algorithme de régression appelé *EVreg* utilisé dans le cas où les sources d'informations sont imparfaites (imprécises, incertaines ou partiellement contradictoires), ce qui correspond ici au cadre de notre problématique.

Dans la suite de ce chapitre, nous décrivons l'algorithme de régression *EVreg* (section 2.3.2). Les caractéristiques utilisées sont décrites dans la section 2.3.3 et une version multi-source de *EVreg* en combinant les différentes masses de croyances obtenues par les différents ensembles de caractéristiques est proposée en section 2.3.4.

2.3.2 Algorithme *EVreg* pour l'évaluation sans référence de la qualité d'image

Plusieurs auteurs ont proposé une extension de la théorie des fonctions de croyance aux ensembles flous [70, 71]. L'idée de base est de considérer que les éléments focaux du cadre de discernement pourraient être des ensembles flous ou des nombres réels. Avant de décrire plus précisément la modélisation proposée pour le *DMOS*, nous décrivons rapidement ce nouveau cadre théorique.

Soit A un ensemble flou du cadre de discernement Ω , nous notons $\mu_A()$ sa fonction d'appartenance. L'ensemble des ensembles flous de Ω est $[0, 1]^\Omega$. La masse de croyance floue \hat{m} est alors définie de $[0, 1]^\Omega$ vers $[0, 1]$ pour un ensemble fini $F(\hat{m})$ des sous-

ensembles flous tels que :

$$\begin{cases} \hat{m}(A) > 1 & \forall A \in F(\hat{m}) \\ \hat{m}(A) = 0 & \text{si } A \notin F(\hat{m}) \\ \sum_{A \in F(\hat{m})} \hat{m}(A) = 1 \end{cases} \quad (2.6)$$

Les opérations de révision (combinaison et affaiblissement) et la prise de décision vues précédemment dans la section 1.3.2 sont généralisées aux fonctions de croyances floues.

Étant donné cette représentation, l'algorithme *EVreg* [69] est décrit comme suit. Soit \mathbf{x} un vecteur arbitraire et \mathbf{y} sa valeur de sortie à prédire en utilisant l'ensemble de données d'apprentissage $\mathcal{L} = \{(x_i, \hat{m}_i)\}_{i=1}^L$ de taille L . La valeur connue \hat{m}_i est la fonction de croyance floue associée à un élément de l'ensemble d'apprentissage x_i . L'algorithme *EVreg* considère chaque élément de \mathcal{L} comme étant une source d'informations indépendante. Chaque (x_i, \hat{m}_i) permet de calculer une fonction de croyance floue $\hat{m}_{\mathbf{y}}[\mathbf{x}, (x_i, \hat{m}_i)]$ donnée par :

$$\begin{cases} \hat{m}_{\mathbf{y}}[\mathbf{x}, (x_i, \hat{m}_i)](A) = \alpha \cdot \hat{m}_i(A) \phi(d(\mathbf{x}, x_i)) & \text{si } A \in \mathcal{F}(\hat{m}_i) \setminus \mathcal{Y} \\ \hat{m}_{\mathbf{y}}[\mathbf{x}, (x_i, \hat{m}_i)](A) = 1 - \alpha \cdot \hat{m}_i(A) \phi(d(\mathbf{x}, x_i)) & \text{si } A = \mathcal{Y} \\ \hat{m}_{\mathbf{y}}[\mathbf{x}, (x_i, \hat{m}_i)](A) = 0 & \text{sinon} \end{cases} \quad (2.7)$$

avec \mathcal{Y} le cadre de discernement, ϕ une fonction décroissante de \mathfrak{R}^+ vers $[0, 1]$ telle que $\phi(0) = 1$ et $\lim_{d \rightarrow \infty} \phi(d) = 0$. α est le coefficient d'affaiblissement, d est une distance à déterminer en fonction des données. Pour combiner chaque élément, la règle de combinaison de Dempster peut être utilisée pour les fonction de croyance floues et la fonction de croyance floue finale est donnée par : $\oplus_{i=1}^L \hat{m}_{\mathbf{y}}[\mathbf{x}, (x_i, \hat{m}_i)]$. Petit-Renaud et al. [69] montrent que cette fonction de croyance floue peut être utilisée pour estimer la valeur de sortie \mathbf{y} . Nous rappelons que dans le cadre général \mathbf{y} peut être un ensemble flou, un intervalle ou un nombre entier.

Notre proposition est alors d'appliquer *EVreg* pour estimer le score de qualité \mathbf{y} d'une image inconnue I en utilisant notre ensemble de données d'apprentissage \mathcal{L} . L'ensemble \mathcal{L} est composé de L couples (x_l, y_l) avec $l = 1, \dots, L$ où x_l représente un vecteur de caractéristiques (nous reviendrons dans la partie 2.3.3 sur ces vecteurs) de l'image I_l et y_l sa valeur *DMOS* connue.

La première question à résoudre est de savoir comment construire une fonction d'appartenance pour représenter les scores *DMOS*. Les bases d'images utilisées dans

ce travail [34, 72, 36] contiennent des images avec plusieurs types et degrés de dégradation. Ces bases d'images ont été construites pour permettre aux auteurs de calibrer les algorithmes et de valider leurs travaux en utilisant des mesures perceptuelles de dégradations. Pour chaque image I_l une valeur *DMOS* y_l est attribuée. Cependant, aucune information n'est donnée sur la variabilité du *DMOS*, en d'autres termes, nous n'avons pas accès aux notes brutes des différents observateurs attribuées à chaque image. Le *DMOS* est l'unique valeur scalaire que nous possédons. Dans cette situation, nous proposons de réduire la fonction d'appartenance floue μ_l associée à I_l à une fonction Dirac μ_l telle que :

$$\begin{cases} \mu_l(y) = 1 & \text{si } y = y_l, y \in \mathcal{R}^+ \\ \mu_l(y) = 0 & \text{sinon} \end{cases} \quad (2.8)$$

où y_l est la valeur du *DMOS* associée à l'image.

Dans le cas où nous pourrions construire ou disposer de plus d'informations sur les valeurs *DMOS* et leur variabilité, la fonction d'appartenance floue pourrait prendre différentes formes telles que Sigmoidale, Gaussienne, exponentielle ou trapézoïdale. Cette modélisation est une perspective de ce travail. Pour cela, nous aurons à construire notre propre base de données d'images (avec une évaluation de la qualité par des observateurs).

L'algorithme *EVreg* (équation 2.7) a été conçu pour le cas général de données imprécises et non fiables représentées par des sous-ensembles flous pour effectuer une régression. Notre choix quant à la forme des fonctions d'appartenance μ_I permet d'appliquer *EVreg* dans le cas simple où les valeurs de sortie sont des valeurs uniques réelles. Dans ce cas, la fonction de croyance floue $\hat{m}_y[\mathbf{x}, (x_i, \hat{m}_i)]$ est équivalente à une fonction de croyance classique $m_y[\mathbf{x}, (x_i, y_i)]$. La fonction de croyance $m_y[\mathbf{x}, (x_i, y_i)]$ possède deux éléments focaux $\{y_i, \mathcal{Y}\}$ définis dans \mathcal{R}^+ . Elle est donnée par :

$$\begin{cases} m_y[\mathbf{x}, (x_i, y_i)](A) = \alpha \cdot \phi(d(\mathbf{x}, x_i)) & \text{si } A = y_i \\ m_y[\mathbf{x}, (x_i, y_i)](A) = 1 - \alpha \cdot \phi(d(\mathbf{x}, x_i)) & \text{si } A = \mathcal{Y} \\ m_y[\mathbf{x}, (x_i, y_i)](A) = 0 & \text{sinon} \end{cases} \quad (2.9)$$

Les paramètres α , ϕ et d ont les mêmes descriptions que pour l'équation 2.7. Les valeurs utilisées pour notre application seront spécifiées dans la section 2.4.

La masse de croyance finale $m_y[\mathbf{x}, \mathcal{L}]$ de l'image inconnue I peut être obtenue en utilisant le règle de combinaison de Dempster [69]. Pour éviter de prendre en compte

les éléments les plus éloignés et minimiser le temps de calcul, nous pouvons réduire la combinaison à un ensemble \mathcal{L}^K composé des K plus proches voisins de \mathbf{x} . Ainsi, la fonction de croyance finale $m_y[\mathbf{x}, \mathcal{L}^K]$ est donnée par :

$$m_y[\mathbf{x}, \mathcal{L}^K] = \oplus_{k=1}^K m_y[\mathbf{x}, (x_k, y_k)] \quad (2.10)$$

Si toutes les valeurs y_k sont différentes, l'ensemble des éléments focaux de $m_y[\mathbf{x}, \mathcal{L}^K]$ sera composé de $K + 1$ éléments. En supposant que le domaine \mathcal{Y} de y soit un intervalle borné $[y_{\text{inf}}, y_{\text{sup}}]$, la valeur *DMOS* estimée \tilde{y} de y est donnée par :

$$\tilde{y} = \sum_{k=1}^K m_y[\mathbf{x}, \mathcal{L}^K].y_k + m_y[\mathbf{x}, \mathcal{L}^K](\mathcal{Y}).\bar{y} \quad \text{avec} \quad \bar{y} = \frac{y_{\text{sup}} - y_{\text{inf}}}{2} \quad (2.11)$$

La figure 2.7 représente schématiquement l'algorithme *EVreg* pour une mesure sans référence d'évaluation de la qualité d'image.

2.3.3 Définition des descripteurs

À partir de l'état de l'art sur les mesures d'évaluation de la qualité d'image, et à travers toutes les études menées au cours des 20 dernières années, nous avons pu constater le grand nombre de méthodes d'évaluation de la qualité basées sur l'apprentissage automatique, où plusieurs caractéristiques porteuses d'informations pertinentes caractérisant les images ont été proposées. Notre objectif ici n'est pas de proposer de nouveaux descripteurs et de les comparer face à ceux déjà existants, mais de proposer une nouvelle méthode pour prédire le score reflétant la qualité de l'image. Puisque l'un des critères principaux aujourd'hui est la corrélation avec les résultats subjectifs des tests psycho-visuels, nous proposons de tester notre algorithme en utilisant les descripteurs déjà proposés et validés par la littérature. Dans la table 2.1, nous décrivons les différents descripteurs utilisés pour mener nos expériences¹. Ces caractéristiques issues de différents espaces de représentation ont été sélectionnées pour des fins de comparaisons et pour la disponibilité de leurs codes.

Nous proposons ainsi d'utiliser dans notre méthode chacune de ces caractéristiques pour l'évaluation de la qualité d'image. Dans l'algorithme *EVreg* décrit dans la figure 2.7, x représente l'un des éléments de l'ensemble $\mathcal{F}_v(x^i) = \{x^\phi, x^{\text{NSS-DCT}}, x^{\text{NSS}}, x^{\text{CURVE}}, x^{\text{BIQI}}, x^{\text{SSE}}\}$ des vecteurs de caractéristiques.

1. Les codes pour extraire toutes les caractéristiques présentées ci-dessous sont disponibles pour le téléchargement sur <http://live.ece.utexas.edu/research/quality/>

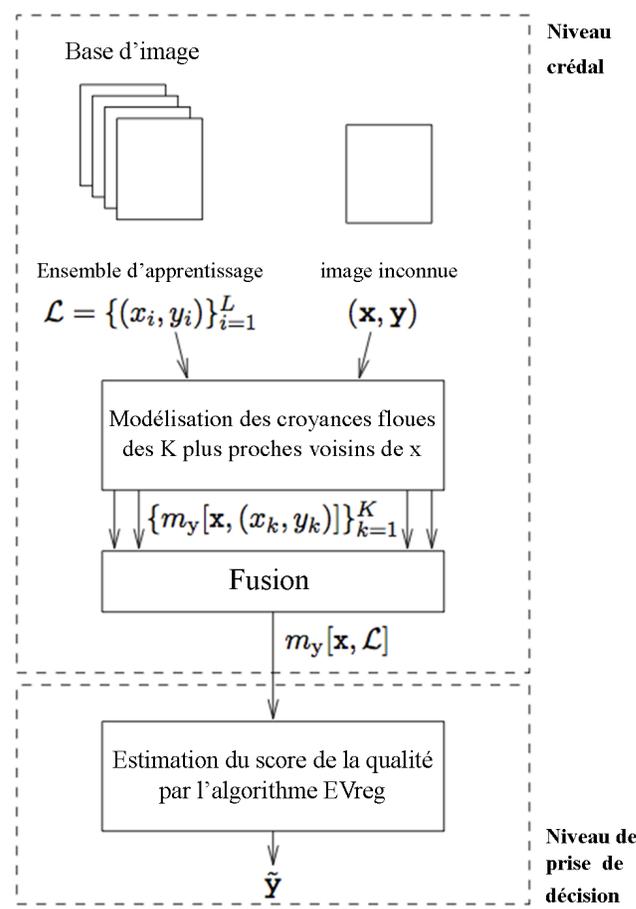


FIGURE 2.7 – L'algorithme *EVreg* appliqué pour une mesure d'évaluation de la qualité d'image

2.3.4 Un algorithme de régression évidentiel multi-sources

Une simple utilisation de l'algorithme *EVreg* est d'utiliser l'un des vecteurs de caractéristiques x^i d'une image I à partir de l'ensemble $\mathcal{F}_v(x^i)$. Pour chaque vecteur de caractéristiques, nous sommes alors en mesure d'estimer un score de qualité y d'une image inconnue.

Étant donné les scores de qualité estimés par chaque vecteur de caractéristiques, la question est maintenant de déterminer le «bon» vecteur de description à considérer. Dans un processus aveugle, lorsque l'image originale ou la nature de la dégradation sont inconnues, il est pratiquement impossible de faire ce choix car la précision et l'efficacité d'un vecteur de caractéristiques varient en fonction la dégradation de l'image.

Caractéristique	Dimension	Nature et domaine
x^ϕ	4	Représentation : le domaine de Fourier et spatial. La valeur moyenne et l'entropie de l'image en congruence de phase, l'entropie et le gradient de l'image dégradée [62].
$x^{NSS-DCT}$	24	Représentation : la transformée en cosinus discrète (DCT). Caractéristiques issues de la statistique de scènes naturelles (densité gaussienne généralisée des coefficients DCT, coefficients des variations de fréquence ...) [63].
x^{NSS}	36	Représentation : le domaine spatial. Caractéristiques issues de la distribution gaussienne généralisée des coefficients de luminance normalisés entre les pixels voisins [64].
x^{CURVE}	12	Représentation : le domaine curvelet. Comprend les coordonnées des maxima des histogrammes des coefficients curvelet, les distributions d'énergie en fonction de l'orientation et de l'échelle [67].
x^{BIQI}	18	Représentation : le domaine de transformée en ondelettes. La variance et le paramètre de forme sont calculés sur trois échelles et trois orientations (3 échelles x 3 orientations x 2 paramètres) [73].
x^{SSE}	12	Représentation : le domaine spatial et spectral. Les caractéristiques représentent les entropies spatiales et spectrales calculées à partir de chaque bloc $M \times M$ de l'image [74].

TABLE 2.1 – Les différentes caractéristiques utilisées

L'un des avantages de la théorie des fonction de croyance est qu'elle offre un outil mathématique pour la fusion de l'information. Afin de bénéficier de l'information apportée par chaque vecteur de caractéristiques, nous proposons un processus de fusion des fonctions de croyance produites par la règle de combinaison de Dempster (voir l'équation (2.10)) avant d'appliquer la régression.

Plus formellement, pour un élément inconnu \mathbf{x} et à partir de chaque vecteur de caractéristiques, l'application de l'équation (2.10) produit une croyance $m_v[\mathbf{x}^f, \mathcal{L}^{(K,f)}]$ où \mathbf{x}^f représente l'un des vecteurs de caractéristiques de l'ensemble $\mathcal{F}_v(\mathbf{x}^i)$ et $\mathcal{L}^{(K,f)}$ représente les couples des K plus proches voisins (x_k^f, y_k) de \mathbf{x} . Notons que les couples des K plus proches voisins (x_k^f, y_k) pour $k = 1, \dots, K$ pourront être différents selon le vecteur de caractéristiques.

Afin de combiner les informations fournies par chaque vecteur de caractéristiques, nous pouvons utiliser la règle de combinaison de Dempster. L'expression de la fonction

de croyance est donnée par :

$$m_y^*[\mathbf{x}, \mathcal{L}] = \bigoplus_{f=1}^{|\mathcal{F}_v|} m_y[\mathbf{x}^f, \mathcal{L}^{(k,f)}] \quad (2.12)$$

Nous constatons que nous fusionnons les croyances associées à chaque vecteur de caractéristiques. Enfin, la valeur estimée de y est calculée en appliquant l'équation 2.11 à m^* . Cet algorithme est appelé *ms-EVreg* et nous permet donc d'estimer un score de qualité pour chaque image à partir de $|\mathcal{F}_v|$ vecteurs de caractéristiques calculées sur une base d'apprentissage.

Dans la section suivante, plusieurs tests expérimentaux sont réalisés afin d'évaluer la performance des algorithmes *EVreg* et *ms-EVreg* et ainsi de caractériser l'apport de la fusion des différents descripteurs.

2.4 Évaluation des algorithmes *EVreg* et *ms-EVreg*

2.4.1 Protocole expérimental

Nous avons testé nos approches sur la base d'image LIVE [34]. Cette base contient un total de 982 images (dont 203 images ne présentent aucune dégradation) obtenues à partir de 29 images de référence qui ont été affectées par différents types de distorsions (compression JPEG2000 et JPEG, bruit blanc (WN : *White Noise*), flou Gaussien (GBLUR : *Gaussian Blur*) et les erreurs de bits dans le canal de transmission (FF : *Fast Fading*)). La qualité perçue de chaque image de la base LIVE est fournie et représentée par un score *DMOS*.

Pour les expérimentations, nous n'avons conservé que les images dégradées, soit le total de 779 images qui sont divisées en un ensemble d'apprentissage et un ensemble de test. Afin d'assurer l'indépendance de la mesure d'évaluation de la qualité, 80% des images dégradées ont été sélectionnées au hasard pour définir notre ensemble d'apprentissage \mathcal{L} de L images, alors que l'ensemble de test \mathcal{T} est composé des 20% des images restantes. Nous avons répété 1000 fois ce processus aléatoire et nous avons évalué la performance de nos propositions sur chaque série d'images de test. La médiane des indices de performance à travers les 1000 itérations est retenue comme la performance finale de l'algorithme.

En ce qui concerne la modélisation des croyances, la distance Euclidienne d a été utilisée pour calculer la distance entre chaque élément et ses k plus proches voisins.

Après plusieurs tests, la meilleure performance est obtenue avec $k = 13$, c'est cette valeur qui a été retenue pour la suite. Dans cette première version de notre algorithme, la fiabilité de chaque vecteur de caractéristiques n'est pas prise en compte, le coefficient α permet simplement d'éviter une situation de conflit total et la division numérique par zéro. Dans ce cas, nous avons choisis le coefficient $\alpha = 0.95$ (équation 2.9). Comme le propose Denœux dans [75], la fonction décroissante monotone ϕ utilisée est donnée par :

$$\phi(d) = \gamma \exp(-d^2) \quad (2.13)$$

avec $\gamma \in [0, 1]$ un paramètre réel. Nous proposons d'utiliser cette définition de fonction ϕ .

2.4.2 Résultats de l'*EVreg* avec des caractéristiques prises séparément

Dans un premier temps, nous avons utilisé les descripteurs décrits dans la table 2.1 séparément, puis à l'aide de l'algorithme *EVreg* une relation fonctionnelle entre chaque groupe de caractéristiques et les valeurs *DMOS* est construite. Nous avons choisis d'utiliser les coefficients de corrélation de Spearman et les coefficients de corrélation linéaire de Pearson pour évaluer la pertinence des valeurs *DMOS* estimées \tilde{y}_i par nos algorithmes, par rapport à celles fournies dans la base d'images LIVE. Notons que ces deux coefficients de corrélation sont les plus utilisés dans la littérature.

Comme proposé dans l'article [76], pour calculer les coefficients de corrélation linéaire de Pearson, la relation entre les prédictions et les scores subjectives sera estimée à l'aide d'une régression non linéaire entre les valeurs prédites par *EVreg* et les *DMOS* correspondant pour éliminer toute non linéarité et faciliter la comparaison avec la littérature. Pour souci de comparaison, nous suivons la même procédure pour la non-linéarité en utilisant une fonction logistique à cinq paramètres définie par :

$$\text{Qualité}(\tilde{y}) = \beta_1 \left(\frac{1}{2} - \frac{1}{1 + \exp(\beta_2(\tilde{y} - \beta_3))} \right) + \beta_4 \tilde{y} + \beta_5 \quad (2.14)$$

avec les mêmes valeurs $\beta_1, \beta_2, \beta_3, \beta_4$ et β_5 présentées dans [76].

Nous avons comparé nos résultats à deux autres méthodes de régression. La première appelée *GRNN* (General regression Neural Network) repose sur la régression générale basée réseaux de neurones est proposée dans le cadre d'une mesure sans référence d'évaluation de la qualité d'image[62]. La seconde méthode est basée sur

la régression par les vecteurs de support de Steve Gunn [77] avec comme noyau la fonction de base radiale (*RBF* : Radial Basic Function). Nous avons choisi ces deux méthodes de régression car elles sont des mesures de référence dans le contexte des mesures de qualité d'images basées sur l'apprentissage automatique.

Nos résultats sont synthétisés dans les tableaux 2.2 et 2.3. Ils montrent les résultats expérimentaux obtenus par chaque groupe de caractéristiques sur 1000 itérations en termes de médiane du coefficient de corrélation de Spearman et de Pearson. Les meilleurs résultats sont affichés en gras dans les tableaux. Chaque colonne correspond à un type de dégradation. La colonne "LIVE" présente les résultats obtenus sur l'ensemble de la base LIVE qui comprend donc l'ensemble des dégradations.

		JPEG2000	JPEG	WN	GBLUR	FF	LIVE
x^ϕ	EVreg	0,8514	0,9168	0,9729	0,8352	0,8139	0,8811
	GRNN	0,8464	0,9168	0,9762	0,8317	0,8090	0,8807
	SVM	0,8385	0,8906	0,9556	0,8772	0,8089	0,8782
$x^{NSS-DCT}$	EVreg	0,9120	0,9412	0,9401	0,9165	0,8657	0,9193
	GRNN	0,8595	0,9252	0,9091	0,8973	0,8404	0,9001
	SVM	0,9212	0,8707	0,9550	0,9280	0,8862	0,9125
x^{NSS}	EVreg	0,8423	0,9176	0,9773	0,9052	0,8743	0,9069
	GRNN	0,8071	0,8837	0,9655	0,8264	0,8528	0,8837
	SVM	0,9039	0,9527	0,9439	0,9234	0,8308	0,9185
x^{CURVE}	EVreg	0,8816	0,9340	0,9718	0,9029	0,8660	0,9098
	GRNN	0,8419	0,9149	0,9637	0,8305	0,8400	0,8915
	SVM	0,9212	0,9417	0,9754	0,8603	0,8666	0,9189
x^{BIQI}	EVreg	0,7697	0,7739	0,9329	0,8800	0,7431	0,7600
	GRNN	0,7475	0,8010	0,9333	0,8419	0,7244	0,7845
	SVM	0,8794	0,9333	0,8010	0,8044	0,6871	0,8309
x^{SSE}	EVreg	0,8838	0,9356	0,9468	0,9090	0,8434	0,8959
	GRNN	0,8777	0,9114	0,9117	0,8736	0,8390	0,8835
	SVM	0,9233	0,8550	0,8597	0,9435	0,8641	0,8840

TABLE 2.2 – Les résultats des coefficients de corrélation de Spearman pour chaque groupe de caractéristiques

À partir de ces résultats, en utilisant les différents vecteurs de caractéristiques

		JPEG2000	JPEG	WN	GBLUR	FF	LIVE
x^ϕ	EVreg	0,8787	0,9493	0,9865	0,8708	0,8680	0,8934
	GRNN	0,8725	0,9537	0,9899	0,8686	0,8750	0,8959
	SVM	0,8730	0,9410	0,9781	0,9097	0,8858	0,8943
$x^{NSS-DCT}$	EVreg	0,9349	0,9594	0,9727	0,9410	0,9100	0,9279
	GRNN	0,8956	0,9520	0,9546	0,9245	0,9207	0,9181
	SVM	0,9484	0,9221	0,9772	0,9422	0,9328	0,9242
x^{NSS}	EVreg	0,8785	0,9512	0,9896	0,9316	0,9170	0,9200
	GRNN	0,8495	0,9319	0,9839	0,8831	0,9077	0,9020
	SVM	0,9337	0,9729	0,9687	0,9444	0,8911	0,9276
x^{CURVE}	EVreg	0,9084	0,9591	0,9863	0,9278	0,9126	0,9189
	GRNN	0,8784	0,9499	0,9821	0,8827	0,8928	0,9061
	SVM	0,9452	0,9681	0,9887	0,8972	0,9284	0,9345
x^{BIQI}	EVreg	0,8197	0,8239	0,9623	0,9140	0,8209	0,7716
	GRNN	0,8060	0,8529	0,9630	0,9004	0,8202	0,8005
	SVM	0,9128	0,9566	0,8618	0,8406	0,7746	0,8405
x^{SSE}	EVreg	0,9070	0,9588	0,9678	0,9328	0,9033	0,9088
	GRNN	0,9087	0,9421	0,9561	0,9061	0,8958	0,9007
	SVM	0,9469	0,9169	0,9169	0,9530	0,9074	0,8947

TABLE 2.3 – Les résultats des coefficients de corrélation de Pearson pour chaque groupe de caractéristiques

séparément, nous remarquons les différents comportements des trois méthodes sur les divers types de distorsions ainsi que sur le cas général de dégradation (en prenant en compte toute la base d'image). À partir du tableau 2.2, nous remarquons que notre algorithme *EVreg* surpasse les deux autres méthodes sur la base d'image LIVE entière, en utilisant l'un des descripteurs suivants $\{x^\phi, x^{NSS-DCT}$ et $x^{SSE}\}$. Cependant, en prenant distinctement chaque type de distorsion, les résultats diffèrent selon le groupe de caractéristiques utilisé. Par exemple, *EVreg* obtient de meilleurs résultats en utilisant les caractéristiques issues de la congruence de phase sur les distorsions de type JPEG2000, JPEG et FF. Nous pouvons également constater que pour le bruit blanc *EVreg* obtient un résultat satisfaisant en utilisant les caractéristiques x^{NSS} et x^{SSE} . Cette même constatation peut être faite à partir de l'analyse du tableau 2.3 (coefficient de Pearson). Sur la base de ces résultats, le point le plus significatif est que

EVreg obtient de meilleurs résultats au moins une fois sur chaque type de dégradation.

Cependant, le vecteur de caractéristiques obtenant la meilleure corrélation avec le *DMOS* est variable suivant la fonction de la distorsion sans que nous puissions définir ou préconiser l'utilisation a priori de l'un des vecteurs de caractéristiques parmi l'ensemble des vecteurs à notre disposition, ce qui reste un constat non satisfaisant.

Ceci nous amène à proposer un mécanisme de fusion profitant des variabilités de comportement de *EVreg* en fonction des différentes caractéristiques pour améliorer les résultats de corrélation. C'est la méthode *ms-EVreg*, que nous avons précédemment introduite.

2.4.3 Évaluation de *ms-EVreg*

Afin d'améliorer les résultats et de tirer pleinement profit des outils de fusion que propose la théorie des fonctions de croyance, notre seconde proposition est de combiner les différentes masses de croyances obtenues par chaque vecteur de caractéristiques. L'estimation du *DMOS* est alors basée sur les informations fournies par l'ensemble des vecteurs de caractéristiques.

Jusqu'à présent l'opération d'affaiblissement introduite dans l'équation 2.9 avant la combinaison n'avait pas de signification liée à la fiabilité des sources. Pour rappel, cette opération est représentée par un facteur $\alpha \in [0, 1]$, appelé coefficient d'affaiblissement. Un coefficient d'affaiblissement $\alpha = 1$ signifie qu'il n'y a pas d'affaiblissement, et la source est connue pour être entièrement fiable. Au contraire, un taux d'atténuation nulle $\alpha = 0$ signifie que nous ne pouvons guère faire confiance à la source.

Notre proposition est maintenant de lier cet affaiblissement à la notion de fiabilité des sources d'informations. Différents propositions et expérimentations sont réalisées pour démontrer la pertinence de cette approche.

Comme dans la section 2.4.2, nous prenons dans un premier temps une valeur classique du taux d'affaiblissement pour chaque groupe de caractéristiques ($\alpha^f = 0,95$). Nous considérons donc ici que toutes les sources d'informations ont la même fiabilité.

Le même protocole utilisé dans les expérimentations précédentes a été suivi, à savoir 80% des images sélectionnées aléatoirement pour l'apprentissage et 20% pour les tests sur 1000 itérations. Nous comparons les résultats obtenus par *ms-EVreg* avec le meilleur score obtenu par chaque groupe de caractéristiques exploitant l'un des méthodes :

EVreg, GRNN ou SVM. Par exemple, en utilisant le vecteur de caractéristiques x^ϕ , le meilleur résultat sur JPEG2000 est obtenu par GRNN=0.8514 (tableau 2.2). Les résultats obtenus par la fusion des descripteurs à l'aide d'un taux d'affaiblissement $\alpha^f = 0,95$ sont présentés dans les tableaux 2.4 et 2.5.

	JPEG2000	JPEG	WN	GBLUR	FF	LIVE
x^ϕ	0,8514	0,9168	0,9762	0,8772	0,8139	0,8811
$x^{NSS-DCT}$	0,9212	0,9412	0,9550	0,9280	0,8862	0,9193
x^{NSS}	0,9039	0,9527	0,9773	0,9234	0,8743	0,9185
x^{CURVE}	0,9212	0,9417	0,9754	0,9029	0,8666	0,9189
x^{BIQI}	0,8794	0,9333	0,9333	0,8800	0,7431	0,8309
x^{SSE}	0,9233	0,9356	0,9468	0,9435	0,8641	0,8959
ms-EVreg	0,8736	0,9246	0,9685	0,8553	0,8376	0,9062

TABLE 2.4 – La médiane des coefficients de corrélation de Spearman sur 1000 ensembles apprentissage-test sur la base LIVE

	JPEG2000	JPEG	WN	GBLUR	FF	LIVE
x^ϕ	0,8787	0,9537	0,9899	0,9097	0,8858	0,8959
$x^{NSS-DCT}$	0,9484	0,9594	0,9772	0,9245	0,9328	0,9279
x^{NSS}	0,9337	0,9729	0,9896	0,9444	0,9170	0,9276
x^{CURVE}	0,9452	0,9681	0,9887	0,9278	0,9284	0,9345
x^{BIQI}	0,9128	0,9566	0,9623	0,9140	0,8209	0,8405
x^{SSE}	0,9469	0,9588	0,9678	0,9530	0,9074	0,9088
ms-EVreg	0,9063	0,9575	0,9840	0,9046	0,9001	0,9209

TABLE 2.5 – La médiane des coefficients de corrélation de Pearson sur 1000 ensembles apprentissage-test sur la base LIVE

Nous pouvons constater que la fusion en utilisant un coefficient d'affaiblissement qui considère chaque source d'information de façon équivalente ne permet pas d'améliorer les coefficients de corrélation par rapport à ceux obtenus avec $x^{NSS-DCT}$, x^{NSS} et x^{CURVE} . Ce résultat semble indiquer que de considérer de la même façon les informations portées par les différentes caractéristiques n'est pas une stratégie efficace.

Nous proposons d'introduire à un affaiblissement adaptatif en utilisant des coefficients d'atténuation optimisés qui tiennent compte de la fiabilité de chaque ensemble de caractéristiques. C'est l'objet de la proposition suivante.

2.5 Les coefficients d'affaiblissement optimisés pour *ms-EVreg*

2.5.1 Le principe

Pour l'algorithme *ms-EVreg*, la fonction de croyance finale est obtenue en fusionnant l'information fournie à partir de plusieurs ensembles de caractéristiques indépendantes avec la même puissance de discrimination. En d'autres termes, leurs poids sont les mêmes pendant le processus de fusion. Cependant, les tableaux 2.4 et 2.5 montrent clairement que les coefficients de corrélation varient en fonction du vecteur de caractéristiques utilisé. Par exemple la meilleure performance sur FF et LIVE en terme de corrélation de Spearman est obtenue par $EVreg(x^{NSS-DCT})$, par contre sur JPEG2000 et GBLUR $EVreg(x^{SSE})$ est meilleur.

Ainsi, nous proposons de prendre en compte cette dépendance des résultats aux caractéristiques en affaiblissant la fonction de croyance en fonction de la fiabilité de chaque ensemble de caractéristiques. La notion de fiabilité et l'estimation de α^f en fonction du vecteur de caractéristiques sont discutées dans la section suivante.

Soit α^f le coefficient d'affaiblissement associé à un vecteur de caractéristiques f . Nous proposons de substituer dans l'équation (2.9) le coefficient α par α^f . Pour un voisin k , la fonction de croyance floue est notée $m_y^\bullet[\mathbf{x}^f, (x_k, y_k)]$ avec :

$$\begin{cases} m_y^\bullet[\mathbf{x}^f, (x_k^f, y_k), \alpha^f](A) = \alpha^f \cdot \phi(d(\mathbf{x}^f, x_k^f)) & \text{si } A = y_k \\ m_y^\bullet[\mathbf{x}^f, (x_k^f, y_k), \alpha^f](A) = 1 - \alpha^f \cdot \phi(d(\mathbf{x}^f, x_k^f)) & \text{si } A = \mathcal{Y} \\ m_y^\bullet[\mathbf{x}^f, (x_k^f, y_k), \alpha^f](A) = 0 & \text{sinon} \end{cases} \quad (2.15)$$

et par conséquent

$$m_y^\bullet[\mathbf{x}^f, \mathcal{L}^{(K,f)}] = \bigoplus_{k=1}^K m_y^\bullet[\mathbf{x}^f, (x_k^f, y_k), \alpha^f] \quad (2.16)$$

$$m_y^\bullet[\mathbf{x}, \mathcal{L}] = \bigoplus_{f=1}^{|\mathcal{F}_v|} m_y^\bullet[\mathbf{x}^f, \mathcal{L}^{(K,f)}] \quad (2.17)$$

Cette nouvelle version est appelée *wms-EVreg* pour *weighted ms-EVreg*.

2.5.2 Estimation des coefficients d'affaiblissement

Nous proposons deux façons d'estimer les paramètres d'affaiblissement α^f pour l'algorithme *wms-EVreg*.

La première proposition utilise directement les valeurs des coefficients de corrélation obtenues pour chaque vecteur de caractéristiques comme coefficient d'affaiblissement α^f . Ce choix est justifié par le fait que les coefficients de corrélation reflètent la capacité de notre méthode à estimer les valeurs de *DMOS* en corrélation avec la perception humaine. De plus, leurs valeurs sont bornées par l'intervalle $[0, 1]$, correspondant à l'intervalle de définition des coefficients d'affaiblissements.

Pour cela, l'algorithme *EVreg* est d'abord appliqué avec un coefficient $\alpha = 0,95$ pour chaque vecteur de caractéristiques et les coefficients de corrélation de Spearman sont calculés. Ensuite, *wms-EVreg* est appliqué en utilisant un coefficient α^f égal au coefficient de corrélation obtenu. Les bases de données d'images \mathcal{L} et \mathcal{T} sont toujours les mêmes dans les deux étapes.

La deuxième proposition consiste à estimer les coefficients d'affaiblissement en se basant sur la matrice de confusion, matrice souvent utilisée pour évaluer les performances dans le contexte de la classification (voir section 1.3.2).

Comme l'algorithme de régression *EVreg* est issu de l'algorithme de classification *EV-KNN*, notre hypothèse est que la capacité d'un vecteur de descripteurs x^f à fournir une bonne estimation du score de qualité est liée à sa capacité à identifier correctement une dégradation spécifique inconnue en utilisant *EV-KNN*. Ainsi, nous proposons tout d'abord d'appliquer *EV-KNN* sur notre base de données pour chaque vecteur de caractéristiques précédemment utilisé pour identifier la dégradation présente sur une image inconnue. Pour chaque caractéristique f , une matrice de confusion C^f est alors estimée. Nous proposons de définir α^f par

$$\alpha^f = \frac{\sum_{i=1}^l n_{ii}}{n} \quad (2.18)$$

où pour un vecteur de caractéristiques fixé, n_{ii} représente le nombre d'images correctement classées par *EV-KNN*, n le nombre total d'images et l le nombre de classe.

Ce coefficient définit la fiabilité que l'on peut attendre de la classification à partir d'un vecteur de caractéristiques particulier. Nous proposons ainsi d'utiliser cette seconde formulation pour affaiblir les fonctions de masse dans l'algorithme *wms-EVreg*.

Ces deux formulations sont expérimentées et étudiées dans la section suivante.

2.5.3 Évaluation de *wms-EVreg*

Nous présentons dans cette section les différents résultats obtenus en utilisant les coefficients d'affaiblissement optimisés proposés précédemment.

La première expérimentation considère les coefficients de corrélation de Spearman obtenus à partir des mesures avec *EVreg* sur la base LIVE comme taux d'affaiblissement. Dans ce cas, nous prenons les valeurs extraites du tableau 2.2. Celles-ci sont synthétisées dans le tableau 2.6.

	α^ϕ	$\alpha^{NSS-DCT}$	α^{NSS}	α^{CURVE}	α^{BIQI}	α^{SSE}
Valeur α^f	0.8811	0.9193	0.9185	0.9189	0.8309	0.8959

TABLE 2.6 – Première proposition des coefficients d'affaiblissement

Nous avons appelé cette méthode *wms-EVreg*. Nous remarquons déjà d'après ces coefficients que x^{BIQI} est considéré comme étant le moins fiable, sa fonction de masse associée à ce vecteur de caractéristiques sera donc plus affaiblie que les autres qui ont un poids plus au moins équivalents compris entre 0.8811 et 0.9189 avant l'étape de fusion.

La seconde expérimentation considère les coefficients de fiabilité α_f obtenus à partir de la matrice de confusion. Pour chaque vecteur de descripteurs, nous appliquons *EV-KNN* sur la base de données LIVE.

Une validation croisée à 5 groupes est appliquée, où 80% des images ont été prises pour l'apprentissage et les 20% restantes pour le test. Pour chaque vecteur de caractéristiques, une classification des images de test estimant la nature de la distorsion est réalisée. Ensuite nous calculons la matrice de confusion qui servira à obtenir les coefficients de fiabilité. Ce processus a été répété 100 fois pour chaque groupe de caractéristiques. Enfin, nous prenons la valeur moyenne du coefficient de fiabilité sur les 100 itérations. Un exemple de la matrice de confusion du vecteur x^ϕ pour une seule itération est présenté dans le tableau 2.7. Le coefficient de fiabilité pour cette itération et pour ce vecteur de caractéristiques est ensuite calculé comme suit :

$$\alpha^\phi = \frac{112 + 107 + 107 + 105 + 85}{779} = 0.6624.$$

Nous avons obtenu les valeurs présentées dans le tableau 2.8 pour chaque groupe

Décision	Vérité	JPEG2000	JPEG	WN	BLUR	FF
JPEG2000		112	41	10	3	3
JPEG		29	107	43	4	4
WN		2	26	107	46	6
BLUR		0	0	25	105	14
FF		0	0	0	7	85

 TABLE 2.7 – Exemple du calcul du coefficient de fiabilité du vecteur x^ϕ

de caractéristiques.

	α^ϕ	$\alpha^{NSS-DCT}$	α^{NSS}	α^{CURVE}	α^{BIQI}	α^{SSE}
Valeur α^f	0.6218	0.6591	0.6098	0.6261	0.5111	0.5961

TABLE 2.8 – Deuxième proposition des coefficients de fiabilité issus de la matrice de confusion

En comparaison par rapport aux coefficients de fiabilité obtenus de la première proposition (tableau 2.6), nous constatons la même tendance pour x^{BIQI} , donc sa fonction de masse sera plus affaiblie par rapport aux autres vecteurs de caractéristiques. Nous pouvons aussi remarquer que la valeur des coefficients de fiabilité de tous les vecteurs a diminué, pour atteindre le maximum de 0.6591 pour $x^{NSS-DCT}$. De plus la variabilité entre ces coefficients d'affaiblissement est plus importante : chaque type de caractéristique pondère ainsi de façon bien plus distincte les fonctions de masse initiales, ce qui apporte un comportement moins homogène du processus de fusion. Ces valeurs sont utilisées ci-après avec *wms-EVreg*. Les résultats associés sont appelés *wms-EVreg2*.

Nous avons comparé les résultats de *wms-EVreg* et de *wms-EVreg2* avec le meilleur score obtenu par chaque groupe de caractéristiques au cours des premières expériences.

Les médianes des résultats expérimentaux sont synthétisés dans les tableaux 2.9 et 2.10. Nous pouvons remarquer que le taux d'affaiblissement contribue pleinement à la performance et à l'amélioration de la corrélation en désavantageant le groupe de caractéristiques le moins fiable et c'est le cas pour toutes les distorsions. Le point le plus important que nous pouvons soulever est l'évolution de la performance de la méthode

($ms-EVreg < wms-EVreg < wms-EVreg2$) quelque soit le type de la distorsion. Avec ces résultats, nous pouvons mettre en évidence l'apport de l'affaiblissement des fonctions de croyance avec des coefficients de fiabilité optimisés issus de la matrice de confusion.

Cette optimisation des coefficients de fiabilité conduit à l'obtention d'une meilleure performance en termes de corrélation avec les scores subjectifs humain par rapport à $ms-EVreg$ et $wms-EVreg$; il s'agit d'un bon compromis pour les différents types de distorsions. Par exemple, le meilleur résultat de corrélation sur JPEG2000 est obtenu par SVM en utilisant x^{SSE} , par contre pour les autres distorsions $wms-EVreg2$ surpasse SVM en particulier sur JPEG, WN et sur toute la base LIVE. Les SVM basés sur la caractéristiques x^{NSS} obtiennent la meilleure corrélation sur JPEG, mais elle est surpassée par $wms-EVreg2$ sur JPEG2000, WN, FF et toute la base LIVE.

	JPEG2000	JPEG	WN	GBLUR	FF	LIVE
x^ϕ	0,8514	0,9168	0,9762	0,8772	0,8139	0,8811
$x^{NSS-DCT}$	0,9212	0,9412	0,9550	0,9280	0,8862	0,9193
x^{NSS}	0,9039	0,9527	0,9773	0,9234	0,8743	0,9185
x^{CURVE}	0,9212	0,9417	0,9754	0,9029	0,8666	0,9189
x^{BIQI}	0,8794	0,9333	0,9333	0,8800	0,7431	0,8309
x^{SSE}	0,9233	0,9356	0,9468	0,9435	0,8641	0,8959
$ms-EVreg$	0,8736	0,9246	0,9685	0,8553	0,8376	0,9062
$wms-EVreg$	0,8885	0,9350	0,9701	0,8675	0,8475	0,9151
$wms-EVreg2$	0,9168	0,9486	0,9788	0,9270	0,8651	0,9309

TABLE 2.9 – La médiane des coefficients de corrélation de Spearman sur 1000 répétitions ensembles apprentissage-test sur la base LIVE

L'étude de la performance des mesures d'évaluation de la qualité d'image obtenues individuellement sur chaque type de distorsion a été réalisée pour examiner la faiblesse ou la robustesse de l'algorithme envers un type de dégradation spécifique. Cependant, en pratique, nous ne connaissons pas le type de dégradation subie par l'image inconnue et pour laquelle nous souhaitons évaluer la qualité : c'est le cas général présenté dans la colonne LIVE qui considère toutes les dégradations de la base. Dans les tableaux 2.9 et 2.10, nous pouvons constater que l'approche $wms-EVreg2$ donne de meilleurs résultats, sachant que la mesure proposée ne nécessite aucune connaissance préalable de la distorsion d'image.

	JPEG2000	JPEG	WN	GBLUR	FF	LIVE
x^ϕ	0,8787	0,9537	0,9899	0,9097	0,8858	0,8959
$x^{NSS-DCT}$	0,9484	0,9594	0,9772	0,9245	0,9328	0,9279
x^{NSS}	0,9337	0,9729	0,9896	0,9444	0,9170	0,9276
x^{CURVE}	0,9452	0,9681	0,9887	0,9278	0,9284	0,9345
x^{BIQI}	0,9128	0,9566	0,9623	0,9140	0,8209	0,8405
x^{SSE}	0,9469	0,9588	0,9678	0,9530	0,9074	0,9088
<i>ms-EVreg</i>	0,9063	0,9575	0,9840	0,9046	0,9001	0,9209
<i>wms-EVreg</i>	0,9165	0,9632	0,9852	0,9116	0,9053	0,9285
<i>wms-EVreg2</i>	0,9382	0,9711	0,9898	0,9489	0,9144	0,9404

TABLE 2.10 – La médiane des coefficients de corrélation de Pearson sur 1000 répétitions ensembles apprentissage-test sur la base LIVE

2.5.4 Comparaison avec d'autres mesures d'évaluation

Nous proposons maintenant de confronter notre mesure de qualité sans référence à différents algorithmes de la littérature. Trois mesures d'évaluation avec référence ont été utilisées pour la comparaison : rapport PSNR [38], l'indice de similarité structurelle SSIM [78] et VIF (Visual Information Fidelity) [46]. Nous avons également effectué des comparaisons de performance entre *wms-EVreg2* et deux autres mesures sans référence : BLIINDS-II [63] et NIQE [66]. Puisque les mesures avec référence et NIQE ne nécessitent pas d'étape d'apprentissage, pour assurer une comparaison équitable entre les méthodes, les coefficients de corrélation des scores prédits avec les jugements humains de la qualité ne sont rapportées que sur les images de l'ensemble de test. Les résultats sont présentés dans les tableaux 2.11 et 2.12. À partir de ces résultats expérimentaux, nous pouvons dire que :

- *wms-EVreg2* est supérieur à la mesure avec référence PSNR sur toutes les catégories de distorsion.
- *wms-EVreg2* surpasse également SSIM pour les images avec du bruit blanc WN et toute la base LIVE. *wms-EVreg2* présente une corrélation pertinente avec les opinions subjectives de la qualité d'image ; sa performance est très compétitive avec des approches avec référence.
- Comparé aux mesures de qualité sans référence, *wms-EVreg2* est plus performant que BLIINDS-II et NIQE sur toutes les distorsions individuelles exceptées

JPEG2000 ainsi que sur toute la base LIVE.

	JPEG2000	JPEG	WN	GBLUR	FF	LIVE
PSNR	0.8877	0.8716	0.9791	0.7737	0.8788	0.8735
SSIM	0.9428	0.9616	0.9701	0.9316	0.9453	0.9129
VIF	0.9538	0.9142	0.9857	0.9731	0.9659	0.9526
BLIINDS-II	0.9199	0.9382	0.9518	0.9027	0.8277	0.9131
NIQE	0.8088	0.8930	0.9686	0.8944	0.8599	0.8219
wms-EVreg2	0,9168	0,9486	0,9788	0,9270	0,8651	0,9309

TABLE 2.11 – La médiane des coefficients de corrélation de Spearman sur 1000 ensembles apprentissage-test sur la base LIVE

	JPEG2000	JPEG	WN	GBLUR	FF	LIVE
PSNR	0.8786	0.8661	0.9802	0.7786	0.8773	0.8601
SSIM	0.9401	0.9406	0.9726	0.8939	0.9486	0.9076
VIF	0.9625	0.9430	0.9904	0.9745	0.9697	0.9504
BLIINDS-II	0.9415	0.9559	0.9723	0.9268	0.8831	0.9189
NIQE	0.8178	0.9249	0.9754	0.9207	0.8861	0.8316
wms-EVreg2	0,9382	0,9711	0,9898	0,9489	0,9144	0,9404

TABLE 2.12 – La médiane des coefficients de corrélation de Pearson sur 1000 ensembles apprentissage-test sur la base LIVE

Afin d'analyser la performance de *wms-EVreg2* et d'illustrer sa bonne corrélation avec les scores subjectifs, un exemple d'une image originale extraite de la base LIVE est présenté dans la figure 2.8, avec différentes distorsions. Nous pouvons voir que ces images ont une qualité perceptuelle nettement différente.

Dans le but de faciliter l'analyse des résultats, nous présentons ici l'échelle d'interprétation des valeurs numériques des différentes mesures de qualité étudiées.

Il est à noter qu'il est difficile de comparer les résultats numériques des différentes mesures, par exemple les valeurs NIQE représentent les distances entre les caractéristiques du modèle de scène naturelle d'un corpus d'image et les caractéristiques extraites de l'image dégradée. Par conséquent nous ne sommes pas en mesure de définir une échelle normalisée de qualité pour cette mesure.

	Très bonne qualité	Mauvaise qualité
PSNR	≥ 40	< 30
SSIM	1	0
VIF	1	0
BLIINDS-II	0	100
NIQE	Calcul de distance	
<i>wms-EVreg2</i>	0	100
DMOS	0	100

TABLE 2.13 – Échelle d'interprétation des résultats

Comme nous pouvons le voir, les images (c) et (e) ont pratiquement le même PSNR, tandis que visuellement l'image compressée JPEG (c) présente plus de distorsions que l'image (e), ce qui ne correspond pas avec les valeurs *DMOS*. Le même comportement a été constaté pour SSIM et NIQE. La seule mesure avec référence VIF arrive à évaluer la qualité en adéquation avec les scores subjectifs et surpasse tous les autres mesures avec référence.

Nous pouvons également constater que *wms-EVreg2* est bien corrélé avec les scores subjectifs par rapport à BLIINDS-II notamment dans l'évaluation des images (b), (d) et (f). Par exemple, avec un score de 90.3334, *wms-EVreg2* est plus proche du *DMOS* par rapport à BLIINDS-II et très compétitif avec la mesure avec référence VIF. Pour l'image compressée JPEG (c) et le flou Gaussien (e), les scores de *wms-EVreg2* et BLIINDS-II sont corrélés avec les valeurs *DMOS*.

Nous proposons l'étude d'un autre exemple où dans ce cas le score prédit par *wms-EVreg2* ne correspond pas au score subjectif (figure 2.9). Nous remarquons par exemple pour l'image (f) lorsque la dégradation est trop impactante, les trois mesures sans référence à savoir BLIINDS-II, NIQE et *wms-EVreg2* n'arrivent pas à obtenir des résultats corrélés avec le score subjectif.

Nous pouvons également remarquer que *wms-EVreg2* reste plus performant que NIQE qui donne le même score aux images (c) et (d) qui ne correspond pas aux valeurs *DMOS*. Bien que la mesure VIF surpasse toutes les autres mesures de qualité, les résultats *wms-EVreg2* restent concurrentiels. Nous rappelons que *wms-EVreg2* est un algorithme d'évaluation sans référence basé sur un apprentissage, à l'inverse du

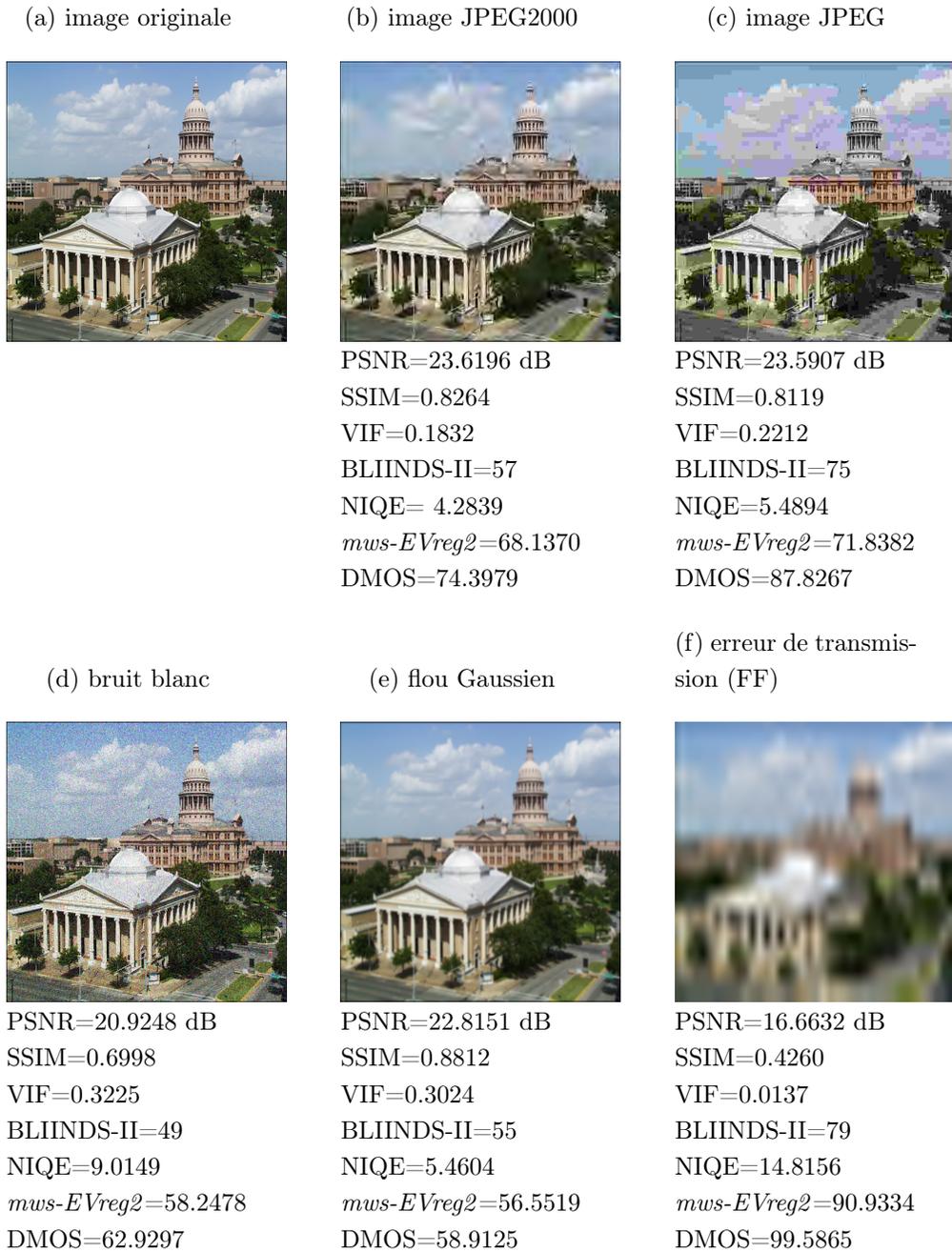


FIGURE 2.8 – Exemples de résultats obtenus sur une image originale ("churchandapitol") et ses cinq versions dégradées de la base d'image LIVE.

VIF qui est une méthode avec référence. Toutefois à partir de ces deux exemples, nous constatons qu'il est difficile d'avoir une analyse au cas par cas.

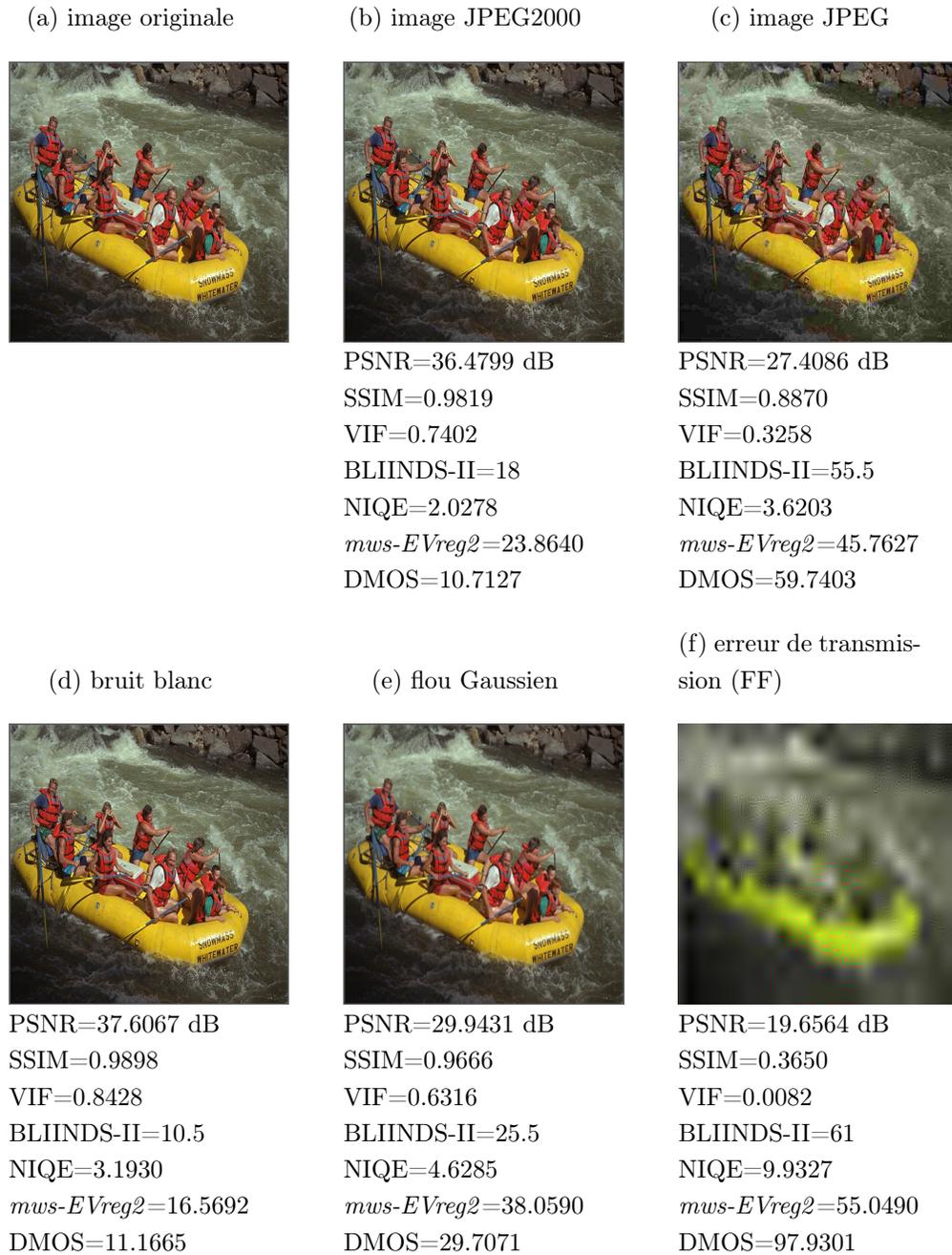


FIGURE 2.9 – Exemples de résultats obtenus sur une image originale ("rapids") et ses cinq versions dégradées de la base d'image LIVE.

Toutes les expérimentations précédentes ont été réalisées sur des images issues de la même base de données. En d'autres termes, $wms-EVreg2$ a été paramétré sur une

partie de la base d'image LIVE et testé sur la partie restante. Pour étudier si les performances de l'algorithme dépendent de la base de données, nous réalisons dans la section suivante un test complet en faisant apprendre *wms-EVreg2* sur la base LIVE et en testant les performances sur une autre base de données indépendante.

2.5.5 La robustesse et l'indépendance de la base

Pour démontrer la robustesse de la mesure proposée et montrer son indépendance par rapport à la base d'images d'apprentissage, nous avons testé la performance de *wms-EVreg2* sur la base d'images TID2013 [72]. La base TID2013 contient 25 images de référence dont 24 images ont été prises de la base de données Kodak et une image a été créée artificiellement (synthétique). 24 types de distorsions avec 5 niveaux pour chacun ont été utilisés pour produire un total de 120 images dégradées pour chaque image de référence, pour enfin obtenir un total de 3000 images dégradées dans la base de données ($24 \times 5 \times 25$).

Nous avons lancé l'apprentissage de *wms-EVreg2* sur la base LIVE et l'avons testé sur TID2013. Sachant que la base de données LIVE ne contient que des images naturelles et 5 types de distorsions, nous avons testé *wms-EVreg2* que sur les 24 images naturelles de TID2013 et en prenant en compte dans un premier temps que les 4 types de distorsions (JPEG2000, JPEG, bruit blanc et le flou Gaussien) qui ont été apprises par *wms-EVreg2* dans la base LIVE.

Les résultats des coefficients de corrélation de Spearman sont présentés dans le tableau 2.14. Nous pouvons remarquer que *wms-EVreg2* surpasse NIQE ainsi que BLIINDS-II dans le cas *aveugle*. En outre, afin de vérifier les performances de notre mesure, la qualité de l'image a été prédite pour chaque type de distorsion, prise en compte dans cette expérience, de la base TID2013. Nous remarquons que la mesure *wms-EVreg2* surpasse NIQE quelque soit la dégradation et elle est plus performante que BLIINDS-II sur le bruit blanc et le flou Gaussien. Enfin, *wms-EVreg2* reste compétitif avec les mesures avec référence populaires telles que PSNR et SSIM.

Nous pouvons soulever un point important dans notre proposition : l'ordonnement de l'évolution de la performance entre les méthodes *ms-EVreg*, *wms-EVreg* et *wms-EVreg2* est conservé quelque soit le type de distorsion.

Il est également important de rappeler que le *wms-EVreg2* est appris sur la base LIVE et testé sur une autre base d'image indépendante. Les résultats obtenus prouvent

que l'approche proposée est robuste aux changements de base de données, et démontre que la performance de *wms-EVreg2* est indépendante de la base d'image d'apprentissage.

	JPEG2000	JPEG	WN	GBLUR	TOUS
PSNR	0.9054	0.9188	0.9531	0.9671	0.9244
SSIM	0.9491	0.9316	0.8762	0.9704	0.9212
VIF	0.9538	0.9289	0.9340	0.9663	0.9336
BLIINDS-II	0.9221	0.8815	0.7560	0.8931	0.8466
NIQE	0.8123	0.7583	0.8104	0.7726	0.7733
ms-EVreg	0.8507	0.8514	0.8436	0.8758	0.8591
wms-EVreg	0.8683	0.8634	0.8436	0.8880	0.8704
wms-EVreg2	0.9044	0.8684	0.8581	0.8974	0.8766

TABLE 2.14 – Les coefficients de corrélation de Spearman sur une partie de la base d'image TID2013 et apprentissage sur LIVE

Nous proposons maintenant d'être encore plus général en intégrant des dégradations non apprises.

2.5.6 Évaluation de *wms-EVreg2* sur toute la base TID2013

Nous proposons maintenant de tester *wms-EVreg2* sur la base d'images TID2013 toute entière en prenant en considération toutes les distorsion présentes dans cette base (donc certaines distorsions ne sont pas présentes dans notre base d'apprentissage). Les résultats sont présentés dans la table 2.15.

- Nous remarquons que notre méthode est plus performante sur 9 des 24 distorsions ainsi sur la base toute entière.
- Parmi ces 9 dégradations, seules deux ont été apprises à savoir le bruit blanc et le flou Gaussien (voir Annexe A pour définir les distorsions).
- Pour JPEG et JPEG2000 (dégradations 10 et 11 : voir Annexe A) les résultats de *wms-EVreg2* et BLIINDS-II sont très proches.
- Les distorsions sur lesquelles NIQE obtiennent de meilleurs résultats représentent principalement le bruit et ses différentes variantes.

Dégradation	BLIINDS-II	NIQE	ms-EVreg	wms-EVreg	wms-EVreg2
Dégradation1	0,7290	0,8104	0,7921	0,8082	0,8263
Dégradation2	0,6569	0,6431	0,6716	0,6799	0,7791
Dégradation3	0,7686	0,4791	0,2137	0,2928	0,4233
Dégradation4	0,5438	0,7728	0,54568	0,5644	0,6135
Dégradation5	0,8249	0,8441	0,8735	0,8907	0,8953
Dégradation6	0,7074	0,7647	0,4798	0,5509	0,6507
Dégradation7	0,7815	0,8391	0,4231	0,5187	0,7022
Dégradation8	0,8625	0,7725	0,8680	0,8776	0,8843
Dégradation9	0,7157	0,6106	0,6656	0,6828	0,7391
Dégradation10	0,8664	0,7583	0,8359	0,8515	0,8582
Dégradation11	0,8983	0,8123	0,8467	0,8645	0,8973
Dégradation12	0,2835	0,1880	0,2572	0,2675	0,2804
Dégradation13	0,6274	0,4333	0,2358	0,3313	0,4523
Dégradation14	0,1989	0,1911	0,2352	0,1793	0,1647
Dégradation15	0,2535	0,2393	0,1245	0,1162	0,2809
Dégradation16	0,1786	0,2469	0,2182	0,1513	0,1504
Dégradation17	0,2666	0,1468	0,2410	0,2575	0,3111
Dégradation18	0,1754	0,3623	0,2961	0,2360	0,2687
Dégradation19	0,7237	0,6862	0,5478	0,5851	0,6705
Dégradation20	0,1093	0,3529	0,3172	0,3413	0,2523
Dégradation21	0,7285	0,7894	0,5929	0,5715	0,6299
Dégradation22	0,8181	0,8960	0,6987	0,7566	0,7980
Dégradation23	0,6303	0,5593	0,6819	0,7248	0,7690
Dégradation24	0,8164	0,7965	0,8436	0,8659	0,8865
TOUS	0,3807	0,2968	0,3340	0,3574	0,3931

TABLE 2.15 – Les coefficients de corrélation de Spearman sur la base d'image TID2013 et apprentissage sur LIVE

— *wms-EVreg2* a un pouvoir de généralisation plus grand que les autres méthodes.

Puisque que la base de d'images LIVE possède 18 images en commun avec la base d'images TID2013, et dans le but de vérifier la robustesse de *wms-EVreg2*, une nouvelle expérience a été réalisée : nous avons paramétré notre algorithme à partir

de la base de données LIVE, et validé sur la base d'image CSIQ [36]. Cette base est composée de 30 images de référence obtenues à partir d'une source de domaine public. Les images ont été choisies pour couvrir cinq catégories : animaux, paysages, personnes, plantes et urbain. Les distorsions utilisées dans CSIQ sont : compression JPEG, compression JPEG2000, décalage de contraste global (CONTRAST), bruit Gaussien rose (PN), bruit Gaussien blanc (WN) et le flou Gaussien (GBLUR). Au total, 866 images dégradées ont été générées.

L'expérience a été réalisée en considérant toutes les distorsions même celles non apprises dans la base LIVE. Les résultats de la comparaison sont présentés dans le tableau 2.16.

	JPEG2000	JPEG	WN	GBLUR	PN	CONTRAST	TOUS
PSNR	0.8499	0.8305	0.8706	0.7290	0.8002	0.8157	0.8431
SSIM	0.9092	0.8707	0.9067	0.8298	0.9067	0.8544	0.8756
VIF	0.9371	0.9503	0.9575	0.9744	0.9425	0.9362	0.9387
BLIINDS-II	0.8946	0.9070	0.8017	0.8914	0.3788	0.0122	0.6121
NIQE	0.8296	0.7769	0.7674	0.8609	0.1546	0.5344	0.6787
ms-EVreg	0,7944	0,8571	0,8734	0,8486	0,4992	0,1348	0,6014
wms-EVreg	0,8141	0,8665	0,8876	0,8542	0,4725	0,0993	0,6097
wms-EVreg2	0,8358	0,8816	0,9044	0,8656	0,4303	0,1314	0,6148

TABLE 2.16 – Les coefficients de corrélation de Spearman sur une partie de la base d'image CSIQ

Nous constatons qu'en dépit d'un déclin des résultats, le *wms-EVreg2* est concurrentiel avec BLIINDS-II sur toutes les distorsions à l'exception du bruit blanc où *wms-EVreg2* dépasse largement BLIINDS-II. NIQE obtient de meilleur résultat sur le contraste et sur la base CSIQ.

2.6 Conclusion

Dans ce chapitre, nous avons proposé une nouvelle mesure sans référence d'évaluation de la qualité d'image capable d'estimer la qualité des images dégradées avec de multiple types de distorsion. Cette approche appelée *wms-EVreg2* est basée sur la fusion de différentes caractéristiques statistiques, extraites de l'image, en utilisant la théorie des fonctions de croyance. Nous avons proposé une première version *EVreg* qui

prédit le score final de qualité des images en prenant en compte chaque vecteur de caractéristiques séparément. Pour évaluer cette stratégie, les résultats obtenus ont été comparés avec deux autres méthodes de régression : GRNN et SVM.

Les résultats sont juste satisfaisants en ce qui concerne la corrélation avec les scores subjectifs. Afin d'améliorer ces résultats, nous proposons une version multi source qui permet de fusionner les masses de croyance issues des différents vecteurs de caractéristiques pour prédire le score de la qualité des images. Nous avons proposé une fusion adaptée par le biais de l'utilisation d'un coefficient d'affaiblissement (fiabilité) des masses de croyance. Trois stratégies d'affaiblissement sont proposées : la première stratégie consiste à pondérer de façon équivalente les informations portées par les différentes caractéristiques des images ; la valeur du coefficient d'affaiblissement utilisée est de 0,95 quelque soit le vecteur de caractéristiques. La deuxième stratégie considère les coefficients de corrélation de Spearman obtenus à partir des premières expériences comme coefficients d'affaiblissement. Enfin, la troisième stratégie utilise la matrice de confusion issue de la classification des images en fonction de leur type de dégradation.

À partir des différentes expérimentations, nous avons constaté que notre troisième stratégie où les coefficients d'affaiblissement sont issus de la matrice de confusion est celle qui nous fournit les meilleures performances. Les scores prédits ont une bonne corrélation avec les scores de qualité subjectifs et fournit des performances de prédiction de qualité compétitives par rapport aux mesures avec référence. Nous avons aussi montré (en testant notre approche proposée sur une partie des bases d'images TID 2013 et CSIQ) que *wms-EVreg2* est robuste aux changements de base d'images et prouve son indépendance vis-à-vis du contenu de la base. À partir de cette dernière expérimentation, nous pouvons constater que *wms-EVreg2* se corrèle bien avec les scores subjectifs et surpasse les méthodes de références proposées dans la littérature.

Chapitre 3

Détection d'une information cachée dans une image

Sommaire

3.1	Introduction	80
3.2	Le contexte de la stéganographie	81
3.2.1	La stéganographie au cours des siècles	82
3.2.2	La stéganographie de nos jours	83
3.3	Stéganographie dans les images	84
3.3.1	Propriétés d'un algorithme de stéganographie	86
3.3.2	Stéganographie dans le domaine spatial	88
3.3.3	Stéganographie dans un domaine transformé	91
3.4	La stéganalyse, détection de stéganographie	94
3.4.1	Les différentes classes de stéganalyse	95
3.4.2	Différents schémas de la stéganalyse	96
3.5	Un schéma de stéganalyse basé sur la théorie des fonctions de croyance	102
3.5.1	Positionnement général	102
3.5.2	Espace de caractéristiques pour la stéganalyse	104
3.5.3	Ev-Knn pour la stéganalyse	105
3.5.4	Discussion sur les espaces de caractéristiques	110
3.5.5	<i>EVEC-Knn pour la stéganalyse</i>	111
3.6	Évaluation de la méthode proposée	113
3.6.1	Paramétrage de la méthode : \mathcal{D}_{red} et R	113

3.6.2	Analyse sur nsF5	115
3.6.3	Introduction d'une classe de rejet	117
3.6.4	Analyse sur <i>J-UNIWARD</i>	118
3.6.5	Capacité de généralisation	120
3.6.6	Évaluation dans le domaine spatial sur HUGO et <i>S-UNIWARD</i>	121
3.7	Conclusion	123

3.1 Introduction

La stéganographie est l'art de la communication secrète. L'objectif est de dissimuler un message secret dans un médium anodin (une image, une vidéo, un son, etc.) de sorte qu'il ne puisse pas être détecté. Avec l'avènement de l'internet à partir des années 90, la stéganographie a pris de l'ampleur dans les supports numériques - privilégiés pour la transmission d'informations - ce qui a donné naissance à la stéganographie numérique, qui est devenue aujourd'hui une pratique populaire et accessible à toute personne souhaitant communiquer de façon discrète.

Comme toute autre technologie, la stéganographie peut être utilisée à des fins malveillantes. La nécessité et le besoin de détecter des objets stéganographiés donnent lieu à la stéganalyse, le processus dual de la stéganographie.

Avec l'avancement de la stéganographie des images, de nombreuses méthodes de stéganalyse ont été développées pour faire face à ces nouvelles technologies de dissimulation de messages secrets. Dans les premiers temps, certaines informations préalables sur les algorithmes stéganographiques ont été supposés disponibles. Il s'agit de la stéganalyse ciblée. Toutefois, les recherches se sont orientées ces dernières années vers des situations dans lesquelles aucune information sur les algorithmes stéganographiques n'est disponible. C'est le cas la stéganalyse aveugle, qui tente de différencier les images *stégo* des images de couverture sans la connaissance des algorithmes stéganographiques.

Dans ce chapitre, nous proposons un schéma de stéganalyse basé sur la théorie des fonctions de croyance construit sur des sous-espaces aléatoires des caractéristiques. Nous allons tout d'abord présenter globalement le concept de la stéganographie. Nous présentons les principales méthodes d'insertion de messages cachés dans des médias numériques. Nous dressons ensuite un bref état de l'art des méthodes de stéganalyse proposées dans la littérature. Ensuite, nous proposons un schéma de stéganalyse

aveugle basé sur la théorie des fonctions de croyance. Enfin, différents tests et résultats expérimentaux de ce schéma seront présentés et commentés avant de conclure ce chapitre.

3.2 Le contexte de la stéganographie

Bien que l'idée de la stéganographie remonte aux périodes antiques, ce n'est que récemment que le nom réel a été utilisé pour la première fois par Johannes Trithemius (1462 à 1516), en *Steganographia* (figure 3.1). *Steganographia* est considéré comme l'un des tous premiers travaux sur la cryptographie et la stéganographie, dans lequel les différentes techniques de cryptographie et de stéganographie primitives sont détaillées. Le terme de stéganographie lui-même était créé des mots grecs "steganos" - couverts - et "graphia" - écrit.

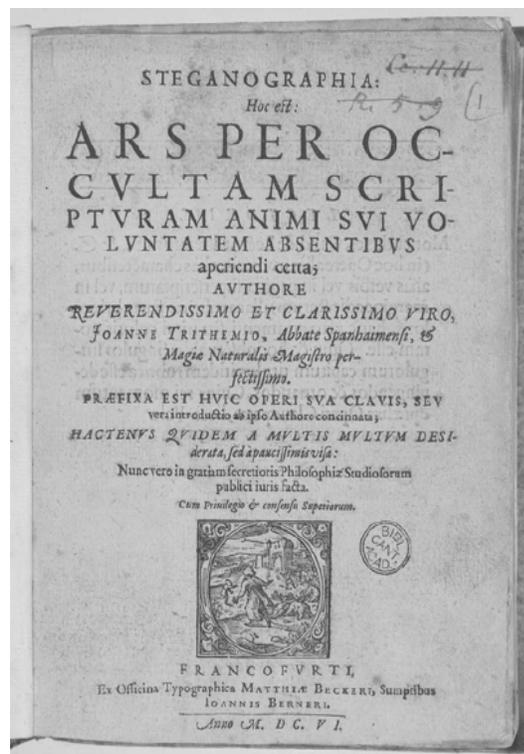


FIGURE 3.1 – La page de garde de *Steganographia* de Johannes Trithemius, l'inventeur du mot "stéganographie"

Il est important de noter la différence entre la stéganographie et la cryptographie. La cryptographie vise à modifier le message de sorte qu'il devienne impossible de le

lire. En cryptographie, il n'est pas important que le message chiffré puisse attirer des soupçons. En stéganographie, l'existence du message est secrète : le canal de communication est considéré comme ouvert et le message lui-même n'est généralement pas modifié (même s'il peut être chiffré). Le principal objectif est donc de cacher le message aussi bien que possible dans un contenu anodin, de sorte que tout intermédiaire n'aurait pas de soupçons quant à sa présence. La figure 3.2 illustre un cas de stéganographie (par substitution de LSB, voir la section 2.4.2) pour laquelle la dissimulation du message est invisible à l'œil humain.

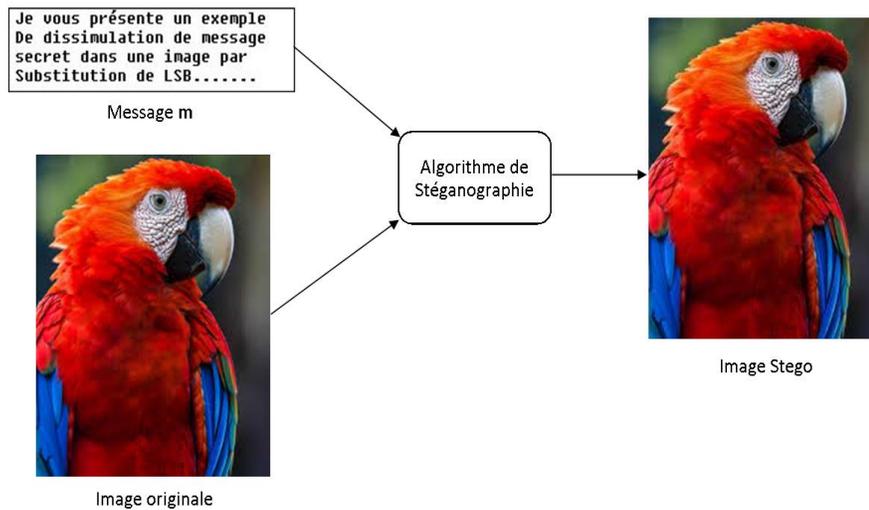


FIGURE 3.2 – Exemple de dissimulation d'un message secret par l'algorithme de substitution LSB

3.2.1 La stéganographie au cours des siècles

Depuis que la communication existe, la nécessité du secret dans cette communication a été présente, que ce soit à des fins de confidentialité bénigne ou malicieuse. Certains des exemples les plus anciens connus sont liés à Hérodote dans son oeuvre intitulé l'enquête [79], qui raconte l'histoire d'Histiaeus, conseiller du roi de Perse, qui incita par le moyen de la stéganographie à son neveu Aristagoras à provoquer une révolte dans le Lonnia. Le message a été transmis par l'intermédiaire d'un esclave de Histiaeus dont la tête avait d'abord été rasée et tatouée avec le contenu du message. Une fois que les cheveux aient repoussé suffisamment pour couvrir le message, l'esclave a pu être envoyé sans aucun risque vers Aristagoras, qui n'avait qu'à raser la tête de l'esclave pour lire le message.

Au fil des siècles, différentes techniques et formes de stéganographie, de plus en plus évoluées, ont été utilisées dans le monde.

En chine antique, des messages secrets étaient écrits sur de rubans de soie puis enrobés dans des petites boules de cire. Ces boules étaient avalées par le messenger, pour arriver d'une manière totalement discrète, jusqu'au destinataire.

L'une des techniques les plus utilisées, dès le 1^{er} siècle avant J.-C, est l'encre sympathique, ou l'encre invisible. Il s'agit d'un procédé chimique utilisé pour écrire le message secret qui sera invisible à l'oeil nu. Pour révéler le message, un simple passage sous une source de chaleur est nécessaire.

Plus récemment, un exemple intéressant est celui du premier ministre du Royaume-Uni, Margaret Thatcher, qui a vu à plusieurs reprises des documents de son propre cabinet rendus publics par la voie de la presse. Afin de révéler l'identité de l'auteur, elle avait intégré un petit message (différent pour chaque copie) dans les documents remis à chaque ministre. Le message a été incorporé dans les propriétés d'espacement des documents (espacements entre les mots et les lignes) ce qui a permis d'identifier une fuite potentielle au sein du gouvernement [80]. Cette forme de dissimulation de message est appelée stéganographie linguistique.

Malheureusement, tous ces exemples reposent sur des sources «non officielles» et il n'y a pas de preuves concrètes d'une telle stéganographie. De toute évidence, les meilleurs cas de stéganographie ne seront peut-être jamais détectés ou connus...

3.2.2 La stéganographie de nos jours

Au cours des dernières années, l'intérêt pour la stéganographie a considérablement augmenté, comme le montre par exemple le nombre de publications liées aux sujets de la stéganographie et aux watermarking tels que rapporté par l'IEEE [81].

Les deux journaux USA Today (USA) et Corriere della Sera (Italie) ont suggéré l'utilisation des moyens de stéganographie (en utilisant des images comme support) pour les communications d'Al-Qaïda. Plusieurs acteurs des médias prétendent même que les attentats du 11 septembre 2001, auraient été organisés en utilisant des messages secrets cachés au sein d'images numériques à caractère pornographique [82][83]. Mais, aucune preuve réelle ou article prouvant ces allégations n'est apparue.

Ainsi, les instituts de recherche et les gouvernements étant inquiets sur l'utilisation

de la stéganographie ont eu un intérêt croissant pour 1) la maîtrise et la compréhension de ce processus de dissimulation, 2) le développement d'outils performants permettant de dissuader les usages malveillants, ce qui a donné naissance à la stéganalyse qui tente de détecter la présence des informations cachées. La découverte d'une section sur la stéganographie et de la communication secrète dans le "Techniques du Mujahid", un manuel de formation pour les djihadistes, a remis en lumière ces questions du possible terrorisme, ce qui peut expliquer en partie cet intérêt.

Dans tous les cas, l'existence de la stéganographie et les possibilités qu'elle ouvre en font un véritable problème lié à la sécurité qui mérite d'être étudié.

La croissance du nombre d'objets numériques accessibles au public a fait de la stéganographie numérique une tâche plus facile, dans le sens où il est devenu difficile pour les autorités de contrôler tout le contenu numérique qui est échangé. En effet, nous pouvons assez facilement trouver des algorithmes de stéganographie pour les médias tels que le texte , des sons et des images.

3.3 Stéganographie dans les images

Les supports numériques actuels peuvent être exploités pour transmettre des messages secrets via les techniques de la stéganographie. Les images et les sons en particulier, car ils sont souvent compressés avec perte.

Les images sont considérées comme les supports numériques les plus communément échangés sur Internet. De ce fait, la grande majorité des outils de stéganographie est spécifiquement dédiée à la dissimulation d'informations dans les fichiers d'images. Dans le reste de ce manuscrit, notre travail se focalise sur la stéganographie dans les images numériques.

La formulation de la stéganographie moderne peut s'illustrer par le scénario de deux prisonniers [84]. Alice et Bob sont deux prisonniers qui travaillent sur un plan de leur évasion. Ils sont autorisés à communiquer, mais leurs communications passent par le gardien, Eve. Eve est chargée de la surveillance des échanges entre Alice et Bob ; si elle suspecte le moindre signe de conspiration entre eux, elle mettra fin à leurs échanges. L'incorporation du message secret dépend d'une clé de stéganographie partagée uniquement par Alice et Bob.

Prenant compte de la situation, Alice et Bob doivent utiliser une technique pour

dissimuler leurs messages afin d'éviter d'éveiller les soupçons d'Eve. Plus précisément, si Alice souhaite envoyer un message m à Bob, elle l'incorpore dans un média de couverture c pour obtenir un objet stégo s , qui ensuite sera envoyé à Bob à travers la gardienne Eve comme illustré dans la figure 3.3.

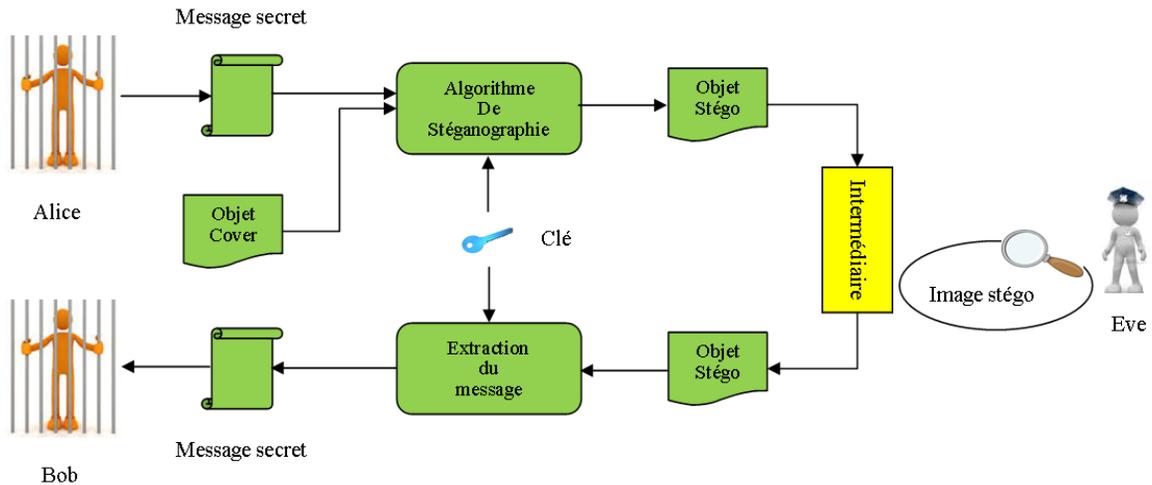


FIGURE 3.3 – Scénario de Stéganographie / Stéganalyse présenté par le problème des prisonniers

À partir de ce scénario, nous pouvons distinguer deux parties :

1. Alice et Bob qui représentent les stéganographes,
2. La gardienne Eve qui représente la stéganalyste, qui peut examiner l'objet intercepté de façon *passive*, *active* ou *malicieuse*.

Une inspection active consiste à altérer suffisamment les objets inspectés afin de rendre impossible la lecture d'un éventuel message caché. La stéganalyse passive consiste simplement à détecter la présence d'un message caché. C'est le scénario typique sur lequel plusieurs travaux de recherches se focalisent principalement. Enfin la stéganalyse malicieuse consiste à extraire le message caché et, éventuellement, à réintroduire un message falsifié.

Dans ce manuscrit, nous nous intéressons à la stéganalyse passive, dont l'objectif est de détecter si une image donnée est une image de couverture c'est-à-dire exempte de message caché ou une image *stégo* présentant un message caché.

Alice et Bob travaillent avec l'ensemble de toutes les images de couverture (*cover*) possibles et les ensembles de clés et des messages qui dépendront, dans le cas le plus général, de chaque image *cover* : soit \mathcal{C} l'ensemble des objets *cover* $c \in \mathcal{C}$, \mathcal{M} est

l'ensemble de tous les messages qui peuvent être communiqués dans c et \mathcal{K} l'ensemble des clés possibles.

Suivant le schéma de la figure 3.3, un schéma de stéganographie est une paire de fonctions d'insertion et d'extraction Emb et Ext (respectivement relatives à *Embedding* et *Extraction*),

$$Emb : \mathcal{C} \times \mathcal{K} \times \mathcal{M} \longrightarrow \mathcal{C}, \quad (3.1)$$

$$Ext : \mathcal{C} \times \mathcal{K} \longrightarrow \mathcal{M}, \quad (3.2)$$

telle que pour tout $\mathbf{c} \in \mathcal{C}$, et pour tout $\mathbf{k} \in \mathcal{K}$, $\mathbf{m} \in \mathcal{M}$,

$$Ext(Emb(\mathbf{c}, \mathbf{k}, \mathbf{m}), \mathbf{k}) = \mathbf{m}. \quad (3.3)$$

En d'autres termes, Alice peut prendre un objet *cover* $\mathbf{c} \in \mathcal{C}$ et cacher n'importe quel message $\mathbf{m} \in \mathcal{M}$ en utilisant n'importe quelle clé $\mathbf{k} \in \mathcal{K}$ (échangée au préalable avec Bob), pour obtenir enfin une image *stégo* $\mathbf{s} = Emb(\mathbf{c}, \mathbf{k}, \mathbf{m})$.

Dans les travaux de recherche présentés dans ce chapitre, nous nous intéressons principalement aux images numériques naturelles, c'est-à-dire acquises à l'aide d'un appareil photographique ou scanner, comme support d'acquisition. Nous rappelons que le but de notre travail n'est pas d'extraire le message \mathbf{m} , mais de détecter la présence ou non de ce message dans une image numérique.

Avant d'entrer dans les détails concernant les techniques de stéganographie usuelles, il est nécessaire de définir quelques propriétés caractérisant un schéma stéganographique.

3.3.1 Propriétés d'un algorithme de stéganographie

Capacité d'insertion

La capacité d'insertion (ou payload) est le nombre de bits significatifs que le stéganographe peut dissimuler dans un support numérique [85].

La capacité d'insertion est destinée à mesurer un rapport entre la quantité de données dissimulées et les propriétés de l'image *cover*. Par conséquent, elle est fortement dépendante de l'image considérée ; le nombre de messages qui peut être échangé sur une image de couverture dépend du schéma stéganographique et de l'image *cover* elle-même.

Par exemple, si \mathcal{C} est l'ensemble de toutes les images en niveau de gris de taille 512×512 , et Alice dissimule un message bit par pixel, alors $\mathcal{M} = \{0, 1\}^{512 \times 512}$ et $|\mathcal{M}(c)| = 2^{512 \times 512}$ pour tout $\mathbf{c} \in \mathcal{C}$. Pour les images compressées JPEG, Alice cache un message bit par coefficient DCT quantifié non nul, donc le nombre de messages qui peut être dissimulé dépend de l'image *cover* elle-même, car le nombre de coefficients DCT quantifiés non nuls dans les images JPEG dépend du contenu. Cela implique une capacité d'insertion différente d'une image à une autre.

Cette capacité d'insertion est définie en bits pour un objet de couverture \mathbf{c} par :

$$payload = \log_2 |\mathcal{M}(\mathbf{c})| \quad (3.4)$$

et la capacité d'insertion relative est définie par :

$$payload_r = \frac{\log_2 |\mathcal{M}(\mathbf{c})|}{n} \quad (3.5)$$

où n est le nombre d'éléments dans \mathbf{c} , tels que le nombre de pixels ou de coefficients DCT non nuls.

Capacité stéganographique

La capacité stéganographique peut être empiriquement décrite comme étant nombre maximal de bits qui peuvent être insérés dans l'image *cover* tout en satisfaisant un ensemble de contraintes sur la distorsion et la probabilité de détection. La capacité stéganographique reste une quantité qui est difficile à estimer ; elle est strictement plus petite que la capacité d'insertion [86] [85].

Sécurité d'un schéma stéganographique

La notion de sécurité ou d'indétectabilité d'un schéma stéganographique se réfère à la question de minimisation de la quantité de distorsions due à la dissimulation d'un message. Elle mesure la capacité ou l'incapacité de la gardienne Eve (stéganalyste) à détecter la présence du message secret au sein de l'image *cover*.

La notion de sécurité en stéganographie a été définie par Cachin dans [87] comme étant l'entropie qui existe entre deux distributions $P_{\mathbf{c}}$ pour un objet *cover* et $P_{\mathbf{s}}$ pour un objet *stégo*. Supposons que la distribution de probabilité $P_{\mathbf{c}}$ d'un objet de couverture $\mathbf{c} \in \mathcal{C}$ existe. Si la clé secrète $\mathbf{k} \in \mathcal{K}$ et le message $\mathbf{m} \in \mathcal{M}$ sont choisis aléatoirement (uniformément), alors le schéma stéganographique définit la distribution de probabilité $P_{\mathbf{s}}$ des objets *stégo* $\mathbf{s} \in \mathcal{C}$.

La différence entre les distributions $P_{\mathbf{c}}$ et $P_{\mathbf{s}}$ est mesurée en utilisant une mesure appelée divergence de Kullbak-Liebler (D_{KL}) et définie par :

$$D_{KL}(P_{\mathbf{c}}, P_{\mathbf{s}}) = \sum_{\mathbf{c} \in \mathcal{C}} P_{\mathbf{c}}(\mathbf{c}) \log_2 \frac{P_{\mathbf{c}}(\mathbf{c})}{P_{\mathbf{s}}(\mathbf{c})} \quad (3.6)$$

Avec cette formulation, Cachin définit un algorithme de stéganographie S comme étant ε -sûr si $D_{KL}(P_{\mathbf{c}}, P_{\mathbf{s}}) \leq \varepsilon$. Par conséquent, plus ε est petit, plus les deux distributions $P_{\mathbf{c}}$, $P_{\mathbf{s}}$ sont plus proches, et plus il est difficile de distinguer une image *stégo* d'une autre *cover*.

Dans le cas où $\varepsilon = 0$, l'algorithme de stéganographie est considéré comme sûr, car aucune différence ne peut être faite entre l'image de couverture et l'image *stégo*.

Théoriquement, cette définition de la sécurité suppose que l'on sache définir précisément ce que c'est une distribution, et nécessite l'estimation de $P_{\mathbf{c}}$ et $P_{\mathbf{s}}$, mais du point de vue pratique, elle est difficile à utiliser. L'espace de toutes les images \mathcal{C} est trop grand (potentiellement infini), ce qui rend l'équation $D_{KL}(P_{\mathbf{c}}, P_{\mathbf{s}})$ difficile à calculer.

Plusieurs modèles de sécurité ont été proposés dans la littérature pour évaluer la sécurité d'un schéma stéganographique. Parmi ces modèles, ceux qui sont consacrés aux schémas à clés privées, d'autres aux schémas à clés publiques [88].

Dans la prochaine section, un état de l'art des algorithmes de stéganographie les plus utilisés est présenté. Il n'est absolument pas exhaustif et vise simplement à présenter les concepts généraux de ces algorithmes.

3.3.2 Stéganographie dans le domaine spatial

La stéganographie dans le domaine spatial concerne les images fixes non compressées qui peuvent être en niveaux de gris ou en couleur (RVB). Pour chaque canal de couleur, les valeurs des bits de chacun des pixels (codés sur b bits) n'ont pas tous la même importance ; en effet le premier bit est pondéré par $2^0 = 1$ (appelé bit de poids

faible), tandis que le dernier bit est pondéré par 2^{b-1} (bit de poids fort). Les premiers algorithmes stéganographiques de l'état de l'art, présentés par la suite, sont issus de cette propriété.

Dans la figure 3.4, nous illustrons les différents plans de bit d'une image en niveau de gris du bit de poids faible LSB (Least Significant Bit) jusqu'aux bit de poids fort (Most Significant Bit). Nous remarquons que les plans des bits de poids faibles sont nettement moins structurés que ceux de poids plus forts, donc la substitution des bits de poids faible des pixels sera moins perceptible compte-tenu de la sensibilité de la vision humaine.

Stéganographie par substitution de LSB

La stéganographie par substitution de LSB (LSB substitution), est l'une des premières techniques de la littérature qui permet de dissimuler un message secret dans une image numérique [89]. La simplicité d'implémentation de cette technique lui a permis d'être la plus utilisée sur internet. Comme son nom l'indique, cette méthode permet d'insérer le message en remplaçant les bits de poids faibles des pixels d'une image directement¹, soit d'une manière complètement séquentielle, ou en choisissant pseudo-aléatoirement les pixels à modifier, ou encore par une clé stéganographique qui indique le sens de parcours des pixels de l'image.

Autrement dit, pour insérer un message $\mathbf{m} = (\mathbf{m}_1, \dots, \mathbf{m}_m)$, le dernier bit de poids faible de chaque pixel de l'image est remplacé par un bit du message à cacher.

Stéganographie par correspondance de LSB (LSB Matching)

Dans le but de réduire l'impact de l'insertion d'informations cachées, de nombreuses améliorations de la stéganographie par substitution LSB ont été proposées. L'amélioration la plus courante et la plus étudiée est la méthode d'insertion par correspondance des LSB appelée aussi *LSB Matching* ou encore \pm *embedding* qui a été proposée pour la première fois dans [90]. Elle fonctionne de la même manière que la stéganographie par substitution LSB. La principale évolution est que si la valeur du LSB du pixel considéré ne correspond pas à celle qui doit y être insérée, la valeur de ce pixel sera incrémentée ou décrétementée de façon aléatoire. Ici encore, le sens de parcours des pixels de l'image doit être préalablement échangé entre l'émetteur et le récepteur de l'image

1. Il faut tout d'abord séparer les trois composantes RVB de l'image

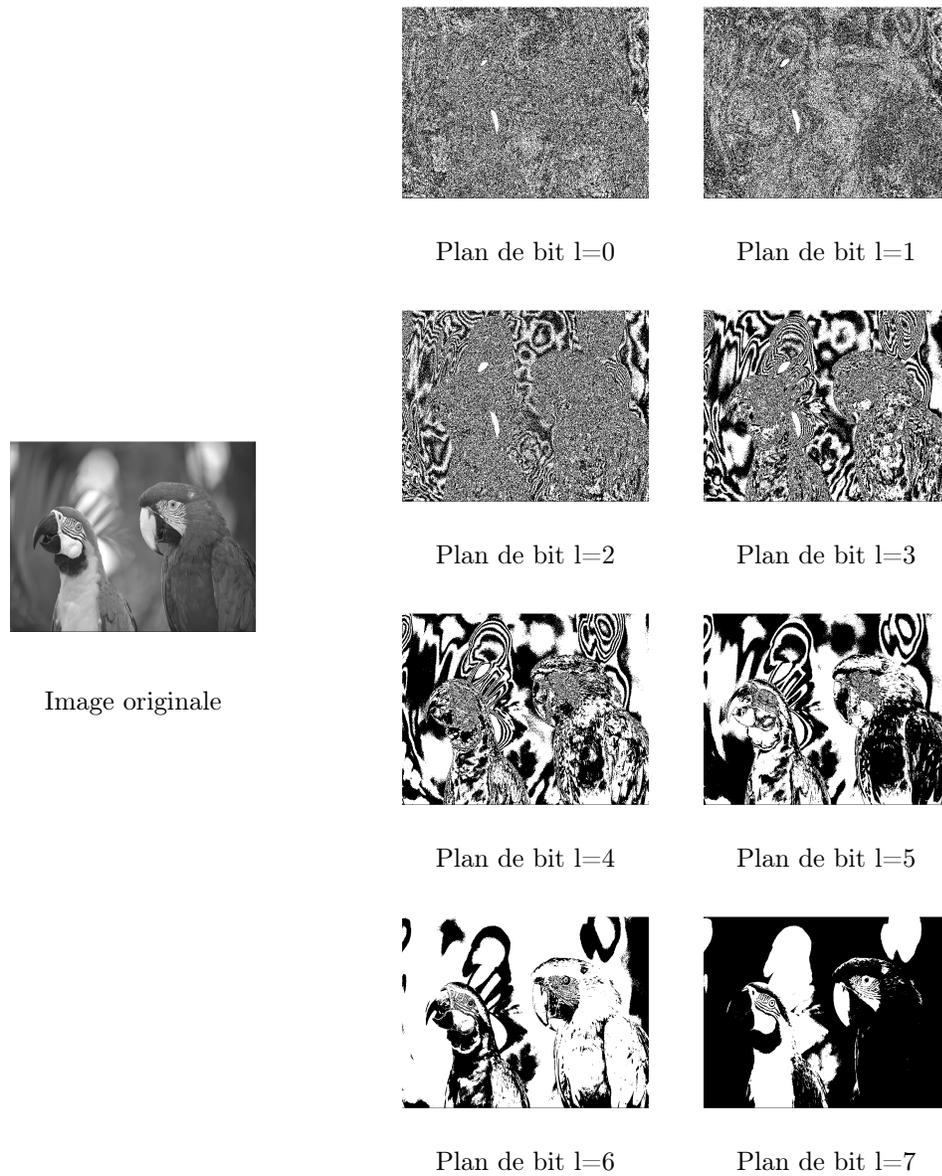


FIGURE 3.4 – Illustration des différents plans de bit d'une image en niveau de gris

stéganographiée.

L'algorithme HUGO

L'algorithme HUGO (Highly Undetectable SteGO), est l'un des systèmes les plus avancés de stéganographie. Il dissimule les messages dans les LSB des images en niveaux de gris représentées dans le domaine spatial. Proposé par Pevný et al.[91] lors

d'une compétition BOSS (Break Our Steganographic System) [92], il fait partie des méthodes dites adaptatives qui reposent sur la minimisation d'impact d'insertion dans les images, autrement dit le message secret est dissimulé dans les zones de l'image de couverture qui sont difficilement détectables afin de conserver le plus possible la statistique de l'image.

Cette stratégie permet de décomposer sa conception en deux parties. La première partie consiste concevoir un modèle d'image qui permet de générer un espace de caractéristiques dans lequel la distance entre les pixels conduit à une bonne fonction de distorsion $\mathcal{D}(\mathbf{c}, \mathbf{s})$ entre une image de couverture \mathbf{c} et une image *stégo* \mathbf{s} qui lui est associée. Cette fonction est ensuite utilisée dans la deuxième partie de l'algorithme, appelée le codeur, qui détermine les pixels de l'image de couverture qui doivent être modifiés pour communiquer le message. De plus, le codeur optimal minimise la distorsion moyenne calculée sur différents messages de même longueur.

Le modèle d'image a été largement inspiré par l'espace des caractéristiques appelé SPAM (SPAM Subtractive Pixel Adjacency Matrix) [93], qui permet de modéliser les dépendances entre pixels voisins en utilisant les chaînes de Markov.

Parmi les différents schémas stéganographiques que nous avons présentés, nous constatons que les méthodes adaptatives, celles qui reposent sur le principe de minimisation d'impact d'insertion, sont théoriquement les plus efficaces. La grande majorité des algorithmes stéganographiques est actuellement basée sur l'utilisation d'une distorsion (également appelée coût de détectabilité) calculée pour chaque pixel, et qui est combinée à un schéma de codage qui minimise la distorsion globale de l'insertion pour un *payload* donné.

3.3.3 Stéganographie dans un domaine transformé

À ce stade, nous avons évoqué les schémas stéganographiques dans le domaine spatial. Les images les plus souvent échangées sur internet sont des images compressées avec perte telles que JPEG ou JPEG 2000. Ces techniques de compression reposent sur une transformée de l'image ; le JPEG exploite la transformée en cosinus discrète (DCT), et le JPEG2000 la transformée en ondelettes discrète.

Compte-tenu de l'utilisation importante de ce type d'images, plusieurs schémas d'insertion d'informations secrètes adaptés à ce genre d'images ont vu le jour.

La plus grande majorité des algorithmes stéganographiques qui opèrent dans le domaine transformé sont dérivés des méthodes stéganographiques du domaine spatial décrites auparavant. Parmi les méthodes stéganographiques de référence pour les images JPEG, nous pouvons citer Outguess [94], F5 [95] et Jsteg [96] dont le principe est l'insertion des bits du message secret par remplacement des LSB appliquée aux coefficients DCT quantifiés et non plus directement aux valeurs des pixels.

Jsteg

L'algorithme JSteg proposé par Upham [96] est basé aussi sur la modification des LSB des coefficients DCT. Il fait parti des premiers algorithmes de stéganographie et intègre tout simplement les bits du message directement dans les LSB qui sont différents de 0 et 1. Au début, les LSB utilisés pour les modifications ont été choisis de façon séquentielle, puis un chemin pseudo-aléatoire a été utilisé.

L'algorithme F5

L'algorithme F5 proposé par Westfeld dans [95] est plus robuste rapport à ses prédécesseurs, F3 et F4, ainsi il a une capacité d'insertion beaucoup plus élevée. Le principe de l'algorithme F5 est qu'il ne modifie jamais directement le LSB des coefficients DCT, mais décrémente leurs valeurs (sauf ceux qui sont déjà à zéro). Une matrice de codage est utilisée afin de réduire le nombre total de modifications à effectuer sur l'objet de couverture. Dans cet algorithme, la décrémentation des LSB créant une valeur nulle conduit à un effet de rétrécissement (*Shrinkage effect*). Comme l'algorithme n'utilise pas les coefficients DCT nuls, l'information doit être insérée à nouveau dans un autre LSB. Cet effet crée des changements détectables dans les histogrammes des coefficients DCT.

L'algorithme nsF5

L'algorithme nsF5 (pour *no-shrinkage* F5) proposé par Fridrich et al [97], est un dérivé de l'algorithme F5, pour corriger certains de ces inconvénients notamment l'effet de rétrécissement. Les auteurs proposent d'utiliser la technique du papier mouillé [98]. Supposons que le message à insérer contient m bits $\mathbf{m} = (\mathbf{m}_1, \dots, \mathbf{m}_m)$, et qu'il existe n LSB des coefficients DCT (qui ont une valeur 0 ou 1), dont n_u sont les LSB utilisables

pour la modification (c'est-à-dire coefficients DCT non nuls). Le problème est alors de trouver le vecteur des LSB modifiés \mathbf{b} tels que le système

$$\mathbf{D}\mathbf{b}=\mathbf{m} \tag{3.7}$$

soit satisfait, avec \mathbf{D} une matrice binaire pseudo-aléatoire partagée entre l'émetteur et le récepteur (via une clé secrète).

UNIWARD

L'algorithme de stéganographie UNIWARD proposé par Vojtěch [99] est un algorithme de dissimulation universelle indépendant du domaine d'insertion. La distorsion due à l'insertion est calculée dans le domaine des ondelettes en tant que la somme des variations des coefficients des ondelettes relatives dans les sous-bandes de haute fréquence. La directionnalité des fonctions de base d'ondelettes permet à l'expéditeur d'évaluer le voisinage de chaque pixel pour la présence de discontinuités dans de multiples directions (textures et régions "bruyantes") et ainsi éviter de faire intégrer des changements dans les parties de l'image qui peuvent être modélisés suivant au moins une direction (contours propres et régions lisses). Cette approche a été mis en oeuvre dans le domaine spatial (S-UNIWARD), ainsi que dans le domaine de transformé JPEG (J-UNIWARD).

Nous avons présenté quelques concepts et notations utilisés dans le domaine de la stéganographie, ainsi que les principales méthodes de dissimulation d'information cachées. Actuellement, l'approche la plus réussie de la stéganographie dans les médias numériques est d'incorporer le message tout en minimisant une fonction de distorsion convenablement définie. La conception de la distorsion est essentiellement la seule tâche laissée au stéganographe, puisqu'il existe des codes pratiques efficaces qui permettant d'insérer des messages. L'objectif du stéganographe est de concevoir la fonction de distorsion pour obtenir un schéma empirique avec un haut niveau de détectabilité statistique.

Dans la section suivante, nous présenterons le processus consistant à détecter la présence d'un message et distinguer si une image est *cover* ou *stégo*. Ce processus est appelé stéganalyse et est décrit plus en détail dans ce qui suit.

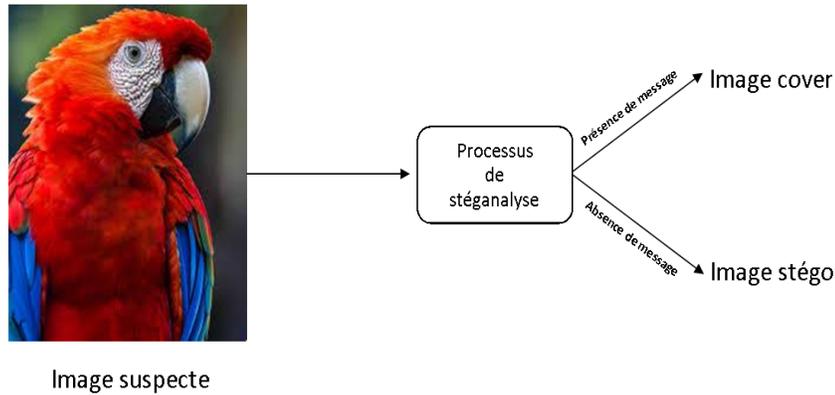


FIGURE 3.5 – Le processus classique de la stéganalyse

3.4 La stéganalyse, détection de stéganographie

La stéganalyse est la discipline duale de la stéganographie et a pour objectif de détecter la présence d'un message dissimulé à l'aide d'un algorithme de stéganographie.

Avant l'apparition des médias numériques, le concept de la stéganalyse n'a pas vraiment existé. Dans l'exemple de la stéganographie du tatouage sur la tête de l'esclave de Histiaeus (voir section 3.2.1), la stéganalyse aurait consisté à regarder l'esclave sous tous les angles possibles et essayer de deviner la présence d'un tatouage caché sur son cuir chevelu.

Tout comme la stéganographie qui diffère de la cryptographie, la notion de la stéganalyse est très différente aussi de celle de la cryptanalyse : en cryptanalyse, le but est de «casser le code», puis extraire le message codé. La stéganalyse n'a pas pour objectif d'obtenir le message caché dans le médium de couverture, mais seulement détecter la présence de celui-ci. L'objectif initial de la stéganalyse était donc de donner une réponse binaire à la question "Y a t-il un message caché dans ce médium?". Il s'agit donc d'une catégorisation des médiums en deux classes : médium sein (*cover*) et médium altéré par un algorithme (*stégo*). La figure 3.5 illustre cette idée de stéganalyse.

Plus tard, avec l'arrivée des médias numériques et la croissance importante de la stéganographie, les moyens de détection de stéganographie ont été développés ainsi que d'autres types de stéganalyse. Dans ce qui suit, nous présentons les différentes classes de stéganalyse. Ensuite nous illustrons quelques approches de stéganalyse actuelles.

3.4.1 Les différentes classes de stéganalyse

Bien que l'objectif principal de la stéganalyse soit de détecter la simple présence d'un message dans un médium suspect, le domaine a évolué et a connu certaines améliorations dérivées de l'idée originale. La forme originale (la classification binaire entre l'image *cover* et *stego*) est qualifiée comme stéganalyse qualitative, même si cette terminologie n'est pas vraiment utilisée. Dans ce qui suit, nous distinguons trois autres types de stéganalyse. Ensuite, nous présenterons dans la section 3.4.2 les différents schémas pratiques de la stéganalyse qui peuvent être appliqués à ces trois classes.

La stéganalyse ciblée

La stéganalyse ciblée, appelée aussi *spécifique*, a pour but d'essayer de déterminer les faiblesses de sécurité d'un algorithme stéganographique particulier. Autrement dit, l'algorithme de stéganographie doit être connu, ce qui permet d'étudier son "implémentation" et/ou ses "failles statistiques". Comme indiqué dans la section 3.3, par exemple, les algorithmes stéganographiques simples qui modifient directement les LSB des coefficients DCT pour insérer le message (comme JSteg) affecteront d'une manière visible les histogrammes de ces coefficients (étant donnée que le message à insérer est de taille raisonnable). Une fois l'algorithme connu, le stéganalyste génère un ensemble de supports stéganographiés avec cet algorithme pour comprendre et analyser ces différentes étapes. Il va ensuite comparer la statistique des images de couverture avec celles qui ont été générées pour pouvoir concevoir un algorithme de stéganalyse spécifique dérivé de la façon dont l'information est insérée, comme pour l'exemple des LSB des coefficients DCT précédemment [100].

La stéganalyse aveugle

La stéganalyse aveugle appelée aussi stéganalyse universelle vise à détecter la présence de messages dissimulés quelque soit le type d'algorithme de stéganographie utilisé. Un exemple d'algorithme de stéganalyse universelle est proposé par Pevný [101] qui est capable de détecter plusieurs schémas de stéganographie dans le domaine JPEG. Les méthodes de stéganalyse universelles les plus pertinentes s'appuient sur l'utilisation des mécanismes d'apprentissage automatique et de la classification basés sur l'extraction des vecteurs de caractéristiques de haute dimension pour pouvoir discriminer les deux classes *cover* et *stégo* [102][103]. Pour cela, une large base de données avec les

deux classes *cover* et *stégo* est utilisée, elle est divisée en deux parties qui sont utilisées respectivement pour l'apprentissage et le test.

La stéganalyse forensic

Enfin, la stéganalyse forensic [104], au-delà de l'étape de détection de présence de message, essaie d'obtenir l'actuel message caché. Il y a plusieurs raisons possibles pour lesquelles la gardienne (l'intermédiaire) souhaiterait obtenir le message. Si la gardienne a un contrôle sur le canal de transmission, au lieu d'éveiller les soupçons de l'expéditeur et du destinataire en coupant simplement le canal, elle tente d'obtenir le message caché sans attirer l'attention des parties communicantes. En ce sens, la stéganalyse forensic est plus proche de la cryptanalyse que de la stéganalyse.

Dans les travaux présentés dans cette thèse, notre objectif final est de développer une méthode de stéganalyse universelle, celle qui repose sur l'apprentissage automatique et la classification.

3.4.2 Différents schémas de la stéganalyse

Dans cette section, nous présentons quelques exemples des schémas de stéganalyse de la littérature afin d'illustrer le principe de fonctionnement.

La stéganalyse χ^2

La stéganalyse χ^2 est une méthode de stéganalyse ciblée qui peut être utilisée pour la détection de tout algorithme utilisant le principe de stéganographie par substitution LSB. Proposée par Westfeld et Pfitzmann dans [105], basée sur le test statistique du χ^2 , elle repose sur l'analyse des fréquences d'occurrences des valeurs de pixels dans une image. Comme décrit précédemment, dans le cas d'insertion par substitution LSB, les bits de poids faible sont directement modifiés pour incorporer le message. Considérons qu'une image est en niveaux de gris codée sur 8 bits, donc les valeurs de ses pixels vont de 0 à 255. Dans le cas d'insertion par substitution LSB, les pixels avec une valeur impaire sont décrémentés et les pixels avec valeur paire sont incrémentés. Ce changement de ± 1 de la valeur de chaque pixel va transformer un pixel de valeur $2i$ en un pixel de valeur $2i + 1$. De la même façon un pixel de valeur $2i + 1$ deviendra

un pixel de valeur $2i$. Par conséquent, les fréquences des deux valeurs de chaque paire $(2i, 2i+1)$ ont tendance à s'égaliser.

Notons par $H_c = (H_c^0, \dots, H_c^{255})^T$, les valeurs de l'histogramme pour toutes les valeurs des pixels de l'image de couverture \mathbf{c} , et par H_s ce même histogramme, pour l'image de *stego* \mathbf{s} . Soit \mathbf{m} le message qui a été caché dans \mathbf{s} . Puisque l'algorithme d'incorporation ne modifie que les LSB, il ne change pas la somme des deux fréquences H_s^{2j} et H_s^{2j+1} . Par conséquent, la moyenne arithmétique est la même pour chaque paire $(2i, 2i+1)$ à la fois, sur l'image originale et sa version stéganographiée. Donc il est possible de calculer le nombre théorique attendu ${}^*H_s^j$ de pixels de valeur i dans l'image \mathbf{s} :

$${}^*H_s^j = \frac{H_s^{2j} + H_s^{2j+1}}{2} \quad (3.8)$$

Ensuite l'utilisation d'un test de χ^2 , par exemple test de corrélation de Pearson entre les valeurs réellement présentes dans l'image et celles calculées théoriquement, permet de vérifier si une image est *stégo* ou non.

L'analyse χ^2 a été adaptée plus tard aux algorithmes stéganographique par substitution dans le domaine transformé sur des images JPEG [106]. Les auteurs proposent d'utiliser le même principe décrit ci-dessus, mais sur les coefficients DCT.

La stéganalyse RS

La stéganalyse RS est aussi une méthode de stéganalyse spécifique à l'algorithme de stéganographie par substitution LSB dans le domaine spatial. Le principe de cette méthode est présenté dans [107]. Le nom de RS est lié à ce que des groupes de pixels sont définis dans cette idée comme étant : Réguliers et Singuliers.

Dans cette méthode, les relations entre les pixels (leurs valeurs) sont prises en compte. Puisque l'algorithme de stéganographie par substitution des LSB modifie plusieurs valeurs de pixel, il augmente très probablement le bruit dans les zones où un message est intégré. L'objectif est de mesurer la quantité de bruit engendré sur l'image entière, en procédant par des petites zones. A cet effet, une fonction de discrimination \mathcal{G} d'un ensemble de n pixels $\mathbf{i} = (i_1, \dots, i_n)^T$ avec $i_j = i(x_j, y_j)$ (x_j et y_j représentent les coordonnées du pixel $i = (x_j, y_j)$ dans l'image i), est calculée par :

$$\mathcal{G}(i) = \sum_{j=1}^n |i_{j+1} - i_j| \quad (3.9)$$

Évidemment, plus la valeur de la fonction \mathcal{G} est grande, plus l'ensemble des pixels \mathbf{i} est bruité et plus il est probable qu'il contienne un message caché. Par la suite, des fonctions réversibles de permutations F_1 , F_{-1} et F_0 sont définies sur l'ensemble des pixels telles que (avec des valeurs de pixel dans $[0, 255]$)

$$\begin{cases} F_1 : 0 \leftrightarrow 1, 2 \leftrightarrow 3, \dots, 254 \leftrightarrow 255 \\ F_0 : 0 \leftrightarrow 0, 1 \leftrightarrow 1, \dots, 255 \leftrightarrow 255 \\ F_{-1} : -1 \leftrightarrow 0, 1 \leftrightarrow 2, \dots, 255 \leftrightarrow 256 \end{cases} \quad (3.10)$$

En utilisant cette notation, les groupes potentiels de pixels \mathbf{i} sont classés en trois groupes

$$\begin{aligned} \mathbf{i} \text{ est régulier si } & \quad \mathcal{G}(F(\mathbf{i})) > \mathcal{G}(\mathbf{i}) \\ \mathbf{i} \text{ est singulier si } & \quad \mathcal{G}(F(\mathbf{i})) < \mathcal{G}(\mathbf{i}) \\ \mathbf{i} \text{ est inutilisable si } & \quad \mathcal{G}(F(\mathbf{i})) = \mathcal{G}(\mathbf{i}) \end{aligned} \quad (3.11)$$

À partir de cette classification, il est clair qu'un groupe régulier contient très probablement un message caché, par rapport à un groupe singulier. Ainsi, après l'insertion d'un message, la quantité relative des groupes réguliers R (divisée par le nombre total des groupes) est supérieure à celle des groupes singuliers S . Les estimations de l'évolution de ces deux valeurs peuvent être obtenues à partir des expériences pour différentes modifications des LSB.

Il suffit alors de comparer les valeurs de R et S d'une nouvelle image suspecte par rapport aux courbes obtenues expérimentalement pour identifier l'image comme *stégo* ou *cover*.

La stéganalyse par calibration

L'une des informations les plus utiles que le stéganalyste pourrait souhaiter obtenir, pour que la stéganalyse devienne facile à effectuer, est l'image de couverture, bien que cela soit pratique peu réaliste. La stéganalyse par calibration est une attaque ciblée proposée dans [108] initialement conçue pour faire face à l'algorithme F5. Le principe de cette méthode consiste à estimer l'histogramme des coefficients DCT de l'image de couverture à partir d'une image stéganographiée. Les auteurs proposent de décompresser tout d'abord l'image suspecte dans le domaine spatial, puis un décalage horizontale et vertical en blocs de $n = 4$ pixels est effectué (n peut être différent

de 4). L'image résultante est ensuite recompressée en utilisant les mêmes paramètres (matrice de quantification, facteur de qualité) que ceux utilisés pour l'image *stégo*. Une fois que les deux images sont disponibles, il est possible d'estimer l'histogramme DCT des deux images, et de les comparer en utilisant la méthode des moindres carrés. Cette méthode de stéganalyse a été ensuite améliorée et adaptée pour d'autres algorithmes stéganographiques dédiés aux images JPEG [109] [110].

La stéganalyse par apprentissage

La stéganalyse par apprentissage est une méthode de stéganalyse non ciblée. Ce schéma de stéganalyse est basé sur les mécanismes d'apprentissage automatique. Il repose sur l'extraction de caractéristiques dans les images *cover* et les images *stégo* qui leurs sont associées. Ce schéma, n'étant pas spécifique à un algorithme particulier, peut être employé pour tout scénario en stéganalyse notamment pour un scénario où l'algorithme employé n'est pas connu d'avance.

Parmi les premières utilisations de l'apprentissage automatique dans le domaine de la stéganalyse, nous citons les travaux de Avcıbaşı et al. [111] où les auteurs ont utilisé un vecteur de 18 caractéristiques constitué de mesures de similarité binaire. Un autre travail réalisé par Farid et al. [112] où les auteurs ont utilisé un vecteur de 72 caractéristiques extraits en transformant l'image par les filtres miroirs en quadrature. Les deux travaux ont utilisé les SVM pour la classification.

D'autres méthodes de stéganalyse universelle basées sur les SVM ont été proposées en exploitant divers vecteurs de caractéristiques [113, 114].

Des techniques à base des réseaux de neurones ont été aussi largement utilisées dans la stéganalyse avec différents vecteurs de caractéristiques. Par exemple dans [115], les auteurs proposent d'extraire les caractéristiques à partir de la décomposition en ondelettes de l'image.

Au début, les anciens algorithmes de stéganalyse basés sur l'apprentissage automatique utilisaient quelques dizaines de caractéristiques. Cependant, la sophistication accrue des algorithmes stéganographiques ont incité les stéganalystes à utiliser des vecteurs caractéristiques de dimension de plus en plus élevée. Par exemple le vecteur de caractéristiques proposé Pevný et al. [110] est composé de 274 caractéristiques, ensuite il a été étendu à deux fois sa taille dans [116]. Ce vecteur de caractéristiques appelé CCPEV sera défini par la suite.

Comme la stéganographie moderne incorpore des changements dans des régions d'images qui sont difficiles à modéliser, des descripteurs statistiques de plus en plus complexes sont nécessaires pour distinguer un plus grand nombre de dépendances parmi les éléments de couverture qui pourraient être perturbés par l'incorporation [117].

Ce court aperçu sous-tend la tendance à la conception d'espace de caractéristiques de grande dimension, qui nécessite en outre l'utilisation de grandes bases d'images pour l'apprentissage. Cependant la complexité de l'apprentissage des SVM ralentit le cycle de développement. Même pour des problèmes de taille modérée, la complexité du calcul de la matrice de Gram qui représente le noyau (kernel) est proportionnelle au carré du produit de la taille du vecteur des caractéristiques par la taille de l'ensemble d'apprentissage. Cela impose des limites sur la taille du problème, et oblige le stéganalyste à concevoir consciemment les caractéristiques en tenant compte des contraintes de complexité définies par les ressources informatiques disponibles.

Partant de l'idée que l'apprentissage automatique ne devrait pas constituer un obstacle à la conception de l'espace des caractéristiques, et pour faciliter le développement ultérieur de la stéganalyse, un outil de classification évolutif appelé *Ensemble Classifier* a été proposé par Kodovský et al [118].

Ensemble Classifier est composé de nombreux apprenants (classifieurs) de faible complexité formés de façon indépendante sur des sous-espaces choisis aléatoirement (uniformément) de l'espace de caractéristiques original. La dimension des sous-espaces aléatoires (\mathcal{D}_{red}) peut être beaucoup plus petite que la dimension complète \mathcal{D} , ce qui diminue considérablement la complexité de l'apprentissage. La décision finale est prise en fusionnant par vote majoritaire les décisions des différents classifieurs individuels.

L'idée de former des sous-espaces aléatoires à partir de l'espace de caractéristiques d'origine n'est pas nouvelle et est connue sous des noms différents. Les stratégies d'ensemble décrites dans la littérature utilisent différentes méthodes de classification. Par exemple, les SVM sont utilisés comme classifieurs dans [119], tandis que dans [120] un ensemble de classifieurs différents sont comparés, y compris la régression logistique, SVM linéaire et FLD.

Cette stratégie d'ensemble ne fonctionnera que si les classifieurs individuels sont suffisamment diversifiés dans le sens où ils commettent des erreurs différentes sur les données non apprises. Afin d'augmenter davantage la diversité mutuelle des classifieurs, chacun d'eux est formé sur un échantillon tiré avec remplacement de l'ensemble

d'apprentissage plutôt que sur l'ensemble d'apprentissage tout entier. Cette stratégie, connue dans la communauté d'apprentissage automatique au nom de *bootstrap* ou *bagging* [121], permet d'obtenir une estimation plus précise de l'erreur de test, ce qui sera important pour déterminer les paramètres optimaux.

Dans *Ensemble Classifier*, on considère $\mathcal{X}^{app} = \{x_i, y_i\}_{i=1}^{i=N}$ un ensemble d'apprentissage composé de N observations des couples d'image (*cover*, *stégo*) représentées respectivement par leur vecteurs de caractéristiques défini par \mathcal{F} . Soit L le nombre de classifieurs individuels, le $l^{\text{ème}}$ sous-espace de caractéristiques est défini par \mathcal{F}_l , avec $|\mathcal{F}_l| = \mathcal{D}_{red}$. Lors de la phase d'apprentissage, chaque classifieur apprend à associer correctement à chaque observation x_i sa classe d'appartenance.

Même si la performance des classifieurs individuels peut être faible, elle s'améliore après la fusion et finit par se stabiliser pour un L suffisamment grand. Le seuil de décision de chaque classifieur est ajusté pour minimiser l'erreur de détection totale sur l'ensemble d'apprentissage.

Les classifieurs de base proposés dans *Ensemble Classifier* sont les classifieurs FLD (Fisher Linear Discriminant) en raison de leurs faibles complexité. Lors de l'apprentissage, chaque classifieur FLD calcule un axe de projection propre permettant de séparer les observations de la classe *cover* de ceux de la classe *stégo*, et le seuil de décision de chaque classifieur FLD est ajusté.

Pour une image de test, le $l^{\text{ème}}$ classifieur prend une décision en projetant sur son axe associé l'image de test et en comparant le résultat avec le seuil (préalablement ajusté lors de l'apprentissage). Après avoir collecté toutes les L décisions, la sortie du classificateur final est formée en les combinant à l'aide du vote majoritaire.

Les auteurs ont réalisé une étude de performance de *Ensemble Classifier* en remplaçant le classifieur FLD par différents algorithmes, en particulier les SVM, les arbres de décision et les classifieurs Baysiens. En conclusion, ils citent qu'aucun de ces choix n'était une alternative viable face au FLD. En outre, la complexité de l'apprentissage de ces apprenants de base alternatifs était beaucoup plus élevée.

Par la suite, les auteurs ont amélioré cette méthode en automatisant le choix de la dimension du sous-espace des caractéristiques ainsi que le nombre de classifieurs de base à apprendre.

Plus récemment, Pibre et al. [122] ont proposé d'utiliser l'apprentissage profond, plus précisément les réseaux de neurones convolutionnels (CNN) pour la stéganalyse.

L'utilisation de l'apprentissage profond permet de gérer les deux étapes des approches classiques à la fois (extraction des caractéristiques et classification). Les résultats présentés sont très satisfaisants, mais sont limités aux expériences en tenant compte d'une seule valeur du *payload* en raison du nombre élevé de calculs et d'expériences sur les CNN.

La liste des méthodes de stéganalyse présentée ci-dessus est loin exhaustive. Nous avons tenté de couvrir certains schémas de stéganalyse les plus largement reconnus et utilisés.

Suite à cet état de l'art, nous pouvons conclure que la stéganalyse universelle fonctionne d'une manière aveugle afin de discriminer les images *cover* et *stégo*. Bien qu'elle ait des performances plus faible que la stéganalyse ciblée (spécifique), elle est de plus en plus exploitée en raison de sa flexibilité et sa capacité d'adaptabilité à des méthodes de stéganographie complètement inconnues.

Dans les travaux présentés dans ce chapitre, nous nous sommes intéressés aux différentes méthodes de stéganalyse basée sur l'apprentissage automatique, plus particulièrement à l'algorithme *Ensemble Classifier* basé sur le principe du boosting.

Dans ce contexte, nous proposons un algorithme de stéganalyse basé sur la théorie des fonctions de croyance, que nous vous présentons et détaillons dans la section suivante.

3.5 Un schéma de stéganalyse basé sur la théorie des fonctions de croyance

3.5.1 Positionnement général

Actuellement, les meilleurs algorithmes de stéganalyse capables d'attaquer les méthodes stéganographiques modernes sont basés sur l'apprentissage supervisé; l'approche générale consiste à sélectionner un ensemble de caractéristiques pertinentes qui peuvent aider à révéler la présence de données cachées dans une image. Ensuite un classifieur est formé à l'aide d'un apprentissage automatique supervisé, pour distinguer les classes *cover* et *stégo*. Partir de ces caractéristiques, l'objectif est donc de classer des nouvelles images, en se basant sur les données d'apprentissage, à partir des

caractéristiques issues de ces images.

La stéganalyse aveugle n'étant pas spécifique à un algorithme particulier, elle est plus généraliste que la stéganalyse ciblée. En outre, deux choix importants ressortent lors de la conception d'un algorithme de stéganalyse universel par apprentissage : la conception de l'espace de caractéristiques et la conception du classifieur. Comme dans le cadre de l'évaluation de la qualité d'image, dans le travail de cette thèse, nous nous focalisons plus sur la façon de former le classifieur plutôt que sur la conception des caractéristiques. De toute évidence, nous ne savons pas à l'avance quelles sont les coefficients des images qui seront perturbés par l'insertion d'un message, donc impossible de sélectionner à priori les bonnes caractéristiques à exploiter. Par conséquent, nous proposons d'exploiter un ensemble de caractéristiques utilisés dans la littérature. Ces caractéristiques seront détaillés dans la section 3.5.2.

Généralement, pour la classification les méthodes statistiques telles que les SVM et les réseaux de neurones sont les plus utilisées. De notre part, nous proposons d'utiliser le cadre théorique des fonctions de croyance comme classifieur. En effet, ce choix est pris car cette théorie, comme nous l'avons vu, fournit une palette d'outils permettant de combiner les informations apportées par chaque élément de l'ensemble d'apprentissage, tout en tenant compte des conflits pouvant exister. Cette méthode repose sur la modélisation de la croyance en un évènement. Cette modélisation est réalisée à partir de fonctions de masse permettant la bonne représentation des connaissances.

Parmi les modèles permettant la construction des structures de croyance m , nous optons pour le modèle de Dencœur basé sur l'algorithme de k -plus proches voisins. Comme mentionné dans le chapitre 1, ce choix s'appuie sur la simplicité de cette méthode qui permet d'obtenir les fonctions de masses directement à partir des distances de l'individu à classer par rapport aux éléments de l'ensemble d'apprentissage. Contrairement aux modèles de vraisemblance, où la difficulté réside dans l'estimation des probabilités de vraisemblance.

Comme nous l'avons dit, ces dernière années, *Ensemble Classifier* a montré sa performance comme un algorithme de stéganalyse. Cette stratégie d'ensemble de classifieurs basée sur le principe du boosting permet d'avoir un prédicteur précis, en combinant plusieurs classifieurs plus faibles, construits sur des sous-espaces de caractéristiques de dimension beaucoup plus faible que la dimension totale du vecteur de caractéristiques.

Nous proposons de dériver l'algorithme des k -plus proches voisins évidentiel, appelé

par la suite Ev-Knn (Evidential k nearest neighbors), vers la stratégie d’ensemble et l’utiliser comme classifieur de base. Avant de décrire plus en détail cette proposition, nous présentons dans les sections 3.5.2 et 3.5.3 les caractéristiques utilisées dans le domaine spatial ainsi que le domaine transformé, ensuite nous démontrons au travers de premiers résultats la pertinence de la méthode proposée.

3.5.2 Espace de caractéristiques pour la stéganalyse

En stéganalyse aveugle, la première étape importante est d’extraire des caractéristiques jugées pertinentes pour la discrimination des classes. Ces caractéristiques doivent être choisies de manière à pouvoir refléter toute modification de l’image originale. En d’autres termes, ces caractéristiques sont le résumé des informations importantes, elles doivent être sensibles au changement stéganographique, et insensibles au contenu de l’image.

La bataille sans fin entre la stéganographie et la stéganalyse pourrait être considérée comme une compétition pour un meilleur modèle. Alice essaie d’insérer un message de façon non-détectable à travers son modèle, alors que Eve essaie d’identifier les relations entre les caractéristiques de l’objet *cover* négligées par Alice afin de les intégrer dans un nouveau modèle. L’évolution de la stéganographie et de la stéganalyse est donc fortement interdépendante et la conception des espaces caractéristiques reflète les succès et les échecs des deux.

La plupart des caractéristiques modernes proposées consistent en plusieurs sous-modèles simples, chacun d’eux abordant un aspect différent de la relation entre les images de couverture et les images *stégo*. Dans ce contexte, plusieurs travaux ont été conduits que ce soit dans le domaine transformé JPEG (CCPEV [116], CCJRM [123]) ou dans le domaine spatial (SPAM [93], SRM [103] et PSRM [124]).

Dans ce manuscrit, nous avons utilisé dans le domaine transformé JPEG les caractéristiques proposées par Fridrich et al [116] nommées CCPEV et CF (Cartesian-calibrated feature) proposé par Kodovský [118]. Ces caractéristiques sont les plus utilisées dans la littérature, en particulier pour détecter nsF5. Dans le domaine spatial les caractéristiques inspirées du *Spatial Rich Model* [103] conçues pour détecter l’algorithme HUGO.

- **CCPEV** : Ce vecteur contient 548 caractéristiques qui proviennent de la fusion d’une part des caractéristiques de Markov qui modélisent les différences entre les

valeurs absolues des coefficients de DCT voisins en tant que processus de Markov et identifie les dépendances intra-blocs de ces coefficients au sein du même bloc (8×8), d'autre part des caractéristiques qui mesurent les dépendances inter-blocs entre les coefficients DCT.

- **CF** (*Cartesian-calibrated feature*) : Un ensemble de 7850 caractéristiques basées sur des matrices de co-occurrence permettant d'identifier les diverses dépendances (inter et intra-bloc de l'image) des coefficients DCT.
- **SRM** (*Full Spatial domain Rich Model*) : Puisque le message secret est dissimulé dans les régions qui sont difficiles à détecter (régions bruitées et texturées), il peut être considéré comme un bruit ajouté à l'image de couverture. De ce fait, différents filtres passe-haut (linéaires, et non-linéaires) sont utilisés afin d'extraire et modéliser le bruit résiduel, et former un vecteur de 34671 caractéristiques.

Nous proposons d'utiliser dans notre méthode chacune de ces caractéristiques. Nous présentons maintenant le principe de l'algorithme *Ev-Knn*.

3.5.3 Ev-Knn pour la stéganalyse

Principe de l'algorithme Ev-Knn

Afin de concevoir ce schéma de stéganalyse, nous commençons tout d'abord par une construction des masses de croyance qui nécessite l'existence préalable d'un apprentissage. Pour cela, nous avons construit une base de données d'images stéganographiées en exploitant diverses méthodes de stéganographie. Ces bases formeront notre ensemble d'apprentissage et de test.

Tout d'abord, nous définissons le cadre de discernement par $\Omega = \{w_c, w_s\}$, tel que w_c représente l'hypothèse qu'une image est *cover*, et w_s représente l'hypothèse que l'image est *stégo*. L'objectif est de déterminer l'appartenance à une classe d'une nouvelle image I associée à un vecteur de caractéristiques x de dimension D en utilisant les connaissances provenant de la base d'apprentissage.

Comme nous l'avons précisé précédemment, nous optons pour l'approche de De-noeux basée sur les K -plus proches voisins pour l'obtention des fonctions de masse. La figure 3.6 représente le schéma de principe de l'algorithme. Pour rappel, partant d'un ensemble d'apprentissage défini par $\mathcal{L} = (x_i, \omega_i)_{i=1}^N$ avec $\omega_i \in \Omega$, l'idée est de consi-

dérer chaque image représentée par son vecteur de caractéristique x_i de l'ensemble d'apprentissage comme une source d'information qui permet de calculer la fonction de masse m_i en fonction de la distance entre x et x_i de la manière suivante :

$$\begin{cases} m_i(\omega_i) = \alpha\phi(d(x, x_i)) \\ m_i(\Omega) = 1 - \alpha\phi(d(x, x_i)) \end{cases} \quad (3.12)$$

avec ϕ une fonction décroissante de \mathbb{R}^+ vers $[0, 1]$ telle que $\phi(0) \in]0, 1[$, $0 < \alpha < 1$ est une constante. d représente la mesure de distance utilisée. Pour la fonction ϕ , nous choisissons ϕ tel que :

$$\phi(d) = \gamma \exp(-d^2) \quad (3.13)$$

où $\gamma \in [0, 1]$ est un paramètre positif comme proposé par Dencœux [75].

Étant donné que les k voisins sont considérés comme k sources d'information indépendantes, la fonction de masse finale m associée à l'individu x est calculée en utilisant la règle de combinaison de Dempster [2] : $m_x(\omega) = \bigcap_{i=1}^k m_i(\omega_i)$.

Enfin, la décision est prise en transformant la fonction de masse m en fonction de probabilité pignistique notée $BetP$ tel que pour tout $\omega \in \Omega$:

$$BetP(\omega) = \frac{1}{1 - m_x(\emptyset)} \left(\sum_{\omega \in A} \frac{m_x(A)}{|A|} \right) \quad (3.14)$$

Nous choisissons ensuite la classe pour laquelle la probabilité pignistique est maximale.

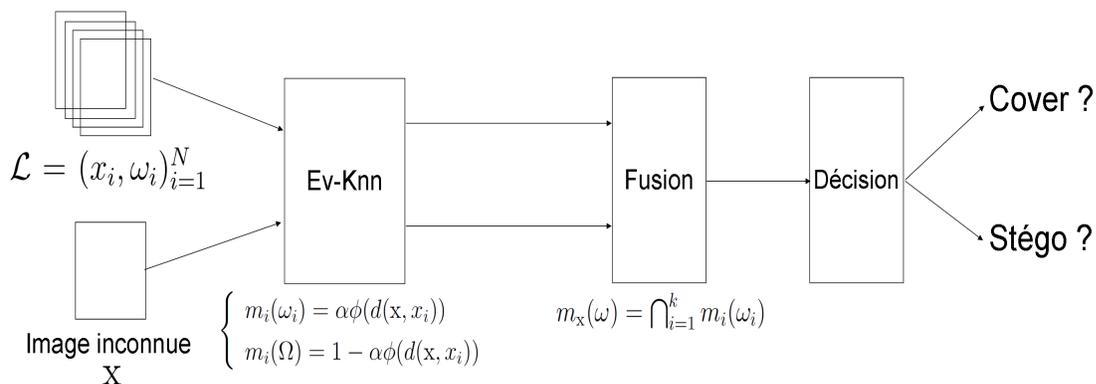


FIGURE 3.6 – Schéma de l'algorithme Ev-Knn

Premiers résultats expérimentaux

Les expérimentations ont été réalisées sur les images provenant de la base d'images BOSS v-1.01². Cette base de données contient 10000 images en niveau de gris de taille 512x512, provenant de 7 caméras différentes. Les images sont au format .pgm.

Les images de la base BOSS sont compressées JPEG avec facteur de qualité de 75 en utilisant la commande de Matlab *imwrite*. L'algorithme de stéganographie *no-shrinkage* F5³ [98] a été utilisé pour simuler l'insertion d'un message avec différents taux d'insertion (*payload* = 0.05, 0.1, 0.15, 0.2, 0.25). Pour cette première expérimentation, nous utilisons le vecteur de caractéristiques CCPEV de dimension 548 proposé par *Fridrich* extrait de chaque image *cover* et *stégo*.

Après avoir extrait les caractéristiques, nous sélectionnons 2000 images *cover* et 2000 images *stégo* de façon aléatoire de la base de données, de telle sorte que chaque image de couverture est associée avec sa version *stégo*.

Une validation croisée à 5 groupes est appliquée pour cette première expérience, où 80% des images de la base ont été sélectionnées pour l'apprentissage, et les 20% restantes constituent l'ensemble de test. Nous souhaitons souligner que l'ensemble de test ne contient aucune image *stégo* dont sa version *cover* fait partie de l'ensemble d'apprentissage et vice versa. En d'autre terme, le classificateur construit n'a aucune information sur les images dans l'ensemble de test.

Ce processus a été répété 1000 fois. À chaque itération, la performance de Ev-Knn est évaluée au travers du calcul de la probabilité d'erreur P_E (définie par la suite). Ensuite, la moyenne des probabilités P_E à travers les 1000 itérations est considérée comme la performance finale de l'algorithme.

En ce qui concerne la modélisation des croyances, nous pouvons remarquer à partir de l'équation 3.13 que le calcul des fonctions de masse et la performance de l'algorithme dépendent fortement de la mesure de distance qui permet de déterminer les k sources d'information d'une image inconnue. Afin de bien choisir la mesure de distance, une étude comparative a été réalisée entre la mesure Euclidienne et la distance de *Mahalanobis* définie par :

$$d(\mathbf{x}, x_i) = \sqrt{(\mathbf{x} - x_i)C^{-1}(\mathbf{x} - x_i)'}$$

où C représente la matrice de covariance de l'ensemble des données d'apprentissage, \mathbf{x}

2. <http://www.ws.binghamton.edu/fridrich/>

3. simulateur nsF5 disponible pour le téléchargement sur <http://dde.binghamton.edu/download/nsf5simulator/>

et x_i représentent deux vecteurs de caractéristiques.

Pour le paramètre k , nous avons réalisé plusieurs tests. Comme le montre la figure 3.7, la probabilité d'erreur se stabilise à partir de $k \geq 300$, elle repart à la hausse avec $k = 600$. Nous remarquons ici que la valeur de k est beaucoup plus grande que celle prise dans le chapitre 2 dans le cadre de l'évaluation de la qualité d'image. Ceci est relative à la grande dimension des bases de données utilisées dans le contexte de la stéganalyse. Pour la suite des expérimentations qui seront présentées, nous avons pris la valeur de $k = 300$.

La fiabilité des sources d'information n'est pas prise en compte ici. Dans ce cas le coefficient $\alpha = 0.95$ permet simplement d'éviter des situations de conflit total, qui se traduit numériquement par une division par zéro lors des opérations de fusion.



FIGURE 3.7 – Le choix du paramètre k

Pour évaluer la performance d'un classifieur, une courbe de la probabilité de fausse détection P_{FD} en fonction de la probabilité de fausse P_{FA} , appelée courbe ROC (Receiver Operating Characteristics) est généralement utilisée car elle offre une vision graphique de la performance de ce classifieur. Cependant, ces courbes peuvent être difficiles à comparer puisqu'elles peuvent se croiser et donc un détecteur peut être meilleur que l'autre uniquement pour certaines fausses alarmes. Par conséquent, une quantité scalaire est calculée à partir de la courbe ROC et est utilisée pour évaluer la performance dans le domaine de la stéganalyse. Cette quantité représente la mesure de probabilité d'erreur de détection P_E :

$$P_E = \frac{P_{FA} + P_{FD}}{2} \quad (3.15)$$

Bien que d'autres mesures de performance existent, nous nous intéressons à cette

probabilité d'erreur utilisée souvent dans la littérature pour évaluer la performance des algorithmes de stéganalyse.

Comme nous pouvons le constater à partir de la figure 3.8 la distance de *Mahalanobis* permet d'obtenir un meilleur résultat en terme de probabilité d'erreur par rapport à la distance Euclidienne quelque soit la valeur du *payload*. En effet la distance Euclidienne ne permet pas de distinguer les deux classes d'images même pour un taux d'insertion plus élevé ($payload = 0.25$) en obtenant une probabilité $P_E = 0.4439$.

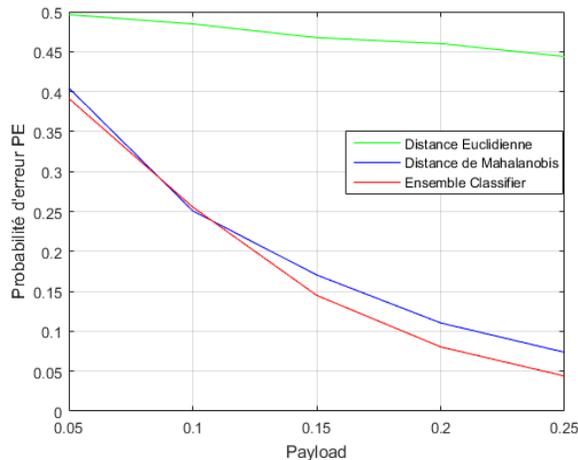


FIGURE 3.8 – Influence de la mesure de distance sur la probabilité d'erreur

Les meilleures performances obtenues avec la distance de *Mahalanobis* peuvent s'expliquer par la prise en compte des corrélations et des dispersions de l'ensemble de données par cette distance grâce à l'utilisation de la matrice de covariance. Pour le reste de ce chapitre, nous proposons donc d'utiliser la distance de *Mahalanobis*.

Maintenant que nous avons défini tous les paramètres liés à Ev-Knn, nous pouvons comparer la performance de l'algorithme par rapport à *Ensemble Classifier*.

Comme le montre la table 3.8, avec $payload < 0.10$ les performances des algorithmes *Ensemble Classifier* et Ev-Knn avec la distance de *Mahalanobis* sont équivalentes. À partir du $payload > 0.10$, un léger avantage à *Ensemble Classifier* est constaté. Nous pouvons conclure de cette première expérience, que l'application de l'Ev-knn nous permet d'obtenir des résultats encourageants.

Pour améliorer la performance, nous proposons d'implémenter la stratégie d'ensemble dans le cadre du classifieur Ev-Knn. Dans la section suivante, nous présentons le principe et les différents étapes de notre proposition.

3.5.4 Discussion sur les espaces de caractéristiques

Rappelons qu'avec le développement de la stéganalyse, les vecteurs de caractéristiques utilisés ont une dimension de plus en plus élevée. Afin de construire un bon classifieur, une étape cruciale consiste à sélectionner les caractéristiques jugées les plus pertinentes séparant les deux classes (*cover* et *stego*).

Ces vecteurs de caractéristiques de très grande dimension contiennent nécessairement des caractéristiques non informatives qui peuvent diminuer de manière significative la précision du classifieur. Cela nous amène à chercher un moyen pour réduire ou sélectionner les données les plus informatives.

Il existe de nombreuses stratégies qui permettent de réduire la complexité de l'ensemble d'apprentissage. Parmi les choix les plus populaires, nous trouvons la technique de réduction de la dimension avec analyse en composantes principales (ACP). L'ACP est appliquée avant la classification comme une étape de prétraitement des caractéristiques. Cependant, de telles méthodes sont rarement adaptées à la stéganalyse car aucun petit sous-ensemble de caractéristiques (basé sur une projection "fixe" de caractéristiques) ne peut fournir des meilleures performances par rapport à celles du vecteur tout entier. Selon Kodovskỳ et al. [123], la réduction de la dimensionnalité et la classification peuvent être effectuées simultanément soit en minimisant directement une fonction objectif, soit en construisant un algorithme itératif pour la réduction de la dimensionnalité avec un *feedback* de la classification après chaque itération.

Parmi les différentes autres approches de la littérature, comme nous l'avons vu, les méthodes d'ensemble permettent d'améliorer les performances des classifieurs en construisant le modèle à partir des sous-espaces aléatoires de caractéristiques.

Nous proposons de nous placer dans ce cadre. Nous décrivons dans les sections suivantes le principe de l'algorithme Ev-Knn basé sur la stratégie d'ensemble, que nous allons appeler par la suite *EVEC-Knn* (*E*Vidential *E*nsemble *C*lassifier *k* nearest *n*eighbors). Ensuite, nous présentons et discutons les différents résultats expérimentaux.

3.5.5 EVEC-Knn pour la stéganalyse

Algorithme

Soit $\mathcal{L} = \{x_i, \omega_i\}$ avec $\omega_i \in \{\omega_c, \omega_s\}$ un ensemble d'apprentissage avec x_i le vecteur de caractéristiques de dimension \mathcal{D} associé à une image I_i , $\omega_i \in \Omega$ est la classe d'appartenance de I_i . Soit \mathcal{D}_{red} la dimension du sous-espace de caractéristiques noté \mathcal{S} avec $\mathcal{D}_{red} < \mathcal{D}$. Notons par $\mathcal{L}_r = \{x_i^r, \omega_i\}$ la restriction de \mathcal{L} dans le sous-espace \mathcal{S}_r avec $x_i^r \in \mathcal{S}_r$ pour $r = 1 \dots R$ avec $R \in \mathbb{N}$.

Supposant maintenant une image I avec son vecteur de caractéristiques \mathbf{x} pour laquelle nous cherchons sa classe d'appartenance à savoir si l'image est *cover* ou *stégo*. \mathbf{x}_r est la restriction de \mathbf{x} dans \mathcal{S}_r . Chaque image de la base d'apprentissage \mathcal{L}_r caractérisée par un sous-espace est considérée comme une source d'information. La masse de croyance associée à \mathbf{x} peut alors être calculée en utilisant l'équation 3.12 (que nous rappelons ci-dessous) avec \mathcal{L}_r comme illustré sur la figure 3.9.

$$\begin{cases} m_i(\omega_i) = \alpha \phi(d(\mathbf{x}, x_i)) \\ m_i(\Omega) = 1 - \alpha \phi(d(\mathbf{x}, x_i)) \end{cases}$$

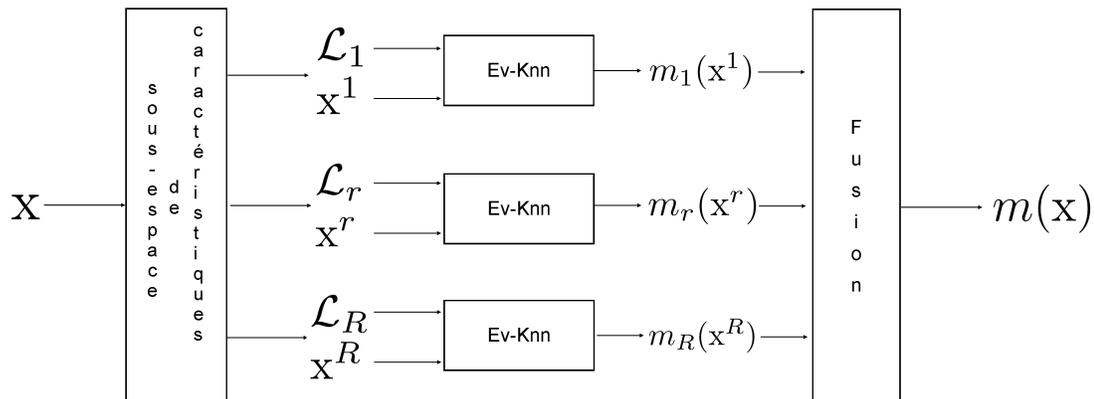


FIGURE 3.9 – Schéma de principe l'algorithme EVEC-Knn

Pour chaque sous-espace \mathcal{S}_r , le vecteur de caractéristiques \mathbf{x}^r peut être considéré comme une source d'information sur laquelle l'EV-knn est appliqué pour produire des masses de croyance m_r concernant la classe d'appartenance de l'image inconnue. L'étape suivante consiste à combiner les différentes masses provenant des différents sous-espaces m_r avec $r = 1, \dots, R$.

Il existe plusieurs stratégies permettant de fusionner les information provenant des différents classifieurs :

1. Prendre la décision pour chaque classifieur puis à travers le vote majoritaire les différentes décisions sont fusionnées.
2. Combiner les information (masses de croyance) avant la prise de décision finale à l'aide des opérateurs de fusion qu'offre la théorie des fonctions de croyance.

Bien que le vote majoritaire soit la méthode la plus simple à mettre en œuvre, il ne permet pas de représenter l'incertitude des différents classifieurs. Par conséquent, nous proposons dans ce travail d'utiliser le cadre théorique des fonctions de croyance, car nous pensons qu'il existe un réel avantage à travailler avec cette théorie, tant pour sa capacité à modéliser l'incertitude et l'imprécision de manière explicite que pour combiner les classifieurs.

Comme nous l'avons vu au , dans la théorie de l'évidence, plusieurs règles de combinaison de croyance ont été proposées. La majorité de ces règles sont basées sur la règle conjonctive ou disjonctive. La différence réside dans la manière à distribuer de la masse conflictuelle. La règle de combinaison de Dempster est l'opérateur de fusion le plus populaire, il distribue le conflit proportionnellement aux masses des éléments focaux non conflictuels.

D'autres règles de combinaisons proposent de supprimer complètement le conflit via une combinaison disjonctive. L'hypothèse faite dans cette règle est que l'un des experts est fiable, à la différence des règles conjonctives où tous les experts doivent être fiables.

La plupart des opérateurs de combinaison ne sont pas idempotents. Ainsi, comme nous l'avons vu dans l'exemple de Zadeh (chapitre 1), la combinaison de fonctions de masse identiques augmente généralement le conflit global. De ce fait, un autre opérateur idempotent de combinaison a été proposé [16]. Cet opérateur permet de de supprimer ce conflit en prenant la moyenne des fonctions de masses :

$$m_{mean}(\omega_j) = \frac{1}{R} \sum_{r=1}^R m_r(\omega_j) \quad \forall \omega_j \in \Omega \quad (3.16)$$

Nous présentons dans ce chapitre, une étude comparative entre la règle de combinaison de Dempster et la combinaison moyenne.

Après la fusion des fonctions de masse, l'étape de la prise de décision intervient.

Il s'agit de choisir l'hypothèse d'appartenance de l'image qui minimise le risque de se tromper. Cette décision est prise en utilisant la probabilité pignistique (équation 3.14).

3.6 Évaluation de la méthode proposée

3.6.1 Paramétrage de la méthode : \mathcal{D}_{red} et R

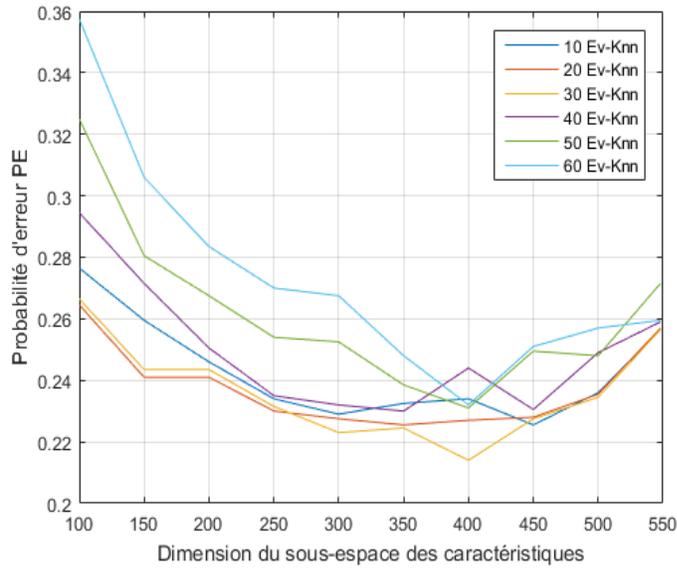
Pour l'algorithme *EVEC-Knn*, deux paramètres sont peut-être avoir un effet sur la performance de l'algorithme : la dimension du sous-espace des caractéristiques \mathcal{D}_{red} et le nombre R de classifieurs Ev-Knn à fusionner pour prendre une décision.

Pour montrer l'influence de ces deux paramètres, nous proposons l'expérimentation suivante : l'algorithme de stéganographie *no-shrinkage* F5 [98] est utilisé pour simuler l'insertion d'un message avec un taux d'insertion de 10% ($payload = 0.1$). Un ensemble de 2000 images de la base BOSS ont été sélectionnées aléatoirement pour l'apprentissage, de telle sorte que chaque image de couverture soit associée avec sa version *stégo*. Les tests ont été effectués sur 2000 nouvelles images de la même base d'images.

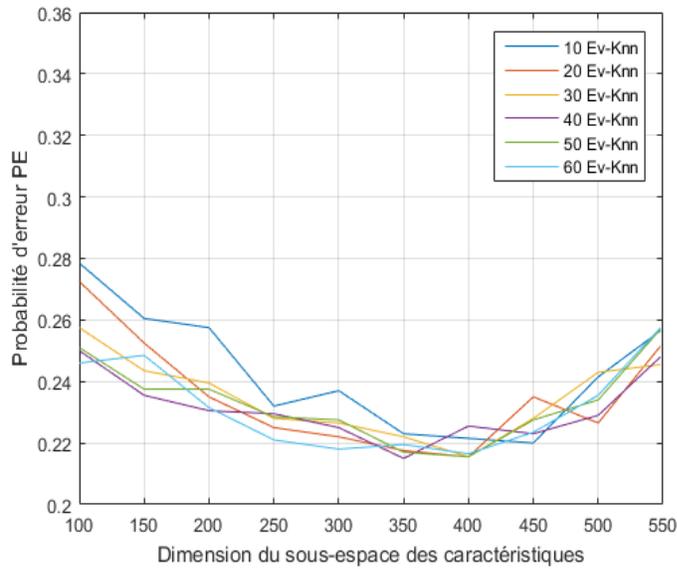
Nous faisons donc varier le nombre de classifieurs R et la dimension du sous-espace de caractéristiques \mathcal{D}_{red} . Les probabilités d'erreur sur l'ensemble des images de test sont présentées dans la figure 3.10.

Les figures 3.10 (a) 3.10 (b) représentent respectivement les résultats obtenus par la règle de combinaison de Dempster et ceux obtenus par l'opérateur de fusion moyenne. À partir de ces résultats nous pouvons soulever différents points :

- L'algorithme *EVEC-Knn* a atteint sa meilleure performance en utilisant un sous-espace de caractéristiques dont la dimension est beaucoup plus petite que la dimension originale ($\mathcal{D}_{red} < \mathcal{D}$) quelque soit l'opérateur de fusion utilisé.
- Les courbes suivent la même tendance, mais l'intervalle de l'erreur de classification avec la combinaison de Dempster est plus grand que celui de la combinaison moyenne. Par exemple pour $\mathcal{D}_{red} = 100$ la probabilité d'erreur $P_E \in [0.2645, 0.3575]$ pour Dempster et $P_E \in [0.2460, 0.2785]$ pour la combinaison moyenne.
- En augmentant le nombre de classifieurs ($R = 40, 50, 60$) pour la combinaison



(a) Combinaison de Dempster



(b) Combinaison moyenne

FIGURE 3.10 – Influence de la dimension du sous-espace des caractéristiques et du nombre de classifieurs Ev-Knn sur le taux d'erreur de classification. Le vecteur de caractéristiques utilisé est CCPEV de dimension $\mathcal{D} = 548$

moyenne, la probabilité d'erreur accroît ce qui est dû probablement à l'augmentation du nombre de situations conflictuelles.

- Le meilleur résultat est obtenu par la combinaison moyenne en prenant ($350 \leq \mathcal{D}_{red} < 400$) quelque soit le nombre de classifieurs $R = (20, 34, 40, 50, 60)$, sauf pour 10 Ev-Knn.
- Pour la combinaison de Dempster une seule courbe ($R = 30$) a réussi à franchir la barre symbolisée par $P_E = 0.2200$ et ceci avec une dimension $\mathcal{D}_{red} = 400$.
- Les meilleurs résultats sont obtenus en utilisant un vecteur de caractéristiques de dimension $\mathcal{D}_{red} = 350$ pour un nombre de classifieurs $R = 40$.

Nous proposons donc d'utiliser la méthode avec le paramétrage suivant : $\mathcal{D}_{red} = 350$ et $R = 40$.

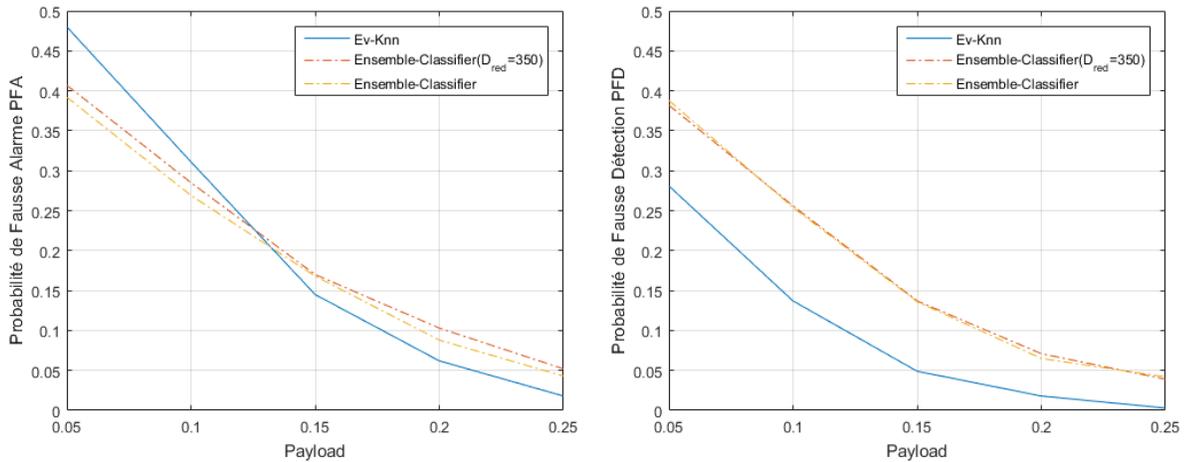
3.6.2 Analyse sur nsF5

Pour évaluer l'efficacité de notre schéma de stéganalyse, les images ont été stéganographiées par l'algorithme d'insertion *nsF5* avec différents taux d'insertion (*payload* = 0.05, 0.10, 0.15, 0.20, 0.25). Nous avons comparé la performance de l'*EVEC-Knn* par rapport à *Ensemble Classifier* avec ces différents *payload*. Pour assurer une comparaison équitable entre les deux méthodes, cette expérience est réalisée sur les mêmes ensembles d'apprentissage et de test. De la même manière, sachant que *Ensemble Classifier* détermine automatiquement les deux paramètres \mathcal{D}_{red} et R , nous proposons d'ajouter une autre comparaison des résultats avec *Ensemble Classifier* en prenant le même paramètre $\mathcal{D}_{red} = 350$ que nous avons utilisé pour *EVEC-Knn*.

Les différents résultats obtenus sont représentés graphiquement dans la figure 3.11. Le premier point que nous pouvons soulever est que la différence en terme de probabilité d'erreur entre *Ensemble Classifier* avec $\mathcal{D}_{red} = 350$ (utilisé par *EVEC-Knn*) et *Ensemble Classifier* avec \mathcal{D}_{red} optimal n'est pas significative, ce qui veut dire que les trois algorithmes ont atteint leurs meilleures performances avec pratiquement la même dimension du sous-espace de caractéristiques \mathcal{D}_{red} .

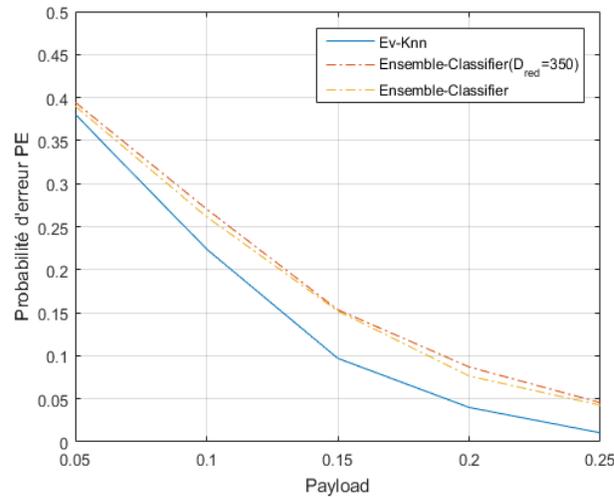
Nous pouvons également remarquer que *EVEC-Knn* déclenche plus de fausses alarmes que *Ensemble Classifier* (figure 3.11(a)) lorsque le *payload* = 0.05, ce qui explique les probabilités d'erreur qui sont proches pour toutes les méthodes.

EVEC-Knn est plus performant en terme de probabilité de fausse de détection (figure 3.11(b)), notamment il arrive à détecter quasiment toutes les images *stégo* quand le *payload* = 0.25.



(a) Probabilité des fausses alarmes

(b) Probabilité des fausses détections



(c) Probabilité d'erreur de classification

FIGURE 3.11 – Comparaison des résultats obtenus par *EVEC-Knn* et Ensemble Classifier avec différents taux d'insertion (payload), $R = 40$, $D_{red} = 350$

En ce qui concerne la probabilité d'erreur (figure 3.11(c)), nous remarquons que les deux méthodes suivent la même tendance en fonction des variations du *payload*, mais *EVEC-Knn* surpasse *Ensemble Classifier* et donne une meilleure précision pour tous les *payload*. *EVEC-Knn* permet d'améliorer la précision de la classification, avec un gain moyen de 3.5%.

3.6.3 Introduction d'une classe de rejet

L'un des avantages de la méthode proposée est qu'elle nous permet d'introduire une classe de rejet qui reçoit les images qui ne peuvent pas être affectées à l'une des deux classes. Ce mécanisme consiste à rejeter une image plutôt que de mal la classer et ainsi induire une erreur, ce qui correspond à un comportement plus prudent vis-à-vis de la décision.

Pour cela, nous définissons un seuil de rejet qui correspond à la limite inférieure de la probabilité pignistique. En dessous de ce seuil, les images seront affectées à la classe de rejet. La figure 3.12 représente ainsi la variation de l'erreur de classification en fonction du seuil et du taux de rejet pour *EVEC-Knn* ($\mathcal{D}_{red} = 350$, $R = 40$) avec le *payload* = 0.1.

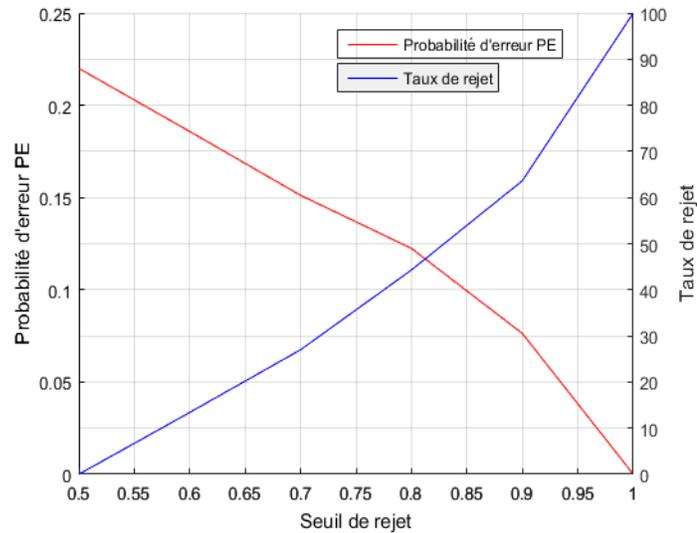


FIGURE 3.12 – Variation de la probabilité d'erreur par rapport au seuil/taux de rejet

Ces résultats montrent bien évidemment l'amélioration de la performance avec la création de la classe de rejet, en effet si nous prenons un seuil de rejet à 0.7, nous remarquons une amélioration des résultats par rapport à la décision sans rejet d'environ 7% en terme de probabilité d'erreur. Cependant, cette amélioration va engendrer un rejet de 27% des images, images dont *EVEC-Knn* ne peut pas les classer avec certitude, ce qui rend le modèle plus prudent. Nous pensons que cette mesure est très importante dans le contexte de stéganalyse, où les images rejetées peuvent nécessiter une attention particulière.

3.6.4 Analyse sur *J-UNIWARD*

Nous proposons maintenant de tester *EVEC-Knn* et de comparer les résultats expérimentaux face à *Ensemble Classifier* sur des nouveaux algorithmes de stéganographie plus récents : *J-UNIWARD* (JPEG Universal Wavelet Relative Distortion) proposé dans [99] pour le domaine de transformé JPEG, HUGO [91] et *S-UNIWARD* (Spatial Universal Wavelet Relative Distortion) pour le domaine spatial. Ces trois algorithmes ont été testés avec différents $payload = \{0.2, 0.4, 0.6\}$.

Nous avons utilisé trois différents vecteurs de caractéristique : le premier contient les caractéristiques *CCPEV* de dimension 548 proposé par *Fridrich*, le deuxième vecteur de dimension 7850 appelé *CF* (Cartesian-calibrated feature) proposé dans [118] pour des images stéganographiées *JPEG*, le troisième vecteur est SRM (*Spatial domain Rich Model*) de dimension $d = 34671$, proposé par [103] pour détecter HUGO et *S-UNIWARD*.

Nous commençons tout d'abord par tenter de détecter *J-UNIWARD*. Rappelons que la performance de *EVEC-Knn* s'appuie sur deux paramètres importants : la dimension du sous-espace de caractéristiques \mathcal{D}_{red} et le nombre de classifieur Ev-Knn R . Pour le vecteur *CCPEV*, comme nous avons obtenu, dans l'expérience précédente, de meilleurs résultats avec $R = 40$, nous avons gardé ce paramètre. Nous conduisons maintenant une étude pour le vecteur de caractéristiques *CF* en prenant $R = 40$ et en faisant varier la dimension du sous-espace \mathcal{D}_{red} afin de trouver la dimension optimale. Cette expérience est réalisée sur un ensemble de 2000 images *cover* et leurs 2000 versions stéganographiées en utilisant l'algorithme *J-UNIWARD* avec un $payload = 0.4$.

À chaque étape, c'est à dire chaque \mathcal{D}_{red} testé, nous conduisons une comparaison de performance avec *Ensemble Classifier*, en prenant le même \mathcal{D}_{red} , sur les mêmes ensembles d'apprentissage et de test.

La figure 3.13 illustre graphiquement les variations de l'erreur de classification en fonction de la dimension du sous-espace. Comme nous pouvons le constater, *EVEC-Knn* est nettement plus performant que *Ensemble Classifier* quelque soit les valeurs \mathcal{D}_{red} testées. Pour *EVEC-Knn*, l'erreur de classification se stabilise avec $750 \leq \mathcal{D}_{red} < 1000$ et le meilleur résultat est obtenu avec 900 caractéristiques au lieu des 7850 caractéristiques de départ.

Après avoir trouvé la dimension optimale du sous-espace, nous proposons d'évaluer la performance de *EVEC-Knn* à détecter *J-UNIWARD* face à *Ensemble Classifier* avec

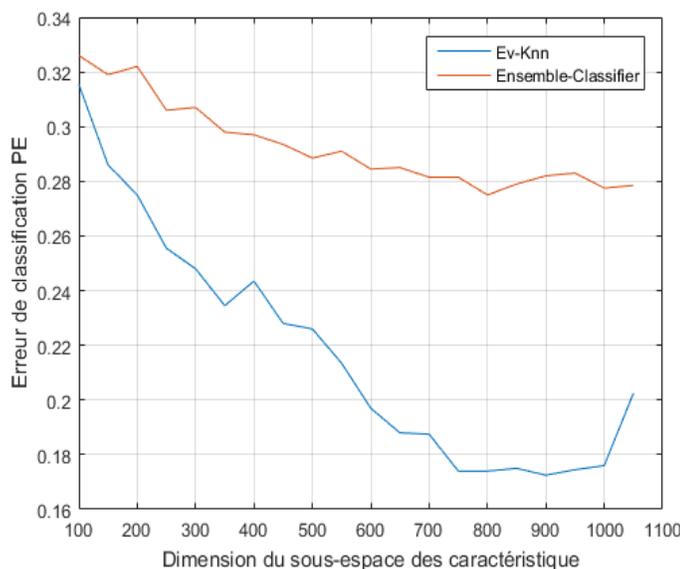


FIGURE 3.13 – Variation de l’erreur de classification en fonction de la dimension du sous-espace de caractéristiques \mathcal{D}_{red}

différentes valeurs du *payload* en utilisant les caractéristiques *CCPEV* et *CF*.

À partir des différents résultats obtenus et qui sont présentés dans le tableau 3.1, nous pouvons remarquer que :

- En utilisant le vecteur *CCPEV*, la performance des deux méthodes est équivalente avec un léger avantage à *Ensemble Classifier* quand le *payload* < 0.6 .
- Avec un *payload* > 0.6 *EVEC-Knn* obtient de meilleurs résultats.
- Le vecteur de caractéristiques *CCPEV* n’est pas le plus adapté pour détecter *J-UNIWARD*.
- En utilisant les caractéristiques *CF*, les résultats sont améliorés par rapport à l’utilisation des caractéristiques *CCPEV*.
- *EVEC-Knn* est plus performant que *Ensemble Classifier* quelque soit la valeur du *payload*.
- Une meilleure détection de *J-UNIWARD* avec une probabilité d’erreur $P_E = 0.1725$ lorsque le *payload* = 0.4, ce qui représente un gain d’environ 11% par rapport à *Ensemble Classifier*.

Après cette série de tests, dans le domaine de transformé *JPEG*, nous pouvons constater que notre méthode permet d’obtenir des résultats performant pour détecter des algorithmes de stéganographie récents tels que *J-UNIWARD* et qu’il résiste com-

<i>Payload</i>	CCPEV		CF	
	<i>EVEC-Knn</i>	<i>Ensemble Classifier</i>	<i>EVEC-Knn</i>	<i>Ensemble Classifier</i>
0.2	0.4425	0.4285	0.3825	0.4185
0.4	0.3245	0.3195	0.1725	0.2830
0.6	0.1545	0.1870	0.1020	0.1635

TABLE 3.1 – Résultats des probabilités d'erreur P_E sur *J-UNIWARD* avec les caractéristiques *CCPEV* et *CF*

paraison avec *Ensemble Classifier* qui est considéré comme l'un des algorithmes les plus performants de stéganalyse de la littérature.

3.6.5 Capacité de généralisation

Nous proposons maintenant un apprentissage plus général en prenant en compte les deux algorithmes de stéganographie nsF5 avec *payload* $\in \{0.05, 0.10, 0.15\}$ et *J-UNIWARD* avec *payload* $\in \{0.2, 0.4, 0.6\}$. Pour chaque *payload*, un ensemble de 1000 images *cover* et leurs 1000 versions stéganographiées sont sélectionnées en tenant compte que les ensembles d'images soient différents pour chaque *payload*. Un total de 12000 images est utilisé pour cette expérimentation dont 6000 images *cover* et 6000 images *stégo*.

La moitié de ces images sont prises aléatoirement pour l'apprentissage et l'autre moitié restante pour le test en respectant toujours que chaque image *cover* soit associée à sa version *stégo*.

	CCPEV	CF
<i>EVEC-Knn</i>	0.3028	0.2723
<i>Ensemble Classifier</i>	0.3273	0.3080

TABLE 3.2 – Résultats des probabilités d'erreur P_E sur nsF5 et *J-UNIWARD* ensemble

Nous remarquons au vu des résultats illustrés dans la table 3.2 qu'en mélangeant les images provenant de deux algorithmes de stéganographie différents avec plusieurs valeurs du *payload*, *EVEC-Knn* reste plus performant que *Ensemble Classifier* quelque soit le vecteur de caractéristiques utilisé. En utilisant les caractéristiques CF, nous arrivons à détecter *J-UNIWARD* avec une probabilité d'erreur $P_E = 0.2723$. Nous constatons que la performance de *EVEC-Knn* baisse par rapport à l'expérience précédente

notamment sur *J-UNIWARD*, cela revient à la présence des images stéganographiées avec un *payload* = 0.2 pour lesquelles nous avons obtenu une probabilité d'erreur $P_E = 0.3825$.

3.6.6 Évaluation dans le domaine spatial sur HUGO et *S-UNIWARD*

Nous proposons maintenant de tester *EVEC-Knn* avec les algorithmes de dissimulation dans le domaine spatial : HUGO [91] et *S-UNIWARD*[99]. Le vecteur de caractéristique utilisé est SRM [103] de dimension $d = 34671$.

De la même manière que les expérimentations précédentes, nous cherchons la valeur optimale de \mathcal{D}_{red} pour *EVEC-Knn*, et nous comparons les résultats avec *Ensemble Classifier* sur l'algorithme HUGO.

Les résultats présentés dans la table 3.3 montrent que *EVEC-Knn* obtient le meilleur résultat sur HUGO avec un vecteur de dimension $\mathcal{D}_{red} = 60$ et $\mathcal{D}_{red} = 60$ pour *S-UNIWARD*. Nous pouvons remarquer qu'avec des sous-espaces de dimension faible, nous obtenons des résultats compétitifs et parfois équivalents à *Ensemble Classifier*, c'est le cas par exemple avec $\mathcal{D}_{red} \in [40, 60]$ pour l'algorithme HUGO et $\mathcal{D}_{red} \in [50, 100]$ pour *S-UNIWARD*.

La table 3.4 présente les résultats de comparaisons des deux méthodes sur HUGO et *S-UNIWARD*. Comme nous pouvons le remarquer, les deux méthodes ont le même comportement, c'est-à-dire elles arrivent à mieux détecter HUGO que *S-UNIWARD*. Cependant *Ensemble Classifier* dépasse *EVEC-Knn* en terme de probabilité d'erreur de classification, sur les deux algorithmes de dissimulation et pour toutes les valeurs du *payload*. Face à ces scénarios mettant en échec la méthode, se pose aussi la question des descripteurs utilisés. En effet, nous avons utilisé les caractéristiques SRM comme étant les meilleurs descripteurs définis dans la littérature, il pourrait donc être envisageable par exemple de tester d'autres caractéristiques dans de futurs travaux.

En résumé, à partir de toutes les expérimentations réalisées, nous pouvons constater que notre schéma de stéganalyse proposé basé sur la théorie de de l'évidence est plus performant et obtient des meilleurs résultats par rapport à l'algorithme *Ensemble Classifier* sur des algorithmes de dissimulation dans le domaine de transformé *JPEG* tels que *nsF5* et *J-UNIWARD*.

\mathcal{D}_{red}	HUGO		S -UNIWARD	
	<i>EVEC-Knn</i>	<i>Ensemble Classifier</i>	<i>EVEC-Knn</i>	<i>Ensemble Classifier</i>
10	0.4055	0.3905	0.4175	0.4250
20	0.3005	0.3310	0.3745	0.3985
30	0.2615	0.2845	0.3550	0.3825
40	0.2590	0.2520	0.3330	0.3500
50	0.2390	0.2495	0.3210	0.3335
60	0.2380	0.2260	0.2950	0.3215
70	0.2455	0.2200	0.2945	0.3120
80	0.2425	0.2150	0.2840	0.3020
90	0.2485	0.2115	0.2965	0.2920
100	0.2635	0.2045	0.2940	0.2850
200	0.2920	0.1795	0.3120	0.2450
300	0.2890	0.1655	0.3235	0.2435
400	0.3095	0.1555	0.3250	0.2410
500	0.3250	0.1545	0.3340	0.2320
600	0.3315	0.1480	0.3300	0.2440
700	0.3355	0.1500	0.3510	0.2425
800	0.3430	0.1485	0.3490	0.2400
900	0.3510	0.1475	0.3585	0.2490
1000	0.3535	0.1580	0.3595	0.2415

 TABLE 3.3 – Variation de P_E en fonction de \mathcal{D}_{red} sur les algorithmes HUGO et S -UNIWARD

En ce qui concerne les algorithmes de stéganographie dans le domaine spatial, nous constatons que *EVEC-Knn* rencontre des difficultés à séparer les images *cover* des images *stégo*, par rapport à *Ensemble Classifier*. Ces résultats nous encouragent à continuer sur cet axe de recherche ; où nous envisageons, dans nos futurs travaux, d'introduire une stratégie permettant d'évaluer la fiabilité de chaque sous-ensemble de caractéristiques afin d'affaiblir le moins fiable avant la fusion et la prise de décision.

	HUGO		S-UNIWARD	
<i>Payload</i>	<i>EVEC-Knn</i>	<i>Ensemble Classifier</i>	<i>EVEC-Knn</i>	<i>Ensemble Classifier</i>
0.2	0.4060	0.3390	0.4365	0.3605
0.4	0.2395	0.1525	0.2840	0.2390
0.6	0.1785	0.1350	0.1835	0.1450

TABLE 3.4 – Résultats des probabilités d’erreur P_E *EVEC-Knn* sur HUGO et *S-UNIWARD*

3.7 Conclusion

Au cours de ce chapitre, nous avons tout d’abord présenté certains concepts et notations utilisées dans la stéganographie/stéganalyse. Nous avons présenté quelques principales méthodes d’insertion de messages cachés et avons dressé un bref état de l’art sur quelques méthodes de détection de ces messages cachées, appelées méthodes de stéganalyse, proposées dans la littérature.

Nous avons proposé un schéma de stéganalyse basé sur la théorie des fonctions de croyance construit sur des sous-espaces aléatoires des caractéristiques. Nous avons tout d’abord tenter d’appliquer l’algorithme des k plus proches voisins évidentiel (appelé *Ev-Knn*) sur le vecteur de caractéristiques des images JPEG en utilisant deux mesures de distance différentes. Bien que les résultats étaient satisfaisants comparaison à *Ensemble Classifier* qui est une des méthodes de référence, l’utilisation triviale de *Ev-Knn* est moins performantes que *Ensemble Classifier*.

Afin d’améliorer ces résultats, nous avons proposé d’implémenter la stratégie d’ensemble basée sur plusieurs classifieurs *Ev-Knn* construits chacun sur un sous-ensemble aléatoire de caractéristiques. Pour la fusion de ces classifieurs, une étude est réalisée afin de trouver le meilleur opérateur de combinaison permettant la fusion de ces classifieurs. Après une série de tests, les meilleurs résultats sont obtenus à travers la combinaison moyenne. Enfin, nous avons proposé d’introduire la classe de rejet qui reçoit les images qui ne peuvent pas être affectées à l’une des deux classes.

La performance de la méthode proposée a été évaluée sur deux algorithmes de dissimulation dans le domaine transformé *JPEG : nsF5, J-UNIWARD*, et sur deux méthodes dans le domaine spatial : HUGO et *S-UNIWARD*. Ces tests expérimentaux ont montré l’efficacité de la méthode proposée dans le domaine transformé. Cependant, il reste de nombreuses configurations qui résident indétectables.

Conclusion & perspectives

Nous avons à travers ce travail abordé deux aspects liés à l'image numérique. Nous nous sommes intéressés tout d'abord à l'évaluation de la qualité d'image, dont le but est de permettre d'évaluer de manière automatique la qualité d'une image, en corrélation avec l'appréciation visuelle humaine, sans avoir recours à sa version originale. Nous avons ensuite abordé le problème de la stéganalyse qui vise à détecter, avec la meilleure fiabilité possible, la présence d'informations cachées dans des images numériques. Le travail que nous avons proposé avait pour objectif, dans les deux cas, de construire des algorithmes que nous avons confrontés et évalués avec différents jeux de données.

Rappel des contributions

Afin d'atteindre ces objectifs, nous nous sommes particulièrement intéressés aux méthodes d'apprentissage automatique notamment dans le cadre théorique des fonctions de croyance qui permet de combiner un ensemble de connaissances ou de croyances des différentes sources d'information d'une manière plus souple, en tenant compte de l'aspect imparfait des mesures.

Nos deux problématiques d'étude ont en commun qu'elles peuvent être abordées par des stratégies d'apprentissage automatiques. Dans ces deux applications, de nombreuses caractéristiques peuvent être extraites des images afin d'estimer soit la qualité des images, soit de détecter les perturbations liées à la présence de messages cachés. Dans une stratégie aveugle, il est très délicat de choisir a priori le ou les meilleures caractéristiques à exploiter en entrée des algorithmes d'apprentissage automatiques. Nous avons ainsi opté pour le cadre théorique des fonctions de croyance qui nous permet de bénéficier d'outils de fusion mais également d'affaiblissement ce qui nous permet de ne pas supprimer certaines caractéristiques avec un a priori. Les bases de cette théorie ont ainsi été présentées suite à l'état de l'art sur les différentes stratégies d'apprentissage en mettant en évidence la souplesse de ce cadre.

Nous avons abordé le problème de l'évaluation de la qualité d'image en présentant cette notion et ses définitions classiques. Nous avons ensuite fourni un bref état de l'art des différentes mesures d'évaluation de la qualité d'image. Nous nous sommes intéressés aux mesures d'évaluation sans référence basées sur l'apprentissage automatique. Ensuite, nous avons proposé une nouvelle mesure sans référence d'évaluation de la qualité d'image appelée *wms-EVreg2* basée sur la fusion de différentes caractéristiques statistiques extraites de l'image, en utilisant la théorie des fonctions de croyance. Les caractéristiques ont été prises de la littérature. Cette fusion est adaptée par le biais de l'utilisation d'un coefficient d'affaiblissement (fiabilité) des masses de croyance. Trois stratégies d'affaiblissement ont été proposées pour affaiblir le groupe de caractéristiques le moins fiable : une valeur du taux d'atténuation égale à 0,95 ; considérer les coefficients de corrélation de Spearman obtenus à partir des premières expériences comme coefficients d'affaiblissement ; utiliser la matrice de confusion issue de la classification des images en fonction de leurs types de dégradation.

Les résultats expérimentaux sur la base d'images LIVE ont montré que *wms-EVreg2* basé sur les coefficients d'affaiblissement issus de la matrice de confusion est la meilleure des trois stratégies. Elle fournit des performances de prédiction de qualité compétitives par rapport aux mesures avec référence. Nous avons aussi montré l'indépendance de *wms-EVreg2* vis-à-vis du contenu de la base d'images en le testant sur les bases d'images TID2013 et CSIQ.

Ce chapitre a montré la pertinence de la prise en compte du taux de fiabilité des différentes caractéristiques dans la construction d'une note finale. Les dernières expérimentations croisant les bases d'image ont constitué une certaine rupture par rapport à la littérature qui d'une manière générale n'utilise pas plusieurs bases dans l'évaluation de l'algorithme.

Nous avons par la suite abordé le deuxième aspect dédié à la recherche d'informations cachées dans les images. Au cours de ce chapitre, nous avons tout d'abord présenté certains concepts et notations utilisées dans la stéganographie et la stéganalyse. Ensuite, nous avons présenté quelques principales méthodes d'insertion de messages cachés et avons dressé un bref état de l'art des méthodes de stéganalyse. Puis, nous avons mis l'accent sur l'importance de l'apprentissage automatique dans la stéganalyse en présentant quelques méthodes proposées dans la littérature. Nous avons tout d'abord proposé un schéma de stéganalyse basé sur la théorie des fonctions de croyance construit avec l'algorithme des k plus proches voisins évidentiel (appelé Ev-Knn) appliqué sur le vecteur de caractéristiques en entier. De nouveau, les mesures

sont directement issues de la littérature. Dans ce cadre, nous avons montré l'influence de la mesure de distance sur le résultat de la classification. Bien que les résultats étaient satisfaisants, nous avons montré qu'une utilisation triviale de Ev-Knn n'est pas satisfaisante car elle ne prend pas en compte la variété des mesures. Nous avons proposé des améliorations à la méthode initiale en implémentant la stratégie d'ensemble basée sur plusieurs classifieurs Ev-Knn construits chacun sur un sous-ensemble aléatoire de caractéristiques. Nous nous plaçons ainsi dans les stratégies de *Boosting* toujours dans le cadre de la théorie de l'évidence. Nous avons réalisé une étude comparative afin de trouver le meilleur opérateur de combinaison permettant la fusion de ces classifieurs. La performance de la méthode proposée a été évaluée sur deux algorithmes de dissimulation dans le domaine de transformé *JPEG : nsF5, J-UNIWARD*, et sur deux autres méthodes dans le domaine spatial : *HUGO* et *S-UNIWARD*. Ces tests expérimentaux ont montré l'efficacité de la méthode proposée dans certains cadres d'applications notamment avec des utilisations en aveugle. Cependant, il reste de nombreuses configurations qui nécessitent d'être étudiées plus en détail. Face à ces scénarios mettant en échec la méthode, se pose aussi la question des descripteurs utilisés. Le cadre de l'algorithme est posé, il faut le rendre maintenant encore plus "robuste".

Perspectives

Évidemment, de nombreuses perspectives naissent de cette contribution. Nous proposons un certain nombre de pistes qui pourraient être explorées dans de futurs travaux.

L'évaluation de la qualité d'image :

- Notre méthode est validée en utilisant les scores subjectifs représentés sous forme de *DMOS* qui se résume à une valeur scalaire. Il serait donc intéressant d'avoir des informations sur la variabilité du *DMOS* afin de les introduire dans la représentation floue utilisée dans notre modèle. La modélisation de *DMOS* par une fonction d'appartenance non restreinte à un Dirac permettrait de mieux considérer les variabilités d'appréciation des observateurs produisant le *DMOS* et pourrait améliorer la précision du score de qualité fourni par notre algorithme. Pour cela une base d'images doit être construite et un test subjectif doit être réalisé tout en gardant la variabilité des scores des observateurs attribués à chaque image.
- Dans notre méthode proposée, nous avons modélisé les masses de croyance en se basant sur la méthode de distance proposée par Denœux, il nous semble intéres-

sant d'étudier d'autres approches permettant la modélisation et la représentation des croyances tels que le modèle de vraisemblance proposé par Appriou et Smets ou bien d'autres approches qui exploitent la matrice de confusion afin d'obtenir ces fonctions de masse permettant d'améliorer les résultats.

Le schéma de stéganalyse :

- La méthode proposée est basée sur une stratégie de fusion de classifieurs construits sur des sous-espaces aléatoires. Il nous semble intéressant d'introduire une stratégie permettant d'évaluer la fiabilité de chaque sous-ensemble de caractéristiques afin d'affaiblir le moins fiable avant la fusion. Cela s'inscrirait dans la même stratégie que la première méthode.
- Dans la stratégie d'ensemble, nous nous sommes restreints à l'utilisation d'un même classifieur (EV-Knn) appliqué sur différents sous-espaces de caractéristiques. Il pourrait être envisageable par exemple d'utiliser différents classifieurs et ensuite fusionner les diverses décisions en fonction de la fiabilité de chaque classifieur. L'idée est que de multiples classifieurs de nature différente permettent, ensemble, d'enrichir la connaissance et améliorer la prise de décision.
- Nous avons montré que notre méthode de stéganalyse est plus performante pour détecter des algorithmes de dissimulation dans le domaine de transformé JPEG. Cependant, cette même approche s'est montré beaucoup moins performante dans le domaine spatial. Ainsi, il nous serait intéressant d'étudier et d'analyser ces algorithmes stéganographiques du domaine spatial, d'étendre les expérimentation en utilisant de nouvelles caractéristiques.

Annexe A

Les différentes dégradations de la base TID2013.

	Nature de la dégradation
Dégradation1	Additive white Gaussian noise
Dégradation2	Additive white Gaussian noise which is more intensive in color components than in the luminance component
Dégradation3	Additive Gaussian spatially correlated noise
Dégradation4	Masked noise
Dégradation5	High frequency noise
Dégradation6	Impulse noise
Dégradation7	Quantization noise
Dégradation8	Gaussian blur
Dégradation9	Image denoising (residual noise)
Dégradation10	JPEG lossy compression
Dégradation11	JPEG2000 lossy compression
Dégradation12	JPEG transmission errors
Dégradation13	JPEG2000 transmission errors
Dégradation14	Non-eccentricity pattern noise
Dégradation15	Local block-wise distortions of different intensity
Dégradation16	Mean shift
Dégradation17	Contrast change
Dégradation18	Change of color saturation
Dégradation19	Multiplicative Gaussian noise
Dégradation20	Comfort noise
Dégradation21	Lossy compression of noisy images
Dégradation22	Image color quantization with dither
Dégradation23	Chromatic aberrations
Dégradation24	Sparse sampling and reconstruction

Les différentes dégradations de la base TID2013

Bibliographie

- [1] G. Shafer *et al.*, *A mathematical theory of evidence*. Princeton university press Princeton, 1976, vol. 1.
- [2] A. P. Dempster, “Upper and lower probabilities induced by a multivalued mapping,” *The annals of mathematical statistics*, pp. 325–339, 1967.
- [3] E. Alpaydin *et al.*, “Introduction to machine learning,” *The MIT press*, 2010.
- [4] K. P. Murphy, *Machine learning : a probabilistic perspective*. MIT press, 2012.
- [5] L. Arnold, “Learning deep representations : Toward a better new understanding of the deep learning paradigm,” Ph.D. dissertation, Université Paris Sud-Paris XI, 2013.
- [6] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT Press, 2016.
- [7] L. A. Zadeh, “Fuzzy sets as a basis for a theory of possibility,” *Fuzzy sets and systems*, vol. 100, pp. 9–34, 1999.
- [8] D. Dubois and H. Prade, “Théorie des possibilités,” 1985.
- [9] I. Bloch, “Information combination operators for data fusion : a comparative review with classification,” *IEEE Transactions on Systems, Man, and Cybernetics-Part A : Systems and Humans*, vol. 26, no. 1, pp. 52–67, 1996.
- [10] J. Kohlas and P.-A. Monney, *A mathematical theory of hints : An approach to the Dempster-Shafer theory of evidence*. Springer Science & Business Media, 2013, vol. 425.
- [11] P. Smets, “The combination of evidence in the transferable belief model,” *IEEE Transactions on pattern analysis and machine intelligence*, vol. 12, no. 5, pp. 447–458, 1990.
- [12] P. Smets and R. Kennes, “The transferable belief model,” *Artificial intelligence*, vol. 66, no. 2, pp. 191–234, 1994.

- [13] E. Lefevre, "Fusion adaptée d'informations conflictuelles dans le cadre de la théorie de l'évidence application au diagnostic médical," Ph.D. dissertation, Institut National des Sciences Appliquées de Rouen, 2001.
- [14] P. Smets, "Belief functions : the disjunctive rule of combination and the generalized bayesian theorem," *International Journal of approximate reasoning*, vol. 9, no. 1, pp. 1–35, 1993.
- [15] T. Reineking, "Belief functions : theory and algorithms," Ph.D. dissertation, Staats-und Universitätsbibliothek Bremen, 2014.
- [16] C. K. Murphy, "Combining belief functions when evidence conflicts," *Decision support systems*, vol. 29, no. 1, pp. 1–9, 2000.
- [17] A. Martin, "Fusion de classifieurs pour la classification d'images sonar," *arXiv preprint arXiv :0806.2006*, 2008.
- [18] A. Martin, "Comparative study of information fusion methods for sonar images classification," in *2005 7th International Conference on Information Fusion*, vol. 2. IEEE, 2005, pp. 7–pp.
- [19] Z. Elouedi, E. Lefevre, and D. Mercier, "Discountings of a belief function using a confusion matrix," in *2010 22nd IEEE International Conference on Tools with Artificial Intelligence*, vol. 1. IEEE, 2010, pp. 287–294.
- [20] Z. Elouedi, K. Mellouli, and P. Smets, "Assessing sensor reliability for multisensor data fusion within the transferable belief model," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 34, no. 1, pp. 782–787, 2004.
- [21] A. Martin, A.-L. Josselme, and C. Osswald, "Conflict measure for the discounting operation on belief functions," in *Information Fusion, 2008 11th International Conference on*. IEEE, 2008, pp. 1–8.
- [22] P. Smets, "Bayes' theorem generalized for belief functions," *Proc. ecai-86*, vol. 2, pp. 169–171, 1986.
- [23] A. Appriou, "Probabilités et incertitude en fusion de données multi-senseurs," *Tiré à part- Office national d'études et de recherches aérospatiales*, 1991.
- [24] T. Denœux, "A k-nearest neighbor classification rule based on dempster-shafer theory," *IEEE transactions on systems, man, and cybernetics*, vol. 25, no. 5, pp. 804–813, 1995.
- [25] L. M. Zouhal and T. Denœux, "An adaptive k-nn rule based on dempster-shafer theory," in *International conference on computer analysis of images and patterns*. Springer, 1995, pp. 310–317.

-
- [26] L. Zouhal and T. Dencœux, “An evidence-theoretic k-nn rule with parameter optimization,” *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 28, no. 2, pp. 263–271, 1998.
- [27] Z. Younes, F. Abdallah, and T. Dencœux, “An evidence-theoretic k-nearest neighbor rule for multi-label classification,” in *International Conference on Scalable Uncertainty Management*. Springer, 2009, pp. 297–308.
- [28] P. Vannoorenberghe and T. Dencœux, “Likelihood-based vs. distance-based evidential classifiers,” in *Fuzzy Systems, 2001. The 10th IEEE International Conference on*, vol. 1. IEEE, 2001, pp. 320–323.
- [29] E. LEFEVRE, “Fonctions de croyance : de la théorie à la pratique,” *Université d’Artois*, 2012.
- [30] A. Chetouani, “Vers un système d’évaluation de la qualité d’image multicritères,” Ph.D. dissertation, paris ; Université Paris-Nord-Paris XIII, 2010.
- [31] I. Recommendation, “500-11, methodology for the subjective assessment of the quality of television pictures,” *International Telecommunication Union, Geneva, Switzerland*, vol. 4, p. 2, 2002.
- [32] P. ITU-T RECOMMENDATION, “Subjective audiovisual quality assessment methods for multimedia applications.”
- [33] L. C. Patrick and A. Florent, “Subjective quality assessment irccyn/ivc database,” 2005, <http://www.irccyn.ec-nantes.fr/ivcdb/>.
- [34] H. Sheikh, Z. Wang, L. Cormack, and A. Bovik, “Live image quality assessment database,” 2006, <http://live.ece.utexas.edu/research/quality/subjective.htm>.
- [35] Z. Parvez, Y. kawayoke, and Y. Horita, “Image quality evaluation database,” 2008, <http://mict.eng.u-toyama.ac.jp/mictdb.html>.
- [36] E. C. Larson and D. M. Chandler, “Most apparent distortion : full-reference image quality assessment and the role of strategy,” *Journal of Electronic Imaging*, vol. 19, no. 1, pp. 011 006–011 006, 2010.
- [37] N. Ponomarenko, O. Ieremeiev, V. Lukin, K. Egiazarian, L. Jin, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, and C.-C. J. Kuo, “Color image database tid2013 : Peculiarities and preliminary results,” 2013, <http://www.ponomarenko.info/tid2013.htm>.
- [38] Z. Wang and A. C. Bovik, “Mean squared error : love it or leave it ? a new look at signal fidelity measures,” *Signal Processing Magazine, IEEE*, vol. 26, no. 1, pp. 98–117, 2009.

- [39] J. Mannos and D. Sakrison, "The effects of a visual fidelity criterion on the encoding of images," *IEEE Transactions Information Theory*, vol. 4, pp. 525–536, 1974.
- [40] T. Mitsa and K. L. Varkur, "Evaluation of contrast sensitivity functions for the formulation of quality measures incorporated in halftoning algorithms," in *Acoustics, Speech, and Signal Processing, 1993. ICASSP-93., 1993 IEEE International Conference on*, vol. 5. IEEE, 1993, pp. 301–304.
- [41] A. Beghdadi and R. Iordache, "A wigner-ville distribution-based image dissimilarity measure," *ISSPA International Symposium on Signal Processing and Its Applications*, vol. 1, pp. 430–433, 2001.
- [42] A. Beghdadi and B. Pesquet-Popescu, "A new image distortion measure based wavelet decomposition," *ISSPA International Symposium on Signal Processing and Its Applications*, vol. 2, pp. 485–488, 2003.
- [43] Z. Wang and A. Bovik, "A universal image quality index," *IEEE Signal Processing Letters*, vol. 2, pp. 81–84, 2002.
- [44] Z. Wang, A. C. Bovik, and E. P. Simoncelli, "Structural approaches to image quality assessment," *Handbook of Image and Video Processing*, vol. 7, p. 18, 2005.
- [45] Z. Wang, E. P. Simoncelli, and A. Bovik, "Multiscale structural similarity for image quality assessment," in *Signals, Systems and Computers, 2004. Conference Record of the Thirty-Seventh Asilomar Conference on*, vol. 2. IEEE, 2003, pp. 1398–1402.
- [46] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *Image Processing, IEEE Transactions on*, vol. 15, no. 2, pp. 430–444, 2006.
- [47] "VQEG :the video quality experts group [online]," available : <http://www.vqeg.org/>.
- [48] H. J. Park and H. Jung, "Reduced-reference quality assessment for jpeg-2000 compressed image," *IEICE Trans. Commun.*, vol. 5, pp. 1287–1294, 2008.
- [49] D. Liu, Y. Xu, Y. Quan, and P. Le Callet, "Reduced reference image quality assessment using regularity of phase congruency," *Signal Processing : Image Communication*, vol. 29, no. 8, pp. 844–855, 2014.
- [50] Z. Wang and E. P. Simoncelli, "Reduced-reference image quality assessment using a wavelet-domain natural image statistic model," *SPIE Human Vis. Electron. Imag.*, vol. 5666, pp. 149–159, 2005.

-
- [51] A. Traoré, P. Carré, and C. Olivier, “Quaternionic wavelet coefficients modeling for a reduced-reference metric,” *Signal Processing : Image Communication*, vol. 36, pp. 127–139, 2015.
- [52] H. R. Sheikh, A. C. Bovik, and L. Cormack, “No-reference quality assessment using natural scene statistics : Jpeg2000,” *IEEE Trans. Image Process.*, vol. 14, pp. 1918–1927, 2005.
- [53] Z. M. P. Sazzad, Y. Kawayoke, and Y. Horita, “No reference image quality assessment for jpeg2000 based on spatial features,” *Image Commun.*, vol. 23, no. 4, pp. 257–268, 2008.
- [54] Z. Wang, H. R. Sheikh, and A. C. Bovik, “No-reference perceptual quality assessment of jpeg compressed images,” in *Proc. IEEE International Conference on Image Processing*, vol. 1, pp. I-477–I-480, 2002.
- [55] A. Bouzerdoum, A. Havstad, and A. Beghdadi, “Image quality assessment using a neural network approach,” in *Signal Processing and Information Technology, 2004. Proceedings of the Fourth IEEE International Symposium on*. IEEE, 2004, pp. 330–333.
- [56] M. Narwaria and W. Lin, “Objective image quality assessment based on support vector regression,” *IEEE Transactions on Neural Networks*, vol. 21, no. 3, pp. 515–519, 2010.
- [57] P. Carrai, I. Heynderickz, P. G. P., and R. Zunino, “Image quality assessment by using neural networks,” *IEEE International Symposium on Circuits and Systems*, vol. 5, pp. 253–256, 2002.
- [58] Y. Dongxue, H. Xinsheng, and T. Hongli, “Image quality assessment based on wavelet coefficients using neural network,” *International Symposium on Neural Networks*, vol. 2, pp. 853–859, 2007.
- [59] P. Gastaldo, R. Zunino, I. Heynderickx, and E. Vicario, “Objective quality assessment of displayed images by using neural networks,” *Signal Process. : Image Commun.*, vol. 20 no.7, pp. 643–661, 2005.
- [60] R. V. Babu, S. Suresh, and A. Perkis, “No-reference jpeg-image quality assessment using gap-rbf,” *Signal Process.*, vol. 87, no. 6, pp. 1493–1503, 2007.
- [61] A. Chetouani, A. Beghdadi, S. Chen, and G. Mostafaoui, “A novel free reference image quality metric using neural network approach,” in *Proc. Int. Workshop Video Process. Qual. Metrics Cons. Electr.*, 2010, pp. 1–4.

- [62] C. Li, A. C. Bovik, and X. Wu, "Blind image quality assessment using a general regression neural network," *IEEE Transactions on Neural Networks*, vol. 22, no. 5, pp. 793–799, 2011.
- [63] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment : A natural scene statistics approach in the dct domain," *IEEE Transactions on Image Processing*, vol. 21, pp. 3339–3352, Aug 2012.
- [64] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Transactions on Image Processing*, vol. 21, pp. 4695–4708, Dec 2012.
- [65] C. C. Chang and C. J. Lin, "Libsvm : A library for support vector machines [online]. available : <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>," 2001.
- [66] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "completely blind" image quality analyzer," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209–212, 2013.
- [67] L. Liu, H. D. H. Huang, and A. C. Bovik, "No-reference image quality assessment in curvelet domain," *ELSEVIER Signal Processing : Image Communication*, vol. 29, pp. 494–505, Dec 2014.
- [68] L. Zadeh, "Fuzzy Sets," *Information and Control*, vol. 8, pp. 338–353, 1965.
- [69] S. Petitrenaud and T. Denceux, "Nonparametric regression analysis of uncertain and imprecise data using belief functions," *International Journal of Approximate Reasoning*, vol. 35, no. 1, pp. 1–28, 2004.
- [70] P. Smets, "The degree of belief in a fuzzy event," *Information Sciences*, vol. 25, no. 1, pp. 1 – 19, 1981. [Online]. Available : <http://www.sciencedirect.com/science/article/pii/0020025581900086>
- [71] R. R. Yager, "Generalized probabilities of fuzzy events from fuzzy belief structures," *Information Sciences*, vol. 28, no. 1, pp. 45 – 62, 1982. [Online]. Available : <http://www.sciencedirect.com/science/article/pii/0020025582900317>
- [72] N. Ponomarenko, O. Ieremeiev, V. Lukin, K. Egiazarian, L. Jin, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti *et al.*, "Color image database tid2013 : Peculiarities and preliminary results," in *Visual Information Processing (EUVIP), 2013 4th European Workshop on*. IEEE, 2013, pp. 106–111.
- [73] A. K. Moorthy and A. C. Bovik, "A two-step framework for constructing blind image quality indices," *IEEE Signal Processing Letters*, vol. 17, no. 5, pp. 513–516, May 2010.

-
- [74] L. Liu, B. Liu, H. Huang, and A. C. Bovik, “No-reference image quality assessment based on spatial and spectral entropies,” *Signal Processing : Image Communication*, vol. 29, no. 8, pp. 856–863, 2014.
- [75] T. Denœux, “Application du modèle des croyances transférables en reconnaissance de formes,” *rn (A)*, vol. 1, p. 2, 1998.
- [76] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, “A statistical evaluation of recent full reference image quality assessment algorithms,” *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3440–3451, 2006.
- [77] S. R. Gunn *et al.*, “Support vector machines for classification and regression,” *ISIS technical report*, vol. 14, 1998.
- [78] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment : from error visibility to structural similarity,” *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [79] L. HÉRODOTE, “Enquête, livres i à iv, texte présenté, traduit et annoté par andrée barguet,” *Paris, Gallimard (Collection : «Folio»)*, 1985.
- [80] R. J. Anderson, “Information hiding : first international workshop, vol. 1174 of,” *Lecture Notes in Computer Science*, pp. 39–48, 1996.
- [81] I. Cox, M. Miller, J. Bloom, J. Fridrich, and T. Kalker, *Digital watermarking and steganography*. Morgan Kaufmann, 2007.
- [82] J. Kelley, “Terrorist instructions hidden online,” *Usa Today*, vol. 19, 2001.
- [83] D. Sieberg, “Bin laden exploits technology to suit his needs,” *CNN, septembre*, 2001.
- [84] G. J. Simmons, “The prisoners’ problem and the subliminal channel,” in *Advances in Cryptology*. Springer, 1984, pp. 51–67.
- [85] J. Fridrich, *Steganography in digital media : principles, algorithms, and applications*. Cambridge University Press, 2009.
- [86] Y. Miche, “Developing fast machine learning techniques with applications to steganalysis problems,” Ph.D. dissertation, Institut National Polytechnique de Grenoble-INPG, 2010.
- [87] C. Cachin, “An information-theoretic model for steganography,” in *Information Hiding*. Springer, 1998, pp. 306–318.
- [88] S. Kouider, “Insertion adaptative en stéganographie : application aux images numériques dans le domaine spatial,” Ph.D. dissertation, Université Montpellier II-Sciences et Techniques du Languedoc, 2013.

- [89] R. Cogranne, “Détection statistique d’informations cachées dans une image naturelle à partir d’un modèle physique,” Ph.D. dissertation, Université de Technologie de Troyes, 2012.
- [90] T. Sharp, “An implementation of key-based digital signal steganography,” in *Information hiding*. Springer, 2001, pp. 13–26.
- [91] T. Pevný, T. Filler, and P. Bas, “Using high-dimensional image models to perform highly undetectable steganography,” in *Information hiding*. Springer, 2010, pp. 161–177.
- [92] T. Filler, T. Pevný, and P. Bas, “Boss (break our steganography system),” July 2010.
- [93] T. Pevný, P. Bas, and J. Fridrich, “Steganalysis by subtractive pixel adjacency matrix,” *information Forensics and Security, IEEE Transactions on*, vol. 5, no. 2, pp. 215–224, 2010.
- [94] N. Provos, “Defending against statistical steganalysis.” in *Usenix security symposium*, vol. 10, 2001, pp. 323–336.
- [95] A. Westfeld, “F5—a steganographic algorithm,” in *Information hiding*. Springer, 2001, pp. 289–302.
- [96] D. Upham, “Jpeg-jsteg. modification of the independent jpeg group’s jpeg software release 4 for 1-bit steganography in jfif output files,” 1997.
- [97] J. Fridrich, T. Pevný, and J. Kodovský, “Statistically undetectable jpeg steganography : dead ends challenges, and opportunities,” in *Proceedings of the 9th workshop on Multimedia & security*. ACM, 2007, pp. 3–14.
- [98] J. Fridrich, M. Goljan, P. Lisoněk, and D. Soukal, “Writing on wet paper,” *Signal Processing, IEEE Transactions on*, vol. 53, no. 10, pp. 3923–3935, 2005.
- [99] V. Holub, J. Fridrich, and T. Denemark, “Universal distortion function for steganography in an arbitrary domain,” *EURASIP Journal on Information Security*, vol. 2014, no. 1, pp. 1–13, 2014.
- [100] J. Fridrich and M. Goljan, “Practical steganalysis : State of the art,” *SPIE photonics West, Electronic Imaging, San Jose, CA*, 2002.
- [101] T. Pevný and J. Fridrich, “Novelty detection in blind steganalysis,” in *Proceedings of the 10th ACM workshop on Multimedia and security*. ACM, 2008, pp. 167–176.

- [102] J. Fridrich, J. Kodovský, V. Holub, and M. Goljan, "Steganalysis of content-adaptive steganography in spatial domain," in *Information Hiding*. Springer, 2011, pp. 102–117.
- [103] J. Fridrich and J. Kodovský, "Rich models for steganalysis of digital images," *Information Forensics and Security, IEEE Transactions on*, vol. 7, no. 3, pp. 868–882, 2012.
- [104] J. Fridrich, M. Goljan, D. Soukal, and T. Holotyak, "Forensic steganalysis : determining the stego key in spatial domain steganography," in *Electronic Imaging 2005*. International Society for Optics and Photonics, 2005, pp. 631–642.
- [105] A. Westfeld and A. Pfitzmann, "Attacks on steganographic systems," in *Information Hiding*. Springer, 1999, pp. 61–76.
- [106] N. Provos and P. Honeyman, "Detecting steganographic content on the internet," Center for Information Technology Integration, Tech. Rep., 2001.
- [107] J. Fridrich, M. Goljan, and R. Du, "Detecting lsb steganography in color and gray-scale images," *IEEE MultiMedia*, no. 4, pp. 22–28, 2001.
- [108] J. Fridrich, M. Goljan, and D. Hoge, "Steganalysis of jpeg images : Breaking the f5 algorithm," in *Information Hiding*. Springer, 2002, pp. 310–323.
- [109] —, "Attacking the outguess," in *Proceedings of the ACM Workshop on Multimedia and Security*, vol. 2002. Juan-les-Pins, France, 2002.
- [110] T. Pevny and J. Fridrich, "Merging markov and dct features for multi-class jpeg steganalysis," in *Electronic Imaging 2007*. International Society for Optics and Photonics, 2007, pp. 650 503–650 503.
- [111] İ. Avcıbaşa, M. Kharrazib, N. Memonc, and B. Sankurd, "Image steganalysis with binary similarity measures," *IEEE International Conference on Image Processing*, pp. 1057–7149, 2002.
- [112] S. Lyu and H. Farid, "Detecting hidden messages using higher-order statistics and support vector machines," in *Information Hiding*. Springer, 2002, pp. 340–354.
- [113] —, "Steganalysis using higher-order image statistics," *IEEE transactions on Information Forensics and Security*, vol. 1, no. 1, pp. 111–119, 2006.
- [114] H. Gou, A. Swaminathan, and M. Wu, "Noise features for image tampering detection and steganalysis," in *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, vol. 6. IEEE, 2007, pp. VI–97.

- [115] Y. Q. Shi, G. Xuan, D. Zou, J. Gao, C. Yang, Z. Zhang, P. Chai, W. Chen, and C. Chen, "Image steganalysis based on moments of characteristic functions using wavelet decomposition, prediction-error image, and neural network," in *Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on*. IEEE, 2005, pp. 4–pp.
- [116] J. Kodovskỳ and J. Fridrich, "Calibration revisited," in *Proceedings of the 11th ACM workshop on Multimedia and security*. ACM, 2009, pp. 63–74.
- [117] J. Kodovskỳ, "Steganalysis of digital images using rich image representations and ensemble classifiers," Ph.D. dissertation, State University of New York, 2012.
- [118] J. Kodovskỳ, J. Fridrich, and V. Holub, "Ensemble classifiers for steganalysis of digital media," *Information Forensics and Security, IEEE Transactions on*, vol. 7, no. 2, pp. 432–444, 2012.
- [119] L. I. Kuncheva, J. J. Rodríguez, C. O. Plumpton, D. E. Linden, and S. J. Johnston, "Random subspace ensembles for fmri classification," *IEEE transactions on medical imaging*, vol. 29, no. 2, pp. 531–542, 2010.
- [120] A. Assareh, M. H. Moradi, and L. G. Volkert, "A hybrid random subspace classifier fusion approach for protein mass spectra classification," in *European Conference on Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics*. Springer, 2008, pp. 1–11.
- [121] L. Breiman, "Bagging predictors," *Machine learning*, vol. 24, no. 2, pp. 123–140, 1996.
- [122] L. Pibre, P. Jérôme, D. Ienco, and M. Chaumont, "Deep learning for steganalysis is better than a rich model with an ensemble classifier, and is natively robust to the cover source-mismatch," *arXiv preprint arXiv :1511.04855*, 2015.
- [123] J. Kodovskỳ and J. Fridrich, "Steganalysis of jpeg images using rich models," in *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2012, pp. 83 030A–83 030A.
- [124] V. Holub and J. Fridrich, "Random projections of residuals for digital image steganalysis," *Information Forensics and Security, IEEE Transactions on*, vol. 8, no. 12, pp. 1996–2006, 2013.

Bibliographie de l'auteur

— Revues Internationales avec comité de lecture

1. **Nadjib Guettari**, Anne-Sophie Capelle-Laizé et Philippe Carré, *Blind image quality assessment based on theory of belief functions*, soumission en Juin 2017.

— Conférences internationales avec comité de lecture

1. **Nadjib Guettari**, Anne-Sophie Capelle-Laizé et Philippe Carré, *Blind image steganalysis based on evidential K-nearest neighbors*, International Conference on Image Processing (ICIP) IEEE, Phoenix 2016.
2. **Nadjib Guettari**, Anne-Sophie Capelle-Laizé et Philippe Carré, *Fusion of imprecise data applied to image quality assessment*, International Conference on Image Processing (ICIP) IEEE, Paris 2014.

Évaluation du contenu d'une image couleur par mesure basée pixel et classification par la théorie des fonctions de croyance

Résumé :

De nos jours, il est devenu de plus en plus simple pour qui que ce soit de prendre des photos avec des appareils photo numériques, de télécharger des images sur l'ordinateur et d'utiliser différents logiciels de traitement d'image pour appliquer des modifications sur ces images (compression, débruitage, transmission, etc.). Cependant, ces traitements entraînent des dégradations qui influent sur la qualité visuelle de l'image. À cet effet, il est nécessaire de disposer d'outils efficaces capables de mesurer l'impact de ces distorsions sur la qualité de l'image. De plus, ces dernières années est apparue une perturbation particulière, à savoir l'insertion de messages "invisibles" à des fins légitimes ou malveillantes pour des communications confidentielles ou secrètes. De nos jours, avec la généralisation de l'internet, la stéganographie devient une pratique populaire et facilement accessible à toute personne désirant dissimuler un message ou communiquer de façon secrète. Par conséquent, la nécessité et le besoin de détecter des objets stéganographiés donnent lieu à la stéganalyse, le processus dual de la stéganographie.

Dans ce manuscrit, nous avons abordé deux problèmes : l'évaluation de la qualité d'image et la détection d'une modification ou la présence d'informations cachées dans une image. L'objectif dans un premier temps est de développer une mesure sans référence permettant d'évaluer de manière automatique la qualité d'une image en corrélation avec l'appréciation visuelle humaine. Ensuite proposer un outil de stéganalyse permettant de détecter, avec la meilleure fiabilité possible, la présence d'informations cachées dans des images naturelles. Dans le cadre de cette thèse, l'enjeu est de prendre en compte l'imperfection des données manipulées provenant de différentes sources d'information avec différents degrés de précision. Dans ce contexte, afin de profiter entièrement de l'ensemble de ces informations, nous proposons d'utiliser la théorie des fonctions de croyance. Cette théorie permet de représenter les connaissances d'une manière relativement naturelle sous la forme d'une structure de croyances.

Nous avons proposé une nouvelle mesure sans référence d'évaluation de la qualité d'image capable d'estimer la qualité des images dégradées avec de multiple types de distorsion. Cette approche appelée *wms-EVreg2* est basée sur la fusion de différentes caractéristiques statistiques, extraites de l'image, en fonction de la fiabilité de chaque ensemble de caractéristiques estimée à travers la matrice de confusion. À partir des différentes expérimentations, nous avons constaté que *wms-EVreg2* présente une bonne corrélation avec les scores de qualité subjectifs et fournit des performances de prédiction de qualité compétitives par rapport aux mesures avec référence.

Pour le deuxième problème abordé, nous avons proposé un schéma de stéganalyse basé sur la théorie des fonctions de croyance construit sur des sous-espaces aléatoires des caractéristiques. La performance de la méthode proposée a été évaluée sur différents algorithmes de dissimulation dans le domaine de transformé *JPEG* ainsi que dans le domaine spatial. Ces tests expérimentaux ont montré l'efficacité de la méthode proposée dans certains cadres d'applications. Cependant, il reste de nombreuses configurations qui résident indétectables.

Mots clés : évaluation de la qualité, stéganographie, stéganalyse, fonction de croyance.

Evaluation of the content of a color image by pixel-based measure and classification through the theory of belief functions

Abstract :

Nowadays it has become increasingly simpler for anyone to take pictures with digital cameras, to download these images to the computer and to use different image processing software to apply modifications on these images (Compression, denoising, transmission, etc.). However, these treatments lead to degradations which affect the visual quality of the image. For this purpose, it is necessary to have effective tools able to measure the impact of these distortions on the image quality. Moreover, in recent years a particular disruption has appeared, namely the embedding of "invisible" messages for legitimate or malicious purposes for confidential or secret communications. Nowadays, with the widespread use of the internet, steganography is becoming a popular practice and easily accessible to anyone who wants to hide a message or communicate in a secret way. Therefore, the need to detect steganographic objects give rise to steganalysis, the dual process of steganography.

In this manuscript we discussed two issues : the image quality assessment and the detection of modification or the presence of hidden information in an image. The first objective is to develop a No-Reference measure allowing to automatically evaluate the quality of an image in correlation with the human visual appreciation. Then we propose a steganalysis scheme to detect, with the best possible reliability, the presence of information embedded in natural images. In this thesis, the challenge is to take into account the imperfection of the manipulated data coming from different sources of information with different degrees of precision. In this context, in order to take full advantage of all this information, we propose to use the theory of belief functions. This theory makes it possible to represent knowledge in a relatively natural way in the form of a belief structure.

We proposed a No-reference image quality assessment measure, which is able to estimate the quality of the degraded images with multiple types of distortion. This approach, called *wms-EVreg2*, is based on the fusion of different statistical features, extracted from the image, depending on the reliability of each set of features estimated through the confusion matrix. From the various experiments, we found that *wms-EVreg2* has a good correlation with subjective quality scores and provides competitive quality prediction performance compared to Full-reference image quality measures.

For the second problem addressed, we proposed a steganalysis scheme based on the theory of belief functions constructed on random subspaces of the features. The performance of the proposed method was evaluated on different steganography algorithms in the *JPEG* transform domain as well as in the spatial domain. These experimental tests have shown the performance of the proposed method in some application frameworks. However, there are many configurations that reside undetectable.

Keywords : quality assessment, steganography, steganalysis, belief function.

