



Université François Rabelais de
Tours
Ecole Doctorale
« Santé, Sciences, Technologies »
Année Universitaire 2003-2004

**THESE POUR OBTENIR LE GRADE DE
DOCTEUR DE L'UNIVERSITE DE TOURS**

Discipline : Informatique
Présentée et soutenue publiquement par :

Yves CARON

Le ?? septembre 2004

Contribution de la loi de Zipf à l'analyse d'images

Directeur de thèse : **Nicole VINCENT**

jury

BASKURT	Attila	professeur	INSA de Lyon
FERNANDEZ	Christine	professeur	Université de Poitiers
STAMON	Georges	Professeur (hémérite)	Université Paris V
FAURE	Alain	Professeur	Université du Havre
VINCENT	Nicole	Professeur	Université Paris V
MAKRIS	Pascal	Maître de Conférences	Ecole Polytechnique de l'Université de Tours

Remerciements

A mes parents

A mon cousin Vincent BENCHECROUN, sans qui je ne serais sans doute pas venu à Tours

A toute ma famille

A mes encadrants, Madame Nicole VINCENT et Monsieur Pascal MAKRIS pour l'aide qu'il m'ont apportés

A Monsieur Christian PROUST, Directeur de l'E3I et du Laboratoire d'Informatique, pour m'avoir accueilli dans son école

A Monsieur Mohand SLIMANE, pour tout ce qu'il a fait pour moi

A l'ensemble des enseignants-chercheurs et doctorants de l'E3I et de Polytech'Tours

Aux étudiants qui ont travaillé sur ce projet, en particulier Johann BRODIER, Michael MESTRE, Harold CHARPENTIER, pour les contributions qu'ils ont apportés à la réalisation de ce projet

A l'ensemble des étudiants de l'E3I et de Polytech'Tours

Table des matières

Remerciements	2
Table des matières	3
Introduction	6
Chapitre 1 : Les lois puissance et leurs applications	7
Introduction	7
1 Les principales lois puissance	7
1.1 Définition	7
1.2 La loi de Pareto ou « loi des 80/20 »	8
1.3 La loi de Zipf	9
1.4 La loi de Zipf inverse	12
2 Interprétation des lois puissance	14
2.1 le modèle stochastique de Simon	14
2.2 La loi de Zipf-Mandelbrot : interprétation par les arbres lexicographiques.....	19
2.3 Les différents processus aboutissant à une loi puissance	21
3 Applications de la loi de Zipf	23
3.1 Loi de Zipf et langage naturel : extension aux groupes de mots	23
3.2 Loi de Zipf et population des villes	23
3.3 La double loi de Pareto	25
3.4 La loi de Zipf et Internet	27
3.5 La loi de Zipf en musique	28
3.6 Application de la loi de Zipf à l'étude de signaux acoustiques médicaux	28
Conclusion	29
Chapitre 2 : Les lois puissance en analyse d'image	31
Introduction	31
1 Présentation des modèles utilisés	32
1.1 La loi de Zipf	32
1.2 La loi de Zipf inverse	35
2 Codage de l'image	36
2.1 Utilisation directe des niveaux de gris de l'image	36
2.2 Rééchantillonnage de l'échelle des niveaux de gris : le codage des n classes	39
2.3 Codage des rangs généraux	40
2.4 Codage des rangs généraux avec seuil	43
2.5 Bilan	43
3 Application des modèles de lois puissance	43
3.1 Utilisation du codage des 9 classes	44
3.2 Utilisation du codage des rangs généraux	47
3.3 Utilisation de la loi de Zipf inverse	51
3.4 Bilan	54
4 Propriétés du modèle	55
4.1 Influence de la résolution de l'image	55
4.2 Comportement de la loi de Zipf face à une translation	57
4.3 Comportement de la loi de Zipf face à une rotation	62
4.4 Relations entre la loi de Zipf et la géométrie fractale	72
Conclusion	76

Chapitre 3 : Evaluation de la qualité de la compression des images.....	77
Introduction	77
1 Généralités sur la compression d'images	77
1.1 Nécessité de la compression des images	77
1.2 La compression sans perte.....	78
1.3 La compression avec perte	79
2 Evaluation de la qualité de la compression	83
2.1 L'évaluation subjective humaine.....	83
2.2 Les critères numériques univariés	85
2.3 Les critères numériques bivariés classiques	89
2.4 Les critères pondérés	90
2.5 Les critères perceptuels	93
2.6 Les critères graphiques	97
2.7 Conclusion.....	100
3 Mesure de la qualité de compression à l'aide de la loi de Zipf	101
3.1 Mise en évidence des distorsions introduites par la compression	101
3.2 Utilisation de la distance euclidienne	102
3.3 Le critère ZQ (Zipf Quality).....	104
3.4 Amélioration du ZQ	107
3.5 Utilisation du ZQ avec la compression SPIHT	110
3.6 Evaluation de la compression JPEG2000.....	113
3.7 Conclusion.....	116
4 Optimisation de la compression	117
4.1 Principe de l'optimisation multicritères	117
4.2 La méthode taboue	118
4.3 Résultats de l'optimisation	119
Conclusion.....	120
Chapitre 4 : Détection d'objets et de zones d'intérêt	121
Introduction	121
1 Etat de l'art	121
1.1 Notion de région d'intérêt	121
1.2 Le système visuel humain	123
1.3 Mesures de saillance basées sur les caractéristiques visuelles de bas niveau	125
1.4 Modèles d'attention visuelle multi-échelle pour l'analyse de scènes	128
1.5 Utilisation de transformations de symétrie.....	129
1.6 Détection de zones d'intérêt à l'aide d'ondelettes.....	131
1.7 Approche multi-échelles basée sur l'entropie	133
1.8 Approche multirésolution pour l'apprentissage du fond.....	134
1.9 Détection d'objets par dimension fractale.....	135
1.10 Méthode des orientations locales dominantes	136
1.11 Conclusion.....	137
2 Détection d'objets artificiels dans les milieux naturels.....	138
2.1 Position du problème.....	138
2.2 Application des modèles de lois puissance	140
2.3 Résultats et interprétation.....	143
2.4 Influence de la résolution de l'image	147
2.5 Utilisation d'un prétraitement	151
2.6 Utilisation d'un codage à seuil	156
2.7 Conclusion.....	159
3 Détection de zones d'intérêt à l'aide des lois puissance	160

3.1	Introduction	160
3.2	Utilisation de la loi de Zipf	161
3.3	Utilisation de la loi de Zipf inverse	165
3.4	Entropie des fréquences d'apparition des motifs	171
3.5	Utilisation d'un codage basé sur l'écart type	172
3.6	Fusion de plusieurs codages	173
3.7	Comparaisons des différents modèles de lois puissance utilisés.....	177
3.8	Classification d'images par réseaux de neurones.....	179
3.9	Conclusion.....	181
4	Comparaison avec d'autres méthodes	182
4.1	Détection par calcul d'une dimension fractale	182
4.2	Détection de zones d'intérêt par carte d'importance.....	184
	Conclusion.....	189
	Conclusion et perspectives	190
	Bibliographie	192

Introduction

Dans notre monde moderne, notamment avec l'apparition du multimédia, l'image a pris une importance considérable comme moyen de transmettre de l'information. De plus en plus, les images se présentent sous forme numérique. Du fait de l'abondance des informations contenues dans ces images, il devient nécessaire de pouvoir effectuer automatiquement certains traitements. C'est pourquoi l'analyse d'images est devenue à l'heure actuelle l'un des thèmes prépondérant de la recherche dans les sciences de l'information.

Les problématiques liées à l'image comptent parmi les plus difficiles à résoudre en raison de leur caractère éminemment subjectif : en effet les images sont en général destinées à être vues par un être humain, et l'on sait le caractère subjectif de l'interprétation humaine. Les problèmes de l'analyse d'image, où l'on cherche à simuler la perception humaine de l'image, sont donc des problèmes ouverts que l'on ne peut espérer résoudre de manière exacte. De nombreuses méthodes très diverses et plus ou moins efficaces ont été proposées pour résoudre ces problèmes, mais aucune ne peut prétendre à une efficacité totale sur toutes les images.

L'image est, au même titre que le langage naturel humain oral ou écrit, un vecteur d'une information sémantique destinée à être interprétée par le cerveau humain, ce qui conduit à l'idée d'utiliser pour l'analyse d'image des méthodes issues de l'analyse linguistique. C'est pourquoi nous allons appliquer aux images un modèle statistique couramment utilisé en analyse de textes : la loi de Zipf.

Parmi les applications envisageables de la loi de Zipf en analyse d'image, nous allons ici nous intéresser à deux catégories de problèmes : l'évaluation de la qualité des images compressées et la détection d'objets et de zones d'intérêt dans les images.

Dans le premier chapitre de cette thèse, nous allons présenter la loi de Zipf et les autres modèles de loi puissance utilisés dans les sciences humaines et naturelles, les interprétations qui peuvent être données de ces modèles et leurs principales applications.

Le second chapitre présente l'application aux images de la loi de Zipf, et en particulier les codages de l'image qui seront utilisés afin de pouvoir adapter au cas des images un modèle conçu pour l'analyse de texte, ainsi que les principales propriétés du modèle.

Le troisième chapitre est consacré à l'utilisation de la loi de Zipf pour l'évaluation de la compression d'images. Après avoir présenté les méthodes existantes, nous définirons une mesure de la qualité de la compression basée sur la loi de Zipf, ainsi qu'une méthode d'optimisation de la compression d'une image.

Enfin, le quatrième chapitre présente l'utilisation de la loi de Zipf pour la détection d'objets et de zones d'intérêt dans les images. Nous allons d'abord aborder ce problème dans un cas particulier, celui de la détection d'objets artificiels dans des milieux naturels, avant d'appliquer les modèles de loi puissance dans le cadre plus général de la détection de zones d'intérêt d'une image quelconque.

Chapitre 1 : Les lois puissance et leurs applications

Introduction

Dans cette section nous allons montrer comment dans de nombreuses études, des lois puissance ont déjà été utilisées pour modéliser des phénomènes extrêmement divers. Nous allons en particulier nous intéresser aux lois puissance du type de la loi de Zipf qui établissent une relation entre les fréquences d'apparition des différentes occurrences d'un phénomène et le rang de ces occurrences dans une suite ordonnée.

Avant de présenter ces différentes études, il convient de préciser ce que nous entendons par le terme de loi puissance. Nous allons ensuite présenter les principaux modèles de lois puissance qui existent et les interprétations qui peuvent être données pour justifier leur existence. Nous verrons ensuite les principales applications de ces modèles.

1 Les principales lois puissance

1.1 Définition

Des répartitions statistiques en lois puissance ont été observées dans de très nombreux domaines, aussi bien dans les phénomènes naturels que dans les productions des sociétés humaines.

Une loi puissance est une fonction qui lie deux quantités et qui prend la forme $y=ax^{-b}$, où a et b sont des constantes. Sa représentation graphique dans un repère (x,y) est la suivante (figure 1):

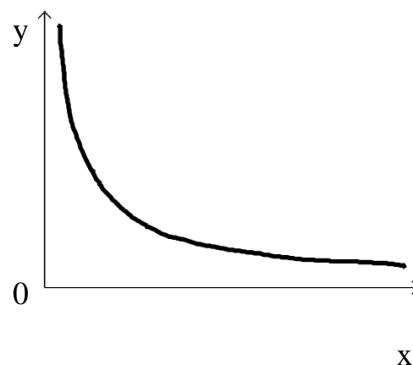


Fig.1 : Représentation d'une loi puissance dans un repère linéaire

Le plus souvent, on représente les lois puissance dans un repère bi-logarithmique, c'est à dire un repère dont aussi bien l'abscisse que l'ordonnée portent des échelles logarithmiques. Dans un tel repère, la représentation graphique d'une distribution en loi puissance est une droite, comme on peut le voir sur la figure 2. On a alors $\ln y = \ln a - b \ln x$. L'exposant $-b$ de la loi puissance est alors lié à la pente de la représentation linéaire dans le repère bi-logarithmique.

Les modèles de loi puissance les plus connus sont la loi de Pareto et la loi de Zipf. Les lois puissance ont fait l'objet d'observations expérimentales dans de très nombreux domaines comme la linguistique, l'économie, la démographie ou la génétique, et diverses interprétations ont été proposées pour expliquer de façon théorique, au moins dans certains cas, les résultats empiriques obtenus.

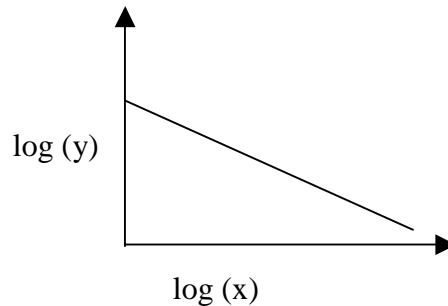


Fig.2 : Représentation d'une loi puissance dans un repère bi-logarithmique

1.2 La loi de Pareto ou « loi des 80/20 »

Historiquement, la première mise en évidence d'une loi puissance a été faite en 1897 par l'économiste italien Vilfredo Pareto [Par-97]. En étudiant la répartition des revenus personnels des individus dans les principaux pays industrialisés, Pareto a constaté que cette répartition suivait une loi puissance.

Plus précisément, si l'on s'intéresse à la probabilité qu'un individu pris au hasard dans la population ait un revenu X supérieur à une certaine somme x , cette probabilité est, selon Pareto, donnée par la formule suivante :

$$P(X > x) = (x/x_{\min})^{-k} \quad (1-2-1)$$

Dans cette formule, x_{\min} représente le revenu minimum et k est une constante positive. L'espérance mathématique du revenu est donnée par la formule suivante :

$$\bar{x} = x_{\min} k / (k-1) \quad (1-2-2)$$

La valeur de k caractérise la pente de la distribution et mesure l'inégalité des revenus. Cette loi est en principe valable pour tous les pays à économie de marché. Elle se vérifie aussi bien pour les revenus des particuliers que pour les chiffres d'affaire des entreprises. En réalité, la loi de Pareto ne modélise correctement que les revenus les plus élevés. En 1931, Gibrat [Gib-31] a montré que les revenus des classes moyennes suivaient une loi lognormale plutôt qu'une loi puissance. Cependant, Mitzenmacher [Mit-01] a mis en évidence un lien entre les distributions lognormales et les lois puissances. Il a montré qu'avec un même modèle mathématique de la variation des revenus, on pouvait obtenir soit une loi puissance, soit une loi lognormale selon qu'il existe ou non un revenu minimum. Donc, les différences constatées entre les travaux de Pareto et les études plus récentes tiennent sans doute aux changements

sociaux intervenus au cours du vingtième siècle, notamment au développement des classes moyennes et à l'instauration de minima sociaux.

Certains ont essayé d'utiliser la loi de Pareto à des fins politiques afin de justifier les inégalités sociales par une sorte de « loi naturelle ». Nous ne nous étendrons pas sur le sujet, mais il est bon de remarquer que la loi de Pareto ne s'applique au sens strict qu'aux revenus les plus élevés, et ne décrit pas par conséquent les inégalités entre riches et pauvres, mais seulement entre les riches eux-mêmes. Ce n'est pas tant l'existence d'inégalités sociales qui pose problème mais l'importance de celles-ci, mesurée par le coefficient k de la formule de Pareto, et par le pouvoir d'achat de la classe la plus pauvre, représenté par le revenu minimum x_{\min} . Donc, une société peut être relativement égalitaire et la répartition des revenus suivre quand même une loi de Pareto.

Pareto a également proposé une formulation très schématisée de sa loi. C'est la loi dite des 80/20 : selon Pareto, les 20% des individus les plus riches possèdent 80% de la richesse d'un pays. D'autres répartitions de ce type ont été mises en évidence dans les domaines les plus divers, comme la production industrielle (20% des produits fabriqués par une entreprise représentent 80% du volume de production de celle-ci), le trafic routier (20% du réseau routier supporte 80% du trafic), la gestion des ressources humaines (20% des individus sont responsables de 80% des problèmes), et bien d'autres exemples encore.

1.3 La loi de Zipf

La *loi de Zipf* [Zip-49] a été mise en évidence par le linguiste américain George K. Zipf (1902-1950). C'est une loi empirique qui s'intéresse à la répartition statistique des fréquences d'apparition des différents mots dans un langage. Contrairement à la loi de Pareto qui s'intéresse à des probabilités, on utilise ici directement les résultats du comptage des mots, des fréquences absolues. Dans les travaux originaux de Zipf, il s'agissait de la langue anglaise écrite. Le premier ouvrage sur lequel a travaillé Zipf pour mettre en évidence sa loi est le roman *Ulysses* de James Joyce. Selon la loi de Zipf, les fréquences d'apparition des mots classés par ordre décroissant de leur fréquence d'apparition s'organisent suivant une loi puissance. Plus précisément, si l'on désigne par F la fréquence d'apparition du mot de rang r dans la suite, la relation est donnée par la formule suivante :

$$F = a r^{-\alpha} \quad (1-3-1)$$

Dans cette formule, a et α sont deux constantes positives. La valeur de la puissance α caractérise la loi puissance. Dans les cas des textes de langue anglaise étudiés par Zipf, la valeur de la puissance est proche de 1. Si l'on suppose que la puissance α est égale à 1, cela conduit à une simplification de la formule que l'on peut quelquefois trouver dans la littérature :

$$F = a / r \quad (1-3-2)$$

Selon cette approximation, *le produit rang-fréquence est constant* et la fréquence des mots est inversement proportionnelle à leur rang.

La *loi de Zipf*, également appelée *loi rang-fréquence* est en général représentée graphiquement dans un repère bi-logarithmique. Si la répartition des fréquences des mots suit une loi puissance, on obtient une droite. On appellera *courbe de Zipf* cette représentation

graphique. Dans la pratique, la loi de Zipf est valable pour les 5000 mots les plus fréquents environ, au-delà de 5000 mots la courbe de Zipf tend à s'incurver vers le bas. Ce résultat a notamment été mis en évidence par Kucera et Francis [Kuc-67].

Depuis les travaux originaux de Zipf, des répartitions en loi puissance de la fréquence des mots ont été observées par de nombreux auteurs dans de nombreuses langues, et cela aussi bien en ce qui concerne le langage parlé qu'écrit. On peut citer par exemple Miller [Mil-58] ou Dahl [Dah-79] pour l'anglais, Lua [Lua-94] pour le chinois, Billmeier [Bil-69] pour l'allemand et Kalinin [Kal-64] pour le russe. On observe bien entendu des différences entre les courbes obtenues selon la langue et le type de textes sur lesquels l'analyse a été effectuée, mais on retrouve, quels que soient la langue et le vocabulaire utilisés, une répartition en loi puissance. La comparaison entre les caractéristiques des courbes de Zipf a été utilisée notamment par Havlin [Hav-95] pour essayer d'identifier l'auteur d'un texte.

En fait, Zipf n'est pas le premier à avoir décrit la répartition des fréquences des mots. Les premiers travaux dans ce domaine ont été réalisés par Estoup en 1916 [Est-16]. Si l'on examine de plus près la répartition des fréquences des mots, quelle que soit la langue, on remarque que les mots les plus fréquents sont les mots-outils grammaticaux comme les articles, pronoms, conjonctions, prépositions, ainsi que les verbes auxiliaires. Ces mots sont parfois appelés « mots vides » car ils n'ont pas de valeur sémantique propre, il n'ont qu'une fonction syntaxique. Ce résultat n'a rien de surprenant, en effet ces mots sont présents en grande quantité dans n'importe quel texte quel que soit le sujet traité, et leur présence est absolument nécessaire. Ces mots grammaticaux sont quelquefois désignés comme faisant partie du *vocabulaire fermé* car ils existent en nombre limité, contrairement au *vocabulaire ouvert* des noms, adjectifs et verbes dont le nombre est virtuellement infini.

Il est également possible de calculer la fréquence des mots en se limitant à une catégorie grammaticale donnée, par exemple celle des noms qui sont les mots les plus signifiants du point de vue sémantique. Là encore, on retrouve une répartition en loi puissance. Ce fait a été mis en évidence pour la première fois par Yule en 1925. [Yul-25]

Un problème qui se pose lors de l'établissement des listes de fréquence des mots, dans le cas des langues naturelles, est la manière de définir ce qui constitue un *mot*. La question n'est absolument pas triviale contrairement à ce que l'on pourrait croire au premier abord. Selon la définition utilisée pour la notion de mot, les résultats des analyses effectuées pour une même langue peuvent être sensiblement différents. Une définition simple, voire simpliste du mot est de considérer que tout groupe de lettres séparé par deux espaces constitue un mot, et que chaque groupe de lettres différent constitue un mot différent. Cette manière de définir et de dénombrer les mots est simple et facile à automatiser, mais ne manque pas de poser des problèmes. Un premier problème est celui des flexions : conjuguons, déclinons, accords de l'adjectif. Se posent également le problème des homonymes et celui des mots pouvant appartenir à plusieurs catégories grammaticales. Ce problème est d'autant plus important que les mots appartenant à plusieurs catégories grammaticales sont souvent des mots-outils extrêmement fréquents. Il y a également le problème des noms composés et des locutions comme *afin de, quant à...* Faut-il les compter comme un ou deux mots ? Tous ces problèmes font qu'il n'est pas toujours simple de compter les mots. Les méthodes de comptage simple, si elles sont facilement automatisables, ne donnent pas une bonne image des fréquences des mots. Pour cela, il est nécessaire d'effectuer une analyse syntaxique et sémantique de la langue, et donc souvent d'utiliser un comptage manuel qui est extrêmement fastidieux.

L'explication donnée par Zipf de cette distribution est basée sur le *principe du moindre effort*.

Le principe du moindre effort consiste à minimiser l'effort à entreprendre afin d'accomplir une tâche donnée, et selon Zipf, la plupart des comportements humains peuvent s'expliquer par ce principe. Dans le cas du langage, cela consiste à minimiser l'effort intellectuel nécessaire à l'élaboration et à la compréhension du message, aussi bien de la part du locuteur (ou de l'auteur) que de l'auditeur (ou du lecteur). Pour le locuteur, l'idéal serait d'avoir un langage qui ne soit composé que de peu de mots, afin de minimiser l'effort de mémoire nécessaire à la composition du message. Quant à ce que chacun de ces mots ait de nombreux sens différents. Pour l'auditeur, l'idéal serait au contraire que chaque mot ait un unique sens très précis, afin de limiter l'effort d'interprétation nécessaire à la compréhension. Afin de respecter dans les deux cas le principe du moindre effort, le langage naturel tend vers l'équilibre entre ces deux exigences contradictoires, ce qui se traduit par l'usage d'un nombre réduit de mots-outils utilisables dans la plupart des situations, ainsi qu'un grand nombre de mots moins fréquents qui ne sont utilisés que lorsque la situation exige une description précise. C'est cela qui, selon Zipf, explique que la répartition observée des fréquences des mots suive une loi puissance. Au-delà de la simple répartition des fréquences d'apparition des mots, d'autres faits linguistiques remarqués par Zipf tendent à prouver que l'évolution linguistique semble être dictée par le principe du moindre effort. En particulier, Zipf a également remarqué que les mots les plus fréquents sont également les plus courts, c'est à dire ceux qui comportent le moins de phonèmes, et que la fréquence des mots décroît proportionnellement au nombre de syllabes. On peut également remarquer que les mots les plus fréquents sont en général les plus anciennement apparus dans la langue. Ces faits sont particulièrement flagrants en anglais, en effet cette langue possède un double vocabulaire : les mots de base qui proviennent du vieil anglo-saxon, d'usage très fréquent mais relativement peu nombreux, et les mots plus savants qui sont pour la plupart d'entre eux des emprunts plus récents à la langue française.

La loi du moindre effort est également responsable d'autres évolutions linguistiques comme le raccourcissement de la longueur des mots, l'usage d'abréviations, la disparition des déclinaisons, l'emploi des temps composés à la place des temps simples...

Cependant, l'interprétation donnée par Zipf de sa loi n'est pas totalement satisfaisante. En effet, des distributions en loi puissance ont été mises en évidence non seulement dans des textes écrits par des humains, mais aussi dans des textes générés aléatoirement, notamment par Miller [Mil-57] ou plus récemment par Li [Li-92]. Or l'interprétation de Zipf est basée sur le comportement humain et ne permettrait pas de rendre compte de tous les phénomènes dans lesquels on rencontre des lois puissance. C'est pourquoi d'autres interprétations ont été proposées qui permettent d'expliquer l'existence des lois puissance par des modèles mathématiques probabilistes. C'est le cas en particulier du modèle proposé en 1955 par Herbert Simon [Sim-55]. Cette interprétation est basée sur un modèle stochastique de la croissance du vocabulaire. D'après ce modèle, les mots nouveaux sont introduits dans un texte suivant une probabilité constante. De plus, la probabilité qu'un mot soit apparu m fois dans un texte est proportionnelle au nombre total d'occurrences des mots qui sont apparus exactement m fois. Ces hypothèses permettent de modéliser l'évolution de la fréquence des mots par un processus de croissance, ce qui, selon Simon, conduit à une répartition en loi puissance. Le modèle de Simon a inspiré d'autres interprétations de distributions en lois puissance dans d'autres domaines.

A titre d'illustration, voici deux exemples de la répartition des fréquences d'apparition des mots en français et en anglais. Les résultats proviennent pour le français d'une étude réalisée

par le ministère de l'Éducation Nationale sur des textes de la littérature française moderne et pour l'anglais du British National Corpus qui comprend des textes de différentes origines. Pour le français, toutes les formes verbales ont été ramenées à l'infinitif et les accords en genre et en nombre ont été éliminés sauf pour les pronoms personnels. La méthode de comptage utilisée distingue entre les différentes catégories grammaticales et des listes de fréquences par catégorie grammaticale sont également disponibles. La figure 3 montre les courbes de Zipf obtenues pour les 1500 mots les plus fréquents dans les deux langues.

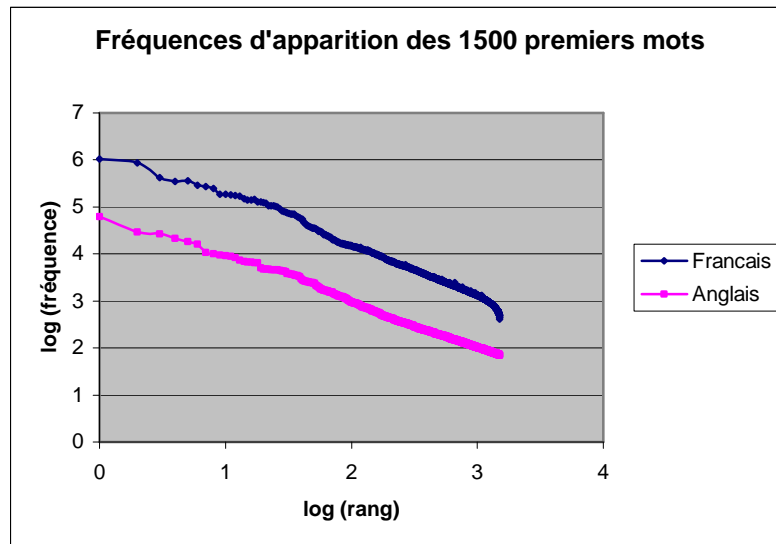


Fig .3 : Fréquence d'apparition des 1500 mots les plus fréquents en français et en anglais.

Dans les deux cas, on remarque que la courbe rang-fréquence représentée dans un repère bi-logarithmique est à peu près linéaire. Néanmoins, les points ne sont pas parfaitement alignés. La différence entre les ordonnées des deux courbes est due essentiellement à la différence de la taille du corpus utilisé pour chacune des deux langues. La pente de la courbe est de $-1,128$ pour le français et de $-0,9682$ pour l'anglais. On retrouve donc les résultats obtenus par Zipf qui prévoient une valeur de l'exposant α proche de 1 dans les langues naturelles.

1.4 La loi de Zipf inverse

En plus de la loi rang-fréquence déjà mentionnée, une autre loi puissance a été mise en évidence par Zipf à partir de ses travaux sur la langue anglaise. Cette seconde loi est nettement moins connue que la précédente, bien que Zipf l'ait en fait découverte dès 1935. Il s'agit de la loi dite *loi de Zipf inverse*. Contrairement à la loi de Zipf précédemment définie qui décrit la répartition statistique des mots, la loi de Zipf inverse décrit la répartition statistique des fréquences des mots. Selon la loi de Zipf inverse, le nombre de mots *distincts* ayant une fréquence d'apparition donnée dans un texte suit une loi puissance. Le nombre $I(f)$ de mots apparaissant f fois dans le texte est donné par la formule suivante :

$$I(f) = A f^{-\gamma} \quad (1-4-1)$$

Dans cette formule, A et γ sont des constantes et la valeur de l'exposant γ est proche de 2 pour les textes écrits en langage naturel. En fait, cette loi empirique n'est valable que pour les mots dont la fréquence d'apparition dans le texte est faible. Contrairement à la loi de Zipf qui

s'intéresse davantage aux mots les plus fréquents, la loi de Zipf inverse permet donc d'étudier la répartition des mots les plus rares dans le texte. De ce fait, la loi de Zipf inverse permet d'analyser la richesse et la diversité du vocabulaire d'un texte. Ce modèle peut notamment être utilisé afin d'identifier l'auteur d'un texte. La loi de Zipf inverse a notamment été utilisée par Cohen et al. [Coh-97] afin de distinguer entre les textes naturels et artificiels. Elle est nettement plus efficace que la loi de Zipf pour cette application. La fig.4 montre un exemple d'application de cette loi. On remarque que la partie gauche de la courbe est une droite, ce qui montre que l'on a bien une loi puissance. L'allure de la partie droite s'explique par le fait que pour les grandes valeurs de la fréquence, il n'y a pas de mot associé à chaque valeur.

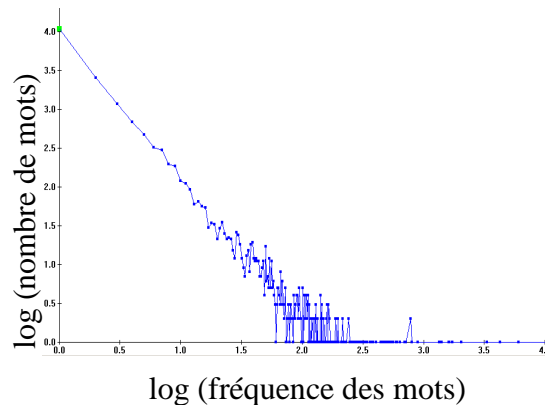


Fig. 4 : Représentation graphique de la loi de Zipf inverse, dite « courbe de Zipf inverse »

La taille du vocabulaire, c'est à dire le nombre de mots différents apparaissant dans le texte, est également fonction de la longueur de celui-ci. Herdan [Her-66] a montré que la croissance de la taille du vocabulaire en fonction de la longueur du texte suit également une loi puissance. Le nombre R de mots différents dans un texte de longueur T est donné par la formule :

$$R = BT^\beta \quad (1-4-2)$$

B et β sont des constantes. La valeur de ces constantes doit permettre d'identifier l'auteur du texte. Bien que β puisse être considéré comme constant sur une grande plage de variation de la longueur des textes, dans la pratique il décroît en fait très lentement en fonction de la longueur du texte.

Un lien peut être fait entre les lois puissance et la notion d'entropie. L'entropie au sens de la théorie de l'information peut être définie à partir des mots du texte selon la formule suivante :

$$\tilde{H}_w = - \sum_{r=1}^R \frac{f(r)}{T} \log_R \frac{f(r)}{T} \quad (1-4-3)$$

Dans cette formule, T représente la longueur du texte, R le nombre de mots distincts et $f(r)$ la fréquence d'apparition du mot r. On utilise ici un logarithme à base R, de manière à permettre de comparer les résultats obtenus avec des textes différents, il permet de donner une valeur de l'entropie comprise entre 0 et 1 quelle que soit la valeur de R. Cette entropie est maximale quand tous les mots du texte ont la même fréquence d'apparition, elle est minimale quand la fréquence relative d'un des différents mots atteint 1. L'entropie permet donc de mesurer l'uniformité de la distribution des différents mots.

Il est maintenant possible de définir une entropie non plus à partir des *mots* du texte mais des *fréquences d'apparition* de ceux-ci. La formule suivante a été proposée par Cohen [Coh-97] afin de calculer l'*entropie relative aux fréquences*.

$$\tilde{H}_f = - \sum_{f=1}^F \frac{I(f)}{R} \log_F \frac{I(f)}{R} \quad (1-4-4)$$

Dans cette formule, $I(f)$ représente le nombre de mots distincts ayant une fréquence d'apparition égale à f et F représente le nombre total d'occurrences des mots dans le texte. L'utilisation d'un logarithme à base F assure que la valeur de l'entropie est comprise entre 0 et 1. L'intérêt de cette définition de l'entropie est, comme pour la loi de Zipf inverse qui est également définie à partir de la fréquence d'apparition des mots, de donner un poids plus important aux mots peu fréquents dans le texte.

Les lois de Zipf et de Zipf inverse ainsi que les entropies qui leur sont associées ont été utilisées par Cohen et al. afin d'analyser les différences entre des textes naturels, c'est à dire écrits par des auteurs humains, et des textes artificiels générés aléatoirement. L'analyse des textes aléatoires avec la loi de Zipf donne des résultats peu différents de ceux obtenus avec des textes naturels, comme l'a déjà montré Miller en 1957 [Mil-57], à savoir que l'on obtient dans les deux cas une loi puissance avec un exposant proche de 1. Par contre, la croissance du vocabulaire en fonction de la taille du texte est beaucoup plus faible dans le cas des textes naturels. L'ordonnée à l'origine de la courbe de Zipf inverse est beaucoup plus élevée dans le cas d'un texte artificiel que pour un texte naturel. Cela signifie que les textes générés aléatoirement comprennent plus de mots rares que les textes naturels. On peut ainsi utiliser la loi de Zipf inverse afin de différencier l'artificiel du naturel.

2 Interprétation des lois puissance

2.1 le modèle stochastique de Simon

Un modèle mathématique a été proposé en 1955 par Herbert Simon [Sim-55] pour expliquer l'existence des distributions en lois puissance (bien que Simon n'ait pas employé ce terme), que ce soit pour les fréquences des mots, la population des villes, la distribution des revenus ou la classification des espèces vivantes. Un raisonnement analogue avait déjà été appliqué par Champernowne [Cha-53] pour expliquer la répartition des revenus suivant la loi de Pareto. Ces distributions peuvent être décrites par une formule du type :

$$f(i) = (a/i^k)b^i \quad (2-1-1)$$

L'exposant k est en général supérieur à 1 et le plus souvent proche de 2 et le rapport $f(2)/f(1)$ est généralement proche de $1/3$ et presque toujours inférieur à $1/2$.

Selon Simon, d'autres distributions comme les distributions en série logarithmique de Fisher et les distributions binomiales négatives peuvent être décrites par la formule $f(i) = (a/i^k)b^i$ pour des grandes valeurs de i mais k est alors inférieur à 1 et $f(2)/f(1)$ ne peut être inférieur à $1/2$.

Une distribution en série logarithmique de Fisher est une distribution dont la loi de probabilité est de la forme :

$$P(X=i) = \frac{1}{-\log(1-\gamma)} \frac{\gamma^i}{i} \quad (2-1-2)$$

avec i entier et γ compris entre 0 et 1. Une distribution binomiale négative a une loi de probabilité de la forme :

$$P(X=i) = C_n^i p^i (1-p)^{n-i} \quad (2-1-3)$$

avec n entier, i compris entre 0 et n et p entre 0 et 1. La figure ci-dessous montre les représentations graphiques de ces deux distributions.

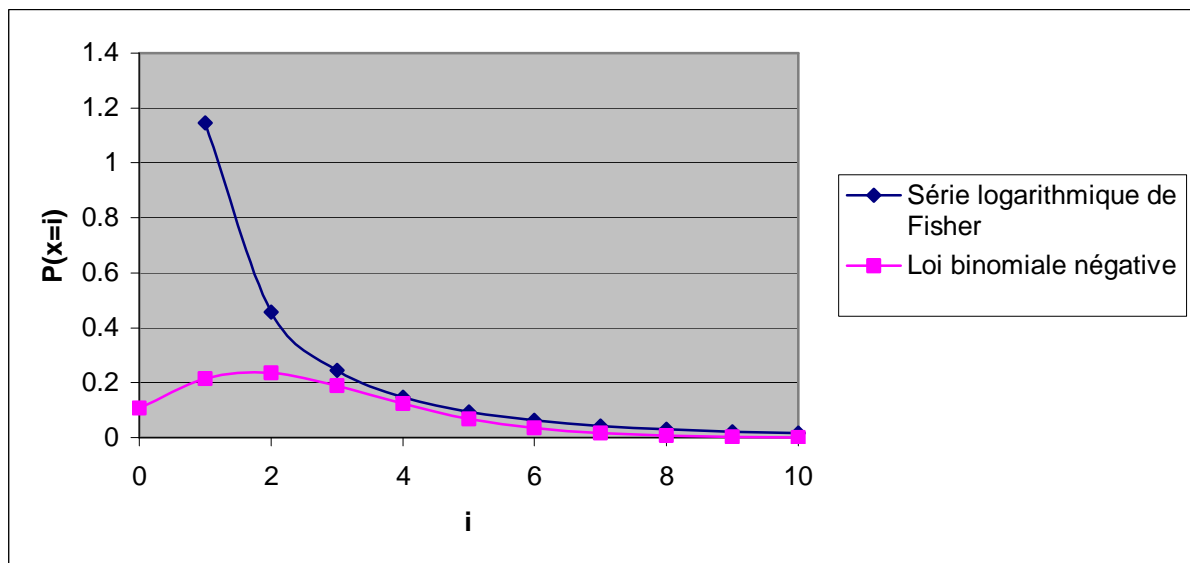


Fig. 2 : Représentation graphique d'une série logarithmique de Fisher et d'une loi binomiale négative, pour $n = 10$ et $p = \gamma = 0,8$.

J.H. Darwin a montré en 1953 [Dar-53] que ce type de distribution peut être obtenu par un processus stochastique de naissance et de mort. Un processus de naissance et de mort est un processus aléatoire qui décrit l'évolution d'une population X en fonction d'un taux de natalité λ et d'un taux de mortalité δ . Il est décrit par les formules suivantes :

$$P(X_{t+h}=n+1|X_t=n) = \lambda n h + o(h) \quad (2-1-4)$$

$$P(X_{t+h}=n-1|X_t=n) = \sigma n h + o(h) \quad (2-1-5)$$

$$P(X_{t+h}=n|X_t=n) = 1 - (\lambda + \sigma) n h + o(h) \quad (2-1-6)$$

Selon Simon, un processus stochastique similaire à ceux qui conduisent à une distribution binomiale négative ou en série logarithmique peut être décrit par une classe de fonctions du type :

$$f(i)=AB(i,\rho+1) \quad (2-1-7)$$

Où A et ρ sont des constantes et $B(i,\rho+1)$ est la fonction Beta calculée en i, et $\rho+1$, elle est donnée par la formule :

$$B(i,\rho+1)=\int_0^1 \lambda^{i-1}(1-\lambda)^\rho d\lambda=\frac{\Gamma(i)\Gamma(\rho+1)}{\Gamma(i+\rho+1)} \text{ avec } (0<i ; 0<\rho<\infty) \quad (2-1-8)$$

Or, une propriété de la fonction gamma est que lorsque i tend vers l'infini, on a :

$$\frac{\Gamma(i)}{\Gamma(i+k)} \sim i^{-k} \quad (2-1-9)$$

Par conséquent, quand i tend vers l'infini, on a :

$$f(i) \sim \Gamma(\rho+1) i^{-(\rho+1)} \quad (2-1-10)$$

Donc, une distribution de la forme $f(i)=AB(i,\rho +1)$ est, pour les grandes valeurs de i, l'approximation d'une loi puissance.

Simon a introduit son modèle en prenant l'exemple d'un texte en cours d'écriture. Le texte déjà écrit a une longueur de k mots, et on désigne par $f(i,k)$ le nombre de mots différents apparus i fois dans les k premiers mots du texte. On considère le mot qui est en train d'être écrit. Simon a admis les deux hypothèses suivantes : la probabilité que le (k+1)ème mot du texte soit un mot déjà apparu i fois est proportionnelle à $if(i,k)$, et il existe une probabilité constante α que le (k+1)ème mot soit un mot nouveau. La probabilité d'un mot dépend donc de ce qui a été écrit auparavant. On distingue deux cas : celui où le mot courant est un mot nouveau, c'est à dire que l'on a $i=1$, et le cas où ce mot est déjà présent dans le texte et $i > 1$. Bien qu'il aurait été plus naturel de calculer l'espérance de $f(i,k)$, Simon a préféré calculer l'espérance de $f(i,k+1)$. Dans le cas où $i=1$, si l'on suppose constante la probabilité α pour qu'un mot nouveau apparaisse, cette espérance est donnée par la formule suivante :

$$E\{f(1,k+1)\} - f(1,k) = \alpha - K(k)f(1,k) \quad (2-1-11)$$

De même, d'après les deux hypothèses faites par Simon, l'espérance de $f(i,k+1)$, dans le cas où le mot est déjà apparu dans le texte, est donné par la formule :

$$E\{f(i,k+1)\} - f(i,k) = K(k)\{ (i-1)f(i-1,k) - if(i,k) \} \text{ pour } (i=2,\dots,k+1) \quad (2-1-12)$$

Comme on suppose que l'on a des distributions statiques, dans les sens où toutes les fréquences d'apparition des mots croissent proportionnellement à k, on peut remplacer les espérances par les fréquences elles-mêmes. On obtient les équations suivantes :

$$f(1,k+1) - f(1,k) = \alpha - K(k)f(1,k) \quad (0 < \alpha < 1) \quad (2-1-13)$$

$$f(i,k+1) - f(i,k) = K(k)\{ (i-1)f(i-1,k) - if(i,k) \} \text{ pour } (i=2,\dots,k+1) \quad (2-1-14)$$

Pour $i=k+1$, on obtient $K(k)=1/k$.

On cherche à évaluer le facteur de proportionnalité $K(k)$. Comme $K(k)if(i,k)$ est la probabilité que le $(k+1)$ ème mot soit apparu i fois dans le texte, on a :

$$\sum_{i=1}^k K(k)if(i,k) = K(k) \sum_{i=1}^k if(i,k) = 1 - \alpha \quad (2-1-15)$$

Or, on sait que $\sum_{i=1}^k if(i,k) = k$, par conséquent on a

$$K(k) = \frac{1 - \alpha}{k} \quad (2-1-16)$$

On pourrait résoudre ces équations directement, mais Simon a proposé une autre méthode.

On considère que l'on a une distribution statique telle que pour toute valeur de i et de k on ait la relation suivante:

$$\frac{f(i,k+1)}{f(i,k)} = \frac{k+1}{k} \quad (2-1-17)$$

Toutes les fréquences d'apparition des mots croissent proportionnellement à k et maintiennent donc la même valeur relative. On peut en déduire que :

$$\frac{f(i,k)}{f(i-1,k)} = \frac{f(i,k+1)}{f(i-1,k+1)} = \beta(i) \quad (2-1-18)$$

où $\beta(i)$ est indépendant de k . Donc les fréquences relatives des mots les uns par rapports aux autres, que l'on désigne par $f^*(i)$ sont indépendantes de la longueur du texte k . Des formules précédentes on en déduit la relation suivante :

$$\left(\frac{k+1}{k} - 1\right) f(i,k) = \frac{(1-\alpha)}{k} \left(\frac{(i-1)}{\beta(i)} - i\right) f(i,k) \quad (2-1-19)$$

Et donc on a :

$$\beta(i) = \frac{(1-\alpha)(i-1)}{1+(1-\alpha)i} = \frac{f^*(i)}{f^*(i-1)} \quad (i=2, \dots, k) \quad (2-1-19)$$

Si l'on pose

$$\rho = \frac{1}{1-\alpha} \quad (1 < \rho < \infty) \quad (2-1-20)$$

et sachant que l'on a $f^*(i) = \beta(i)f^*(i-1) = \beta(i)\beta(i-1)\dots\beta(2)f^*(1)$ on en déduit que :

$$f^*(i) = \frac{(i-1)(i-2)\dots 2.1}{(i+\rho)(i+\rho-1)\dots(1+\rho)} f^*(1) = \frac{\Gamma(i)\Gamma(\rho+1)}{\Gamma(i+\rho+1)} f^*(1) = B(i+\rho+1)f^*(1) \quad \text{avec } (i=2, \dots, k) \quad (2-1-21)$$

en effet on a :

$$\Gamma(i+\rho+1) = (i+\rho)\Gamma(i+\rho) = (i+\rho)(i+\rho+1)\dots(1+\rho)\Gamma(\rho+1) \quad (2-1-22)$$

On remarque que la formule (2-1-21) est identique à (2-1-7) avec $A=f^*(1)$. Or, (2-1-7) est une solution stable de (2-1-12) et admet pour limite une loi puissance. Dans le cas où la formule (2-1-19) n'est pas satisfaite, on peut tout de même écrire que :

$$\frac{f(i,k)}{f(i-1,k)} = \frac{(1-\alpha)(i-1)}{(1-\alpha)i+1+\varepsilon(i,k)} \quad (2-1-23)$$

où $\varepsilon(i,k)$ est une fonction de i et k . Par conséquent on a :

$$\frac{f(i,k+1)}{f(i,k)} = \frac{k+1+\varepsilon(i,k)}{k} \quad (2-1-20)$$

Comme on suppose que les mots nouveaux sont introduits avec un taux constant, la somme des fréquences doit être proportionnelle à k , donc on doit avoir :

$$\sum_{i=1}^{k+1} f(i,k+1) - \frac{k+1}{k} \sum_{i=1}^{k+1} f(i,k) = \frac{1}{k} \sum_{i=1}^k \varepsilon(i,k) f(i,k) = 0 \quad (2-1-24)$$

Simon a interprété ces résultats de la façon suivante : les fréquences croissent en moyenne proportionnellement à k , si une fréquence particulière est plus grande que ce qu'elle devrait être elle va croître plus lentement, si elle est plus petite elle va croître plus rapidement. Ce qui fait que les deux phénomènes se compensent et qu'au final la distribution reste une loi puissance. Ces résultats sont valables si le taux d'introduction α des mots nouveaux reste constant. Maintenant, si le taux d'introduction de mots nouveaux décroît avec k , Simon a démontré que l'on doit observer une décroissance de $f(i,k)$ avec k , ce qui fait que la queue de la distribution est courbée vers le bas. Or, ce phénomène a effectivement été observé dans de très longs textes.

On peut retrouver le résultat de la formule (2-1-21) d'une autre manière. L'équation (2-1-20) peut être réécrite sous la forme :

$$0 = (1-\alpha)[(i-1)f^*(i-1) - if^*(i)] - f^*(i) \quad (i=2, \dots, k) \quad (2-1-25)$$

De même (2-1-13) peut être réécrit :

$$0 = 1 - (1-\alpha)f^*(1) - f^*(1) \quad (2-1-25)$$

Or, une solution des équations (2-1-24) et (2-1-25) est donnée par :

$$\frac{f^*(i)}{f^*(i-1)} = \frac{\lambda(i-1+c)}{\lambda(i+c)+1} = \frac{(i-1+c)}{(i+c+1/\lambda)} \quad (2-1-26)$$

avec $\lambda = k(1-\alpha)/(k+c)$ et c est une constante.

La relation (2-1-26) a une propriété très intéressante, à savoir la limite de l'expression est une loi puissance de la forme $f(i) = (a/i^k)b^i$ avec $b=1$. Donc, pour α et c petits, cette distribution, dite *distribution généralisée de Yule* possède les propriétés d'une loi puissance. Le fait que la forme de la distribution ne soit pas trop altérée par les variations dans les hypothèses de départ permet d'utiliser le modèle stochastique de Simon pour modéliser de nombreux phénomènes où on observe ce type de distribution, tels que la fréquence des mots, la distribution des revenus ou des espèces biologiques au sein des genres.

2.2 La loi de Zipf-Mandelbrot : interprétation par les arbres lexicographiques

Benoît Mandelbrot, surtout connu pour ses travaux sur les fractales, a également travaillé sur la modélisation des fréquences d'apparition des mots dans les textes au moyen des lois puissance. Il a proposé en 1953 [Man-53] une amélioration du modèle de Zipf connue depuis lors sous le nom de loi de Zipf-Mandelbrot. La nouvelle formulation introduit des paramètres supplémentaires afin de mieux correspondre aux données du langage naturel. Selon la loi de Zipf-Mandelbrot, la fréquence $F(r)$ du mot de rang r est donnée par la formule suivante :

$$F(r) = (r + \rho)^{-(1+\alpha)}$$

Dans cette nouvelle formule, ρ et α sont des constantes, α est compris entre 0 et 1 et ρ est supérieur à 1. Mandelbrot a interprété les distributions en lois puissance comme une optimisation au sens de la théorie de l'information. On suppose que le lexique comporte n mots. Le coût d'usage du mot d'indice j de ce lexique est C_j . Ce coût peut par exemple représenter la longueur du mot. La fonction coût est alors du type $C_j \sim \log_N(j)$ pour un alphabet de N caractères et on cherche à optimiser l'information moyenne donnée par chaque mot.

Mandelbrot a proposé une interprétation mathématique de la loi de Zipf basée sur la méthode des *arbres lexicographiques*. [Man-53], [Man-83]. Le lexique est constitué de toutes les suites des lettres qui constituent des mots, séparées par des blancs que l'on peut imaginer placés au début de chaque mot. Mandelbrot a imaginé de représenter le lexique par un arbre, dont le blanc constitue le tronc. Au niveau 1, qui correspond à la première lettre du mot, l'arbre se subdivise en N branches, autant qu'il y a de lettres de l'alphabet, il en est de même pour tous les niveaux suivants. Chaque mot peut donc être représenté par un chemin dans cet arbre. A chaque chemin possible dans l'arbre est associé un poids qui correspond à la probabilité d'apparition du mot. Ce poids est nul pour les séquences de lettres qui ne sont pas admissibles comme mots. Si l'on veut que le codage des mots soit *optimal*, c'est à dire que le nombre moyen de lettres soit le plus petit possible, alors l'arbre lexicographique doit être *régulier*, c'est à dire que chaque chemin possible dans l'arbre doit correspondre à un mot, et que les poids des probabilités au niveau k soient de la forme $U = U_0 r^k$, où r est une constante entre 0 et 1. Si l'on classe les mots par ordre décroissant de leur fréquence d'apparition et si l'on désigne par ρ le rang dans la séquence d'un mot de probabilité U , alors en supposant l'arbre lexicographique régulier, cette probabilité est donnée par la formule suivante :

$$U = P(\rho + V)^{-1/D} \quad (2-3-1)$$

Par conséquent, on a :

$$\rho = -V + U^{-D} P^D \quad (2-3-2)$$

P , V et D étant des constantes. Cette formule obtenue de manière analytique par Mandelbrot se trouve être une généralisation de la loi empirique de Zipf.

La quantité D représente la dimension suivante :

$$D = \log N / \log(1/r) \quad (2-3-2)$$

N représente ici le nombre d'éléments de l'alphabet. On peut mesurer la richesse du vocabulaire par la fréquence relative des mots rares, par exemple par le rapport entre la fréquence du mot de rang 100 et celle du mot de rang 10. A N constant, cette richesse augmente avec r . Plus D est grand et plus le vocabulaire est riche.

A partir de là, la loi de Zipf généralisée a pu être démontrée par Mandelbrot de la façon suivante : au niveau k, ρ varie entre $1+N+N^2+\dots+N^{k-1}=(N^k-1)/(N-1)$ et $(N^{k+1}-1)/(N-1)$ Soit V égal à $1/(N-1)$. Insérant $k=\log(U/U_0)/\log r$ dans ces deux bornes, on trouve

$$(U^D U_0^D)^{-1} < \rho/V \leq N(U^D U_0^D)^{-1} \quad (2-3-4)$$

On obtient le résultat annoncé et on définit la constante P en approximant ρ par la moyenne de ses bornes. Bien que dans la pratique, l'arbre lexicographique ne soit pas régulier, on peut ainsi conclure que la loi de Zipf généralisée est le résultat auquel on aurait dû s'attendre. La dimension D a été définie par Mandelbrot comme étant la *température du discours*. La valeur $D=1$ joue un rôle très spécial qui est dû au fait que l'on a :

$$P^{-1} = \sum_{\rho=(N^k-1)/(N-1)}^{(N^{k+1}-1)} (\rho+V)^{-1/D} \quad (2-3-5)$$

Lorsque $D \geq 1$, la série (2-3-5) diverge. Il est donc nécessaire que ρ soit borné, c'est à dire que le lexique doit être fini. Lorsque $D < 1$, le lexique peut très bien être infini. Dans ce cas, U_0 prend la forme $1 - Nr$ et est inférieur ou égal à 1. On peut alors interpréter U_0 comme la probabilité de l'espace entre les mots et r la probabilité d'une des lettres de l'alphabet. Dans ce cas, on peut considérer que l'on a une suite infinie de blancs et de mots statistiquement indépendant, et l'on peut utiliser les blancs pour découper la suite en mots. Les probabilités des mots suivront la loi de Zipf généralisée ou loi de Zipf-Mandelbrot.

Dans le cas où $D < 1$, l'arbre lexicographique peut être interprété géométriquement sur l'intervalle fermé $[0,1]$. On définit N intervalles ouverts de longueur r qui sont associés aux N lettres de l'alphabet et un intervalle ouvert de longueur $U_0=1-Nr$ associé au blanc. Chaque intervalle correspondant à une lettre sera ensuite divisé en N intervalles « lettre-lettre » et un intervalle « lettre-blanc », l'intervalle « blanc » n'étant pas subdivisé. On remarque que chaque intervalle blanc définit une suite de lettres finissant par un blanc. Par conséquent, il définit un mot, l'un des mots étant réduit à un blanc. Si on le représente précédé d'un point décimal, un mot peut être considéré comme un nombre réel écrit dans la base (N+1) ne contenant aucun zéro intermédiaire. Si l'on représente ces nombres dans l'intervalle $[0,1]$ et que l'on ajoute les bornes de cet ensemble, on construit l'ensemble des nombres réels compris entre 0 et 1 privé des nombres qui contiennent un zéro autre part qu'en position finale. Or, cet ensemble est un ensemble de Cantor dont la dimension fractale est D. La quantité D que l'on a défini comme la température du discours est donc une dimension fractale. Ainsi, on peut établir un lien entre la loi de Zipf généralisée et la notion de dimension fractale, la puissance de la loi puissance étant équivalente à une dimension fractale. Le facteur V défini précédemment correspond alors à la lacunarité fractale.

Le même raisonnement a été appliqué par Mandelbrot au cas de la loi de Pareto. On s'intéresse ici à la hiérarchie des revenus salariaux. On suppose que les salariés d'une entreprise sont classés par niveau hiérarchique. La hiérarchie est régulière si, sauf au niveau le plus bas, chaque salarié a le même nombre N de subordonnés et chaque salarié a un salaire U égal à r fois le salaire de son supérieur immédiat. Ici, r est bien sûr compris entre 0 et 1. Les revenus non salariaux ne peuvent pas être modélisés de cette manière car ils ne comportent aucune hiérarchie pouvant être représentée par un arbre. On classe les individus par ordre décroissant de revenu et on désigne par r leur rang dans la suite. Comme pour les fréquences des mots, on peut montrer que le rang de l'individu touchant le salaire U peut être exprimé par la formule : $r = -V+U^D P^D$. La dimension D représente ici le degré d'inégalité des revenus.. Si

l'on suppose que le revenu peut être différent entre les différents individus d'un même niveau hiérarchique, cela modifie les expressions de V et P mais laisse inchangée celle de D. La loi de Pareto peut s'appliquer également aux revenus spéculatifs, mais avec une valeur différente pour D. La valeur empirique de D est en général proche de 2.

L'interprétation des distributions en loi puissance a donné lieu à un débat passionné entre Mandelbrot et Simon à la fin des années 50. Le débat s'est terminé sans véritable vainqueur, chacun restant sur ses positions. L'interprétation de Simon présente l'avantage de pouvoir s'appliquer à davantage de phénomènes, certains phénomènes dans lesquels on observe des lois puissance, comme la croissance des villes sont difficilement modélisables par des arbres lexicographiques. Les travaux modernes sur les lois puissance du type de la loi de Zipf se basent le plus souvent sur des modèles stochastiques proches de celui proposé par Simon. Cependant, les travaux de Mandelbrot sur les lois puissance ont donné naissance à la théorie des fractales.

2.3 Les différents processus aboutissant à une loi puissance

Selon Reed et Hughes [Ree-02], la plupart des distributions en loi puissance que l'on peut rencontrer, dans quelque domaine que ce soit, ont pour origine des processus stochastiques de croissance exponentielle que l'on observe « tués », c'est à dire dans un état arrêté, à un instant aléatoire par rapport à l'évolution du processus. Reed et Hughes ont déterminé quatre familles de processus stochastiques qui peuvent aboutir à des lois puissance. La plupart des lois puissance observées dans la nature sont la conséquence de l'un ou l'autre de ces processus.

On considère que la date à laquelle ces processus sont tués est aléatoire. Et l'on considère que la probabilité pour que le processus soit tué est constante au cours du temps.

Le premier de ces processus est le mouvement brownien géométrique, où dB_t a une distribution normale avec une moyenne nulle et une variance dt . Il est décrit par la formule :

$$dX = \mu X dt + \sigma X dB_t, \quad E(X_t/X_0) = X_0 \exp(\mu t) \quad (2-3-1)$$

Ce processus conduit à une distribution en double loi de Pareto, c'est à dire que l'on a deux distributions en loi puissance comme expliqué au paragraphe précédent. La densité de probabilité de cette distribution est indiquée par la formule suivante :

$$f_{\bar{x}}(x) = \begin{cases} AX_0^{-\beta} x^{\beta-1} & \text{si } x \leq X_0 \\ AX_0^{\alpha} x^{-\alpha-1} & \text{si } x > X_0 \end{cases} \quad (2-3-2)$$

Le mouvement brownien est responsable de la répartition en loi puissance des revenus personnels et de la population des villes. On suppose que les revenus et les villes évoluent selon un mouvement brownien, et que le temps pendant lequel l'individu a travaillé ou que la ville a existé peut être modélisé par une distribution exponentielle.

Un autre processus pouvant aboutir à une loi puissance est le processus multiplicatif discret, avec $\{Z_n\}$ variables aléatoires indépendantes identiquement distribuées de moyenne μ . Il est décrit par la formule suivante :

$$X_{n+1} = Z_n X_n, \quad E(X_t/X_0) = X_0 \mu^t \quad (2-3-3)$$

Ce processus peut exister sous trois formes, monotone croissant, monotone décroissant ou bidirectionnel. S'il est monotone croissant, c'est à dire que $P(Z_n > 1) = 1$, on observe une distribution en loi puissance pour les valeurs supérieures. S'il est monotone décroissant, à savoir que $P(Z_n < 1) = 1$, on a une distribution en loi puissance pour les valeurs inférieures. Si le processus est bidirectionnel, on a une distribution en loi puissance aussi bien pour la partie supérieure que pour la partie inférieure. Le processus multiplicatif a notamment été utilisé par Mitzenmacher [Mit-01] pour expliquer la distribution de la taille des fichiers dans un système informatique.

Le processus suivant est le processus de naissance et de mort, avec les taux de natalité et de mortalité λ et δ . Il s'exprime par les formules suivantes :

$$P(X_{t+h} = n+1 | X_t = n) = \lambda n h + o(h) \quad (2-3-4)$$

$$P(X_{t+h} = n-1 | X_t = n) = \sigma n h + o(h) \quad (2-3-5)$$

$$P(X_{t+h} = n | X_t = n) = 1 - (\lambda + \sigma) n h + o(h) \quad (2-3-6)$$

Ce processus a été utilisé pour expliquer la distribution des espèces vivantes qui, d'après Yule [Yul-25] suit une loi puissance. On suppose ici que les nouvelles espèces apparaissent à partir d'espèces existantes de manière aléatoire et indépendante. De la même manière on suppose que les extinctions d'espèces sont aléatoires et indépendantes les unes des autres. Les genres apparaissent de la même manière que les espèces. Selon ce modèle, la distribution des espèces au sein des genres est celle d'un processus de naissance et de mort tué aléatoirement. On observe également une répartition en loi puissance pour les espèces fossiles lors des extinctions massives, en effet les extinctions massives représentent un cas où le processus normal de l'évolution des espèces a été tué à un moment donné.

Le dernier processus décrit par Reed et Hughes est le processus de branchement de Galton-Watson, décrit par la formule ci-dessous :

$$X_{n+1} = Z_1 + Z_2 + \dots + Z_{X_n}, \quad E(X_t | X_0) = X_0 \mu^t \quad (2-3-7)$$

avec $\{Z_i\}$ variables aléatoires indépendantes identiquement distribuées qui représentent le nombre de descendants et $E(Z_i) = \mu$. Le branchement de Galton-Watson permet de modéliser la répartition des fréquences des noms de famille, qui selon Reed et Hughes suit également une loi puissance.

Les différents processus décrits par Reed et Hughes permettent d'expliciter de nombreuses répartitions en lois puissance constatées dans divers domaines. Cependant, il y a une omission importante dans leurs travaux, celle de la répartition des mots des textes.

3 Applications de la loi de Zipf

3.1 Loi de Zipf et langage naturel : extension aux groupes de mots

La première application de la loi de Zipf, telle qu'elle a été vue au paragraphe 2-1, concerne la fréquence des mots dans le langage naturel. La plupart des travaux effectués dans ce domaine jusqu'à présent ont porté sur des mots isolés, mais on peut également essayer de voir ce qu'il en est pour des groupes de mots, voire des phrases. Des travaux dans ce domaine ont été effectués dans ce sens par Le Quan Ha et al. [LeQ-02] sur des textes en anglais et en chinois.

Ces travaux ont été effectués sur des corpus comprenant jusqu'à 19 millions de mots. Sur des corpus aussi longs, les courbes de Zipf observées ne sont pas linéaires, mais sont nettement incurvées vers le bas, à partir du 5000^e mot pour l'anglais et du 1000^e mot pour le chinois. Par contre, si l'on calcule la distribution des fréquences de n-uplets de 2 à 5 mots, on retrouve une répartition proche d'une loi puissance dans les deux langues, y compris dans les longs textes.

Et si l'on combine les fréquences des mots isolés et des n-uplets de 2 à 5 mots dans une même courbe de Zipf, en mettant dans une seule liste les mots et les n-uplets de 2 à 5 mots avec leurs fréquences d'apparition et en triant cette liste par ordre décroissant de fréquence, on obtient des résultats très proches de la théorie, à savoir que la courbe est une droite de pente -1 quelle que soit la longueur du corpus.

L'interprétation de ce fait serait que les nouvelles idées qui apparaissent dans un texte quand sa longueur augmente sont exprimées par de nouvelles *combinaisons* de mots plus souvent que par des mots nouveaux, ce qui tend à augmenter la fréquence des mots qui sont utilisés dans ces combinaisons. Ce fait est encore plus flagrant en chinois, en effet les auteurs ont considéré que chaque idéogramme chinois constitue un mot, indépendamment de sa valeur sémantique. Or, le nombre de caractères chinois est moins important que le nombre de mots différents dans un texte en anglais, de plus beaucoup d'idéogrammes chinois ne forment pas une unité sémantique complète et n'ont de sens qu'à l'intérieur d'un mot composé. En ce qui concerne le chinois, d'autres recherches ont été effectuées par Lua [Lua-94] avec une définition différente de la notion de mot ; ici le mot a été défini comme un ensemble de un ou plusieurs caractères formant une unité sémantique. Et dans ce cas, on trouve une distribution des fréquences des mots conforme à la loi de Zipf, ce qui semble confirmer cette interprétation.

3.2 Loi de Zipf et population des villes

En plus de la répartition des fréquences des mots du langage, une autre distribution en loi puissance a été mise en évidence par Zipf [Zip-49] dans un contexte entièrement différent : celui de la population des villes d'un pays. Si l'on classe les villes par ordre décroissant de population, la fonction de répartition de la population en fonction du rang de chaque ville dans la séquence suit une loi puissance. Ici aussi, l'exposant de la loi puissance est proche de 1. Cette loi a été vérifiée par Zipf dans le cas des villes des Etats-Unis. Historiquement, une telle constatation avait déjà été faite par Auerbach en 1913 [Aue-13]. D'autres travaux plus récents ont confirmé l'existence d'une telle distribution, notamment ceux de Gabaix [Gab-99] et Ioannides et Overman [Ioa-01]. On retrouve une telle répartition en loi puissance dans la plupart des pays, à l'exception des plus petits. Plusieurs interprétations ont été proposées pour expliquer cette loi. La plupart d'entre elles se basent sur la modélisation de la croissance des villes par un processus aléatoire, selon le même principe que celui employé par Simon pour les textes.

Selon Gabaix, la croissance des villes peut être modélisée par la loi de Gibrat [Gib-31]. Si les différentes villes croissent aléatoirement avec le même taux de croissance et la même variance, alors la distribution de la taille des villes doit tendre vers la loi de Zipf, c'est à dire une loi puissance avec un exposant égal à -1 . Si l'on note S_i la taille normalisée de la $i^{\text{ème}}$ ville du pays, c'est à dire sa population divisée par la population urbaine totale, la distribution rang-taille $G(S)$ des taille normalisées doit tendre vers :

$$G(S) = \frac{a}{S^\zeta} \quad (3-2-1)$$

où a est une constante et $\zeta = 1$. Selon Gabaix, on obtient ce résultat si la croissance des villes obéit à la loi de Gibrat, c'est à dire si les taux de croissances des villes sont distribués identiquement indépendamment de la taille des villes. On peut le démontrer de la façon suivante : soit γ^i le taux de croissance de la ville i à la date t . La taille de la ville à la date $t+1$ est : $S_{t+1}^i = \gamma_{t+1}^i S_t^i$. Si les taux de croissance sont des variables aléatoires indépendantes distribués identiquement avec une densité $f(\gamma)$ et si l'on suppose que la somme des tailles normalisées reste constante, à savoir que $\int_0^\infty \gamma f(\gamma) d\gamma = 1$ alors l'équation de la variation des tailles normalisées est la suivante :

$$G_{t+1} = \int_0^\infty G_t \left(\frac{S}{\gamma} \right) f(\gamma) d\gamma \quad (3-2-2)$$

Une solution de cette équation est donnée par la loi de Zipf :

$$G(S) = a / S. \quad (3-2-3)$$

Maintenant, on considère que la loi de Gibrat n'est pas complètement respectée et que les villes croissent avec des taux de croissances moyens et des écarts types qui dépendent de la taille de la ville. Dans ce cas la variation de la taille de la ville suit l'équation différentielle :

$$\frac{dS_t}{S_t} = \mu(S_t) dt + \sigma(S_t) dB_t \quad (3-2-4)$$

où μ et σ représentent la moyenne et l'écart type du taux de croissance et B_t est un mouvement brownien géométrique. Dans ce cas, la distribution de la taille des villes converge vers une loi puissance avec un exposant local $\zeta(S)$ qui peut être donné par la formule suivante :

$$\zeta(S) = 1 - 2 \frac{\mu(S)}{\sigma^2(S)} + \frac{\partial \sigma^2(S) / \sigma^2(S)}{\partial S / S}$$

Selon Gabaix, si la croissance urbaine des différentes régions d'un pays est différente, mais si la répartition de la population suit une loi puissance dans chaque région avec un exposant différent, alors la répartition totale doit suivre une loi puissance. Donc on peut parfaitement avoir un exposant local $\zeta(S)$ différent selon les villes et avoir une répartition en loi de Zipf à l'échelle du pays. Le modèle a été testé sur la population urbaine des Etats-Unis. La figure 5 montre une représentation graphique de cette répartition. On remarque que la distribution de la population suit la loi de Zipf, mais que l'on observe des différences par rapport au modèle pour les populations des villes petites et moyennes. Cela s'expliquerait par le fait que ces villes ont des variations de leurs taux de croissance nettement plus importants que les grandes villes.

Une autre variation par rapport au modèle que l'on peut observer pour certains pays dont la France, est « l'effet capitale », à savoir que la capitale, lorsqu'elle est la plus grande ville du pays, a une population beaucoup plus importante que celle prévue par le modèle en loi puissance. En effet, dans ce cas, comme la capitale est le principal centre économique et politique du pays, elle a un pouvoir attracteur beaucoup plus important que les autres villes, ce qui fait qu'elle concentre davantage de population. Ce fait ne s'observe pas aux Etats-Unis où la capitale est une ville relativement petite sans grande importance économique.

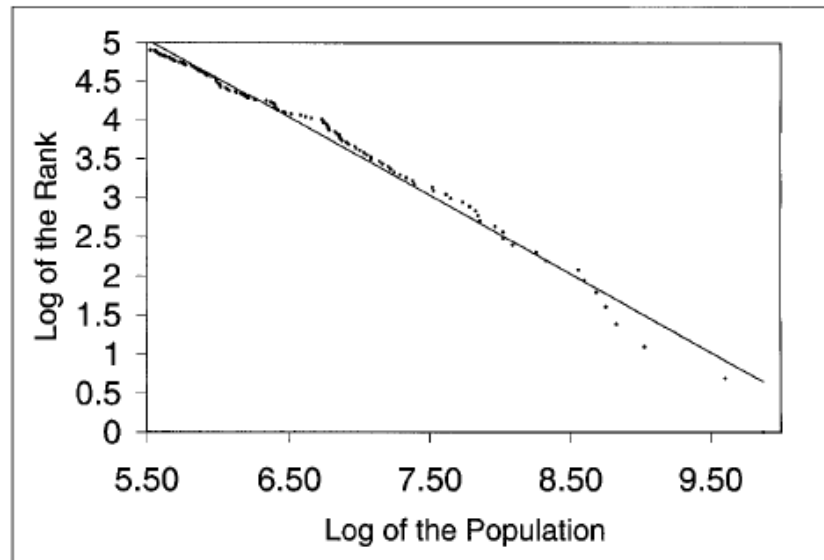


Fig .5 : Distribution de la population des principales villes aux USA (Source : [Gab-99])

3.3 La double loi de Pareto

Une interprétation des distributions en loi puissance donnée par Reed [Ree-01] est basée sur le fait qu'il s'agit d'observations de *processus stochastiques* de croissance exponentielle « tués » aléatoirement, c'est à dire observées « à l'arrêt » à un instant donné. Cette interprétation permet de mettre en évidence des distributions en *double loi de Pareto*, à savoir que l'on peut obtenir des distributions qui suivent une loi puissance à la fois pour les valeurs les plus élevées si on les classe par ordre décroissant et pour les valeurs les plus faibles si on les classe par ordre croissant. Ce type de distribution s'observe dans des phénomènes, comme la population des villes ou la répartition des revenus, qui ont une évolution au cours du temps qui peut être modélisée par le mouvement brownien géométrique selon la formule suivante :

$$dX = \mu X dt + \sigma X dw \quad (3-3-1)$$

Dans cette formule, dw est un bruit blanc ou l'incrément d'un processus de Wiener. L'incrément proportionnel X dans le temps dt est constitué d'une composante systématique μdt et d'une composante aléatoire σdw et le mouvement brownien géométrique peut être considéré comme une version stochastique d'un processus de croissance exponentielle.

Or, l'état d'un tel processus après un temps fixé T suit une distribution lognormale et non une loi puissance. Selon Reed, si l'on observe une loi puissance, c'est que l'instant d'observation T est lui aussi une variable aléatoire dont la distribution suit souvent une loi proche d'une loi exponentielle. Si l'on prend l'exemple de la répartition des revenus dans un échantillon de

population, et si l'on suppose que le revenu de chaque individu peut être modélisé par un mouvement brownien, le temps pendant lequel ce revenu a suivi un tel processus varie d'un individu à l'autre. (On suppose que la population considérée comprend des personnes d'âge différent, entrées à différentes dates dans la vie active). Si l'on suppose pour simplifier que tous les individus ont le même salaire de départ, la distribution des revenus est modélisée par un mouvement brownien observé après un temps T distribué exponentiellement. Cette distribution est appelée double distribution de Pareto. La densité de probabilité de cette distribution est proportionnelle à $x^{-\alpha+1}$ pour $x>1$ et à $x^{-\beta-1}$ pour $x<1$. Or, dans la réalité, le revenu des individus à leur entrée dans la vie active est variable. On pourrait, par exemple, le modéliser par un autre mouvement brownien. De même l'évolution de ce revenu au cours de la carrière professionnelle est aléatoire. Dans ce cas, le revenu d'un individu à une date donnée peut être représenté comme le produit d'une variable aléatoire suivant une double distribution de Pareto et d'une variable aléatoire lognormale indépendante. Ce modèle présente alors une distribution en loi puissance non seulement si l'on s'intéresse aux hauts revenus (comme avec le modèle classique de Pareto) mais aussi pour les revenus les plus bas. Le même raisonnement s'applique à la distribution de la taille des villes. Selon Reed, la croissance des villes peut elle aussi être expliquée par un processus exponentiel. De ce fait, la distribution des populations des différentes agglomérations peut également être modélisée par une double loi de Pareto. Des exemples de doubles lois de Pareto sont donnés sur la figure suivante :

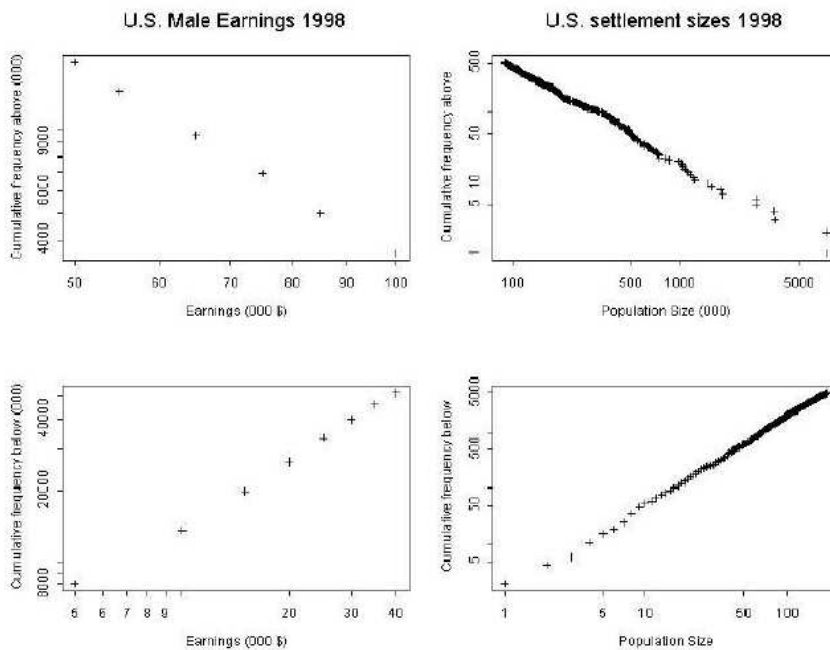


Fig. 6 : Doubles distributions de Pareto pour les revenus personnels et la taille des villes aux USA (source : [Ree-01])

3.4 La loi de Zipf et Internet

De nombreux travaux récents ont montré l'existence de distribution en lois puissance sur Internet. On peut citer en particulier les travaux de Glassmann[Gla-94], Cunha et al. [Cun-95] ou Breslau et al. [Bre-99].

Glassmann a utilisé les propriétés de la loi de Zipf afin de concevoir un cache pour un relais sur le réseau. Ce cache est destiné à stocker les pages Web les plus fréquemment consultées afin d'accélérer leur chargement. La conception de ce cache a été faite en fonction d'études effectuées sur le trafic existant. Ces études ont montré que la fréquence des consultations des pages Internet peut être modélisée par la loi de Zipf, avec toutefois quelques réserves dues au fait que le nombre de pages Web disponibles croît rapidement.

Les travaux de Cunha et al [Cun-95]. cherchent à étudier la manière dont les utilisateurs utilisent le Web. Ils sont basés sur l'analyse de l'historique des pages visitées dans une salle Internet d'une université. Les études ont montré que si l'on classe les pages par ordre décroissant de nombre de consultation, la fréquence de consultation des pages en fonction de leur rang suit une loi puissance. De plus, l'exposant de cette loi puissance est proche de -1 . On peut donc dire que la fréquence de consultation des pages Web suit la loi de Zipf. De plus, les fichiers les plus consultés, à savoir les pages HTML, sont les plus petits, et selon Cunha et al., les utilisateurs tendent à préférer les documents les plus petits. Cela permettrait d'optimiser la conception des caches de manière à cacher en priorité les documents les plus petits. Il faut cependant remarquer que l'étude date de 1995 et que la conception des pages Web a évolué depuis, les pages Web actuelles étant nettement plus complexes que les anciennes, en particulier les pages de titre des sites commerciaux. Donc la stratégie proposée serait sans doute moins valide de nos jours.

Breslau et al. ont mis en évidence le fait que si la pente de la distribution des fréquences de consultation des pages peut varier, elle suit toujours une loi puissance. En revanche, ils n'ont pas mis en évidence de corrélation entre la fréquence de consultation des pages et leur taille ni leur fréquence de mise à jour. En revanche, on n'observe pas de distribution en loi puissance pour le nombre de requêtes aux serveurs.

La mise en évidence des distributions en loi de Zipf ont permis à Breslau et al. de concevoir un modèle de cache pour lequel les requêtes sont considérées comme indépendantes et sont réparties suivant une loi puissance. Avec un tel modèle, le « hit ratio » du cache, c'est à dire le pourcentage de pages demandées, présentes dans le cache, croît linéairement avec le logarithme de la taille du cache. Ce modèle permet d'évaluer les performances de différents algorithmes de remplacement des pages Web dans le cache.

Une distribution en loi puissance s'observe également pour la répartition du nombre de liens hypertextes par page Internet. Selon Mitzenmacher [Mit-01], la présence de cette distribution s'explique par un modèle *d'attachement préférentiel*. Le Web peut être modélisé par un graphe où les pages sont les sommets et les liens hypertextes sont les arcs. Le nombre de liens entrants et sortants d'une page est distribué suivant une loi de Zipf. Cela peut s'expliquer par le fait que les nouvelles pages tendent à être liées à des pages comportant déjà beaucoup de liens. Ce raisonnement est analogue à celui proposé par Yule [Yul-25] pour la distribution des espèces botaniques, à savoir que les nouvelles espèces qui apparaissent au cours de l'évolution ont le plus de chance d'appartenir à des genres comportant déjà beaucoup d'espèces.

3.5 La loi de Zipf en musique

La loi de Zipf peut être également observée dans le domaine de la musique. Voss et Clarke[Vos-75] ont mis en évidence le fait que les signaux sonores musicaux suivent la loi de Zipf. Ils ont étudié des morceaux de différents styles musicaux et ont découvert l'existence d'une loi puissance dans les variations de volume et de fréquences. A partir de là, ils ont conçu un générateur électronique de musique aléatoire délivrant un signal dont le spectre suit une loi puissance en $1/f$. Des essais ont montré que la musique qui suit la loi de Zipf est jugée beaucoup plus esthétique, harmonieuse et agréable à l'oreille par la plupart des auditeurs. Manaris et al. [Man-01] ont imaginé d'utiliser ces résultats afin de définir différentes métriques permettant de caractériser la musique au moyen de la loi de Zipf. On travaille ici sur différents morceaux de musique classique et de jazz ainsi que sur des musiques générées aléatoirement, codés sous forme de fichiers au format MIDI. On calcule les distributions rang-fréquence des différentes caractéristiques des notes, comme la hauteur, la durée, et les intervalles mélodiques et harmoniques. Les résultats montrent que ces différents paramètres suivent la loi de Zipf dans le cas des pièces musicales connues comme étant particulièrement esthétiques, mais qu'en revanche ce n'est pas le cas pour les musiques aléatoires. On remarque également que les paramètres de la loi de Zipf peuvent servir à caractériser un style musical ou un compositeur. Ces résultats semblent montrer que les lois puissance jouent un rôle dans la sensibilité musicale humaine. Les travaux de Manaris et al. [Man-01] ouvrent la voie à des applications dans le domaine de la classification et de l'indexation d'œuvres musicales, ainsi que de l'aide à la composition musicale.

3.6 Application de la loi de Zipf à l'étude de signaux acoustiques médicaux

Une nouvelle application de la loi de Zipf a été proposée par Dellandrea et al. [Del-02] pour l'étude de signaux sonores médicaux, à savoir les bruits xiphoïdiens liés au reflux gastro-oesophagien. Contrairement aux travaux précédents qui s'attachaient surtout à mettre en évidence *l'existence* d'une distribution en loi puissance et à justifier l'existence de celle-ci, on utilise ici la loi de Zipf comme un *outil d'analyse* afin de *détecter* et de *caractériser* un phénomène, l'objectif étant ici de fournir une aide au diagnostic des pathologies liées au reflux gastro-oesophagien. Ce type d'approche permet de dépasser le stade purement descriptif et ouvre la voie à de nouvelles applications des lois puissance.

On travaille ici sur des enregistrements sonores des bruits xiphoïdiens produits lors du passage d'un bolus baryté à travers le sphincter inférieur de l'oesophage. On applique une transformée en ondelettes sur le signal sonore enregistré. Cette transformée est calculée à différentes positions régulièrement espacées dans le temps. Pour chaque position τ les coefficients de la transformée en ondelettes forment un profil. Ce profil est ensuite codé pour constituer un mot. Chaque profil différent est associé à un mot différent. Le texte ainsi obtenu est ensuite découpé en plusieurs sous-textes. Chaque sous-texte correspond à un intervalle de temps d'environ une seconde, l'enregistrement complet totalisant une vingtaine de secondes. On applique ensuite la loi de Zipf sur chacun des sous-textes, en comptant le nombre d'occurrences de chacun des mots et en calculant la courbe rang-fréquence ou courbe de Zipf. Chaque courbe de Zipf est caractérisée par le paramètre β qui est l'opposé de la pente de la courbe et donc l'exposant de la loi de Zipf.

La valeur de β caractérise la structuration du signal. Sa variation au cours du temps va être utilisée pour déterminer les zones d'intérêt qui correspondent aux bruits xiphoïdiens. La figure suivante montre l'évolution de ce paramètre sur une séquence sonore :

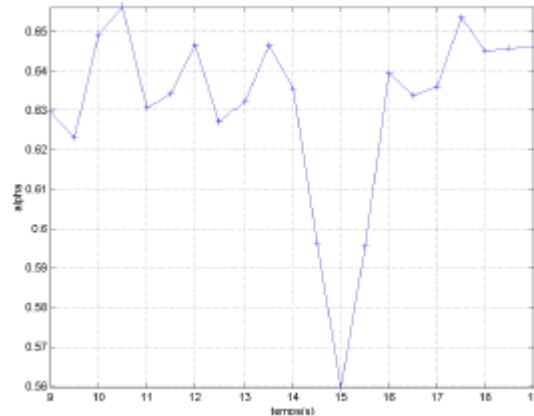


Fig 7: Evolution de l'exposant de la loi de Zipf au cours du temps

Sur cette figure, les zones d'intérêt sont représentées par les minimums locaux de β . La distance entre les différents minimums locaux permet de mesurer la durée du transit du bolus. Les valeurs ont été mesurées sur des patients avant et après une intervention chirurgicale. On observe en général une diminution des valeurs de β après l'intervention. L'utilisation de cette méthode permet de fournir une aide au diagnostic et d'évaluer l'efficacité de l'intervention chirurgicale.

Conclusion

Les distributions statistiques en loi puissance telle que la loi de Pareto et la loi de Zipf sont très communes. Elles ont été mises en évidence, à l'origine, de manière empirique, dans les domaines les plus divers tant dans des phénomènes naturels que dans le domaine des sciences humaines. Plusieurs interprétations ont été proposées pour expliquer ce phénomène. On peut les classer en deux grandes catégories.

Certaines de ces interprétations considèrent que les lois puissance sont le résultat d'une optimisation. C'est le cas en particulier du principe du moindre effort énoncé par Zipf, et de l'interprétation donnée par Mandelbrot. Cependant, de telles interprétations ne permettent pas de justifier l'existence de lois puissance dans des phénomènes aléatoires.

D'autres considèrent que la distribution en loi puissance est le résultat observé d'un phénomène de croissance aléatoire, et présentent des modèles comme celui de Simon, qui permettent de rendre compte de l'existence des lois puissance par de simples modèles probabilistes, sans considérer de finalité.

Alors quelle interprétation est la bonne ? En fait, les deux ne s'excluent pas mutuellement, et la plupart des phénomènes naturels où s'observent des lois puissance ne sont ni totalement aléatoires ni totalement contrôlés, ce qui fait que le hasard et l'optimisation peuvent intervenir tous les deux. Les distributions en loi puissance s'observent principalement dans des phénomènes comme la répartition des mots du langage ou de la population des villes, qui sont des phénomènes humains peu ou pas contrôlés et qui sont laissés à leur évolution naturelle.

Les loi puissances sembleraient alors représenter une sorte d'équilibre obtenu par un effet d'auto-optimisation.

Les loi puissances ont été employés dans de très nombreux domaines très divers comme l'économie, la démographie, la linguistique, la biologie ou l'étude des signaux sonores, mais il reste un domaine où jusqu'à présent elle n'avaient pas fait l'objet d'applications: c'est celui de l'image.

Chapitre 2 : Les lois puissance en analyse d'image

Introduction

Les lois puissance, et en particulier la loi de Zipf [Zip-49], ont été étudiées dans de nombreux domaines, aussi bien pour modéliser des phénomènes naturels que d'autres phénomènes appartenant aux sciences humaines. Cependant, la loi de Zipf a surtout été utilisée jusqu'à présent pour modéliser des phénomènes qui se décrivent de manière unidimensionnelle, comme les mots d'un texte ou des signaux sonores. C'est pourquoi il serait intéressant de rechercher l'existence de répartitions en loi puissance dans un cas bidimensionnel, tel que l'image.

La plupart des modèles utilisés pour décrire une image sont des modèles linéaires. Ces modèles ont l'avantage d'être simples à calculer, mais ils montrent vite leurs limites : en effet il est difficile de modéliser avec précision des objets complexes uniquement à l'aide de modèles linéaires. Par rapport à un modèle linéaire, un modèle en loi puissance présente l'avantage de posséder un degré de liberté supplémentaire, à savoir l'exposant de la loi puissance. De plus, contrairement au cas d'un polynôme ou d'une fraction rationnelle, cet exposant n'est pas nécessairement entier, ce qui permet une plus grande souplesse dans l'application du modèle. Ainsi, une loi puissance permet de modéliser certains phénomènes qui ne pourraient être approximés avec une précision suffisante au moyen de modèles linéaires. Cela justifie de vouloir utiliser ce modèle dans le cas des images, en effet la structure d'une image est en général trop complexe pour être approximée par des lois linéaires.

Il est une autre raison qui justifie le recours à des modèles de lois puissance pour caractériser une image. Selon Reed [Ree-01], des répartitions rang-fréquence en loi puissance du type de la loi de Zipf s'observent dans des phénomènes dont l'évolution peut être modélisée par un mouvement brownien géométrique dans le domaine temporel, comme dans le cas de la croissance des villes. Or, d'après Mandelbrot [Man-83], une image naturelle est caractérisée par une structure fractale, qui est également décrite par un mouvement brownien géométrique, cette fois-ci dans le domaine spatial. De ce fait, la dimension fractale d'une image exprime également au moyen d'une loi puissance un lien entre un paramètre et un facteur d'échelle. Par conséquent, il existe potentiellement un lien entre la dimension fractale d'une image et une distribution en loi de Zipf de motifs d'une image.

Afin d'appliquer les lois puissance à l'analyse d'image, il convient d'abord de définir précisément les modèles que l'on va utiliser. Nous nous limiterons à la loi de Zipf et la loi de Zipf inverse. Nous allons ensuite définir différents codages permettant de modéliser une image à l'aide des lois puissances et nous les appliquerons afin de vérifier quelles sont les caractéristiques de l'image qui peuvent être mises en évidence à l'aide de ces modèles. Enfin, nous étudierons les propriétés des modèles de lois puissance face à une transformation de l'image

1 Présentation des modèles utilisés

1.1 La loi de Zipf

Rappelons sa définition. La loi de Zipf décrit la répartition des n-uplets de symboles d'un ensemble suivant leur fréquence d'apparition. On considère un ensemble de symboles organisés topologiquement, comme un texte ou une image. Cet ensemble est organisé de telle manière qu'il soit possible de définir des n-uplets de symboles et de les dénombrer. Ces n-uplets peuvent être, selon les cas, de longueur fixe ou variable. La loi de Zipf décrit la distribution des fréquences d'apparition de chacun de ces n-uplets, c'est à dire le nombre d'occurrences distinctes entre elles d'un même n-uplet. On peut constater que les fréquences N_1, N_2, \dots, N_m d'apparition des n-uplets ne sont pas quelconques.

Si on classe les n-uplets U_1, U_2, \dots, U_m suivant l'ordre décroissant de leur fréquence d'apparition ($U_{\sigma(1)}, U_{\sigma(2)}, \dots, U_{\sigma(m)}$) et que l'on associe à chaque n-uplet son rang dans la séquence, les fréquences d'apparition des n-uplets sont données en fonction du rang par une loi puissance :

$$N_{\sigma(i)} = k.i^a \quad (1-1-1)$$

Dans cette formule, k et a sont des constantes. Cette loi puissance, dite loi rang-fréquence ou loi de Zipf, peut se représenter dans un graphe bi-logarithmique. Sur un tel graphe où l'abscisse indique $\log_{10}(i)$ et l'ordonnée $\log_{10}(N_{\sigma(i)})$, l'exposant de la loi puissance peut être estimé par le coefficient directeur de la droite de régression approchant au sens des moindres carrés le nuage de points $(\log_{10}(i), \log_{10}(N_{\sigma(i)}))$ avec $i = 1$ à n . On appellera **courbe de Zipf** la représentation graphique en utilisant deux échelles logarithmiques de la fréquence d'apparition des n-uplets en fonction de leur rang. Dans le cas où le phénomène étudié suit une loi puissance, cette courbe de Zipf est une droite.

A titre d'illustration, nous allons prendre l'exemple d'une chaîne de caractères dont on étudie les fréquences d'apparition de tous les couples de deux caractères consécutifs. Si l'on considère la chaîne "bbbaabaabbbb", construite sur un alphabet comportant deux symboles on peut obtenir 4 sortes différents : aa, ab, ba et bb , notés U_1, U_2, U_3, U_4 . On note dans un tableau le nombre d'apparitions de ces couples :

Couple	aa	ab	Ba	bb
Nombre d'apparitions	3	2	2	5

Il suffit ensuite de classer ces couples par ordre décroissant du nombre d'apparitions pour obtenir la courbe de Zipf.

Nous obtenons donc :

$$\begin{array}{lll} \sigma(1)= 4 & U_{\sigma(1)} = "bb" & \text{et } N_{\sigma(1)} = 5 \\ \sigma(2)= 1 & U_{\sigma(2)} = "aa" & \text{et } N_{\sigma(2)} = 3 \\ \sigma(3)= 2 & U_{\sigma(3)} = "ab" & \text{et } N_{\sigma(3)} = 2 \\ \sigma(4)= 3 & U_{\sigma(4)} = "ba" & \text{et } N_{\sigma(4)} = 2 \end{array}$$

La courbe de Zipf est représentée sur le graphique suivant.

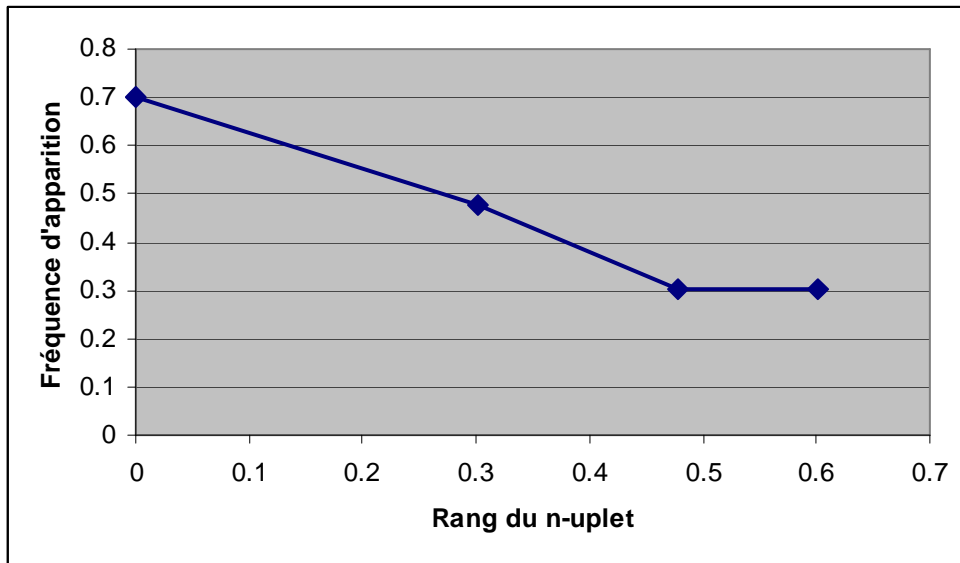


Fig. 1 : courbe de Zipf associée à une séquence de caractères.

On constate sur notre exemple que la loi de Zipf n'est pas vérifiée expérimentalement. Cela est dû au trop petit nombre de caractères dans la chaîne. On va maintenant appliquer ce modèle au cas des images. On s'intéressera principalement aux images en niveaux de gris. Une image est constituée d'un ensemble de pixels organisés dans le plan sous la forme d'une matrice rectangulaire $l \times h$ dont les coefficients représentent les niveaux de gris des pixels de l'image. En général, ces niveaux de gris sont codés sous la forme d'un entier codé sur 8 bits entre 0 et 255. Il existe donc de manière implicite une relation de voisinage entre pixels. Afin d'appliquer la loi de Zipf aux images, il convient de définir les n-uplets de symboles sur lesquels on va travailler. Les n-uplets de symboles seront ici des **motifs** constitués de pixels adjacents de l'image. L'image étant de dimension 2, la définition du motif devra conserver la relation de voisinage entre les pixels d'un motif dans les deux dimensions. C'est pourquoi on définira chaque motif comme une matrice carrée de pixels adjacents de l'image centrée sur un pixel. La matrice du motif sera par conséquent de taille impaire. Le motif $M_{2k+1}(i,j)$ de taille $2k+1$ centré sur le pixel $v(i,j)$ est défini de la manière suivante pour $k > 1$:

$$M_{2k+1} = \left\{ \begin{array}{ccc} v(i-k, j-k) & \dots & \dots & v(i-k, j+k) \\ & & \dots & \\ & & v(i, j) & \\ & & \dots & \\ v(i+k, j-k) & & \dots & v(i+k, j+k) \end{array} \right\} \quad (1-1-2)$$

Afin de récupérer les motifs de l'image, on balaie l'image par un masque de taille $2k+1$. Comme on s'intéresse essentiellement aux fréquences d'apparition de chacun des motifs et non aux relations spatiales entre les différents motifs, le sens du balayage de l'image n'a pas réellement d'importance. On peut par exemple utiliser un balayage séquentiel de l'image, ligne par ligne, de gauche à droite et de haut en bas, comme indiqué sur la figure 2. Ce balayage de l'image peut être chevauchant ou non, c'est à dire que l'on peut considérer des motifs qui se recouvrent ou simplement juxtaposés. L'utilisation d'un balayage entrelacé permet de caractériser la distribution des fréquences des motifs de l'image avec une précision plus grande, mais le temps de calcul est plus long. On applique ensuite un codage aux pixels du masque, différents codages pouvant être utilisés, selon les propriétés de l'image que l'on cherche à mettre en évidence. La définition et l'utilisation de ces différents codages seront détaillées dans un prochain paragraphe.

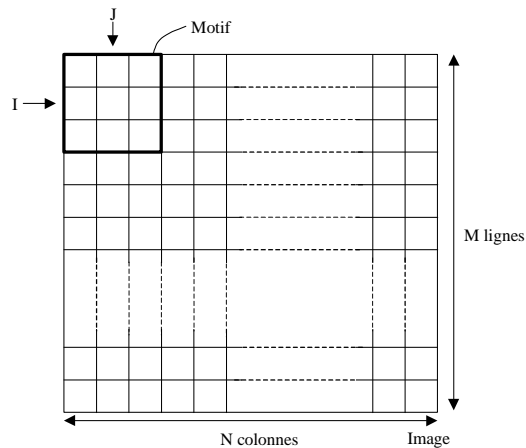


Fig. 2 : Balayage de l'image par un masque 3 x 3

L'algorithme de construction de la courbe de Zipf associée à une image consiste donc à balayer séquentiellement l'image avec un masque de capture de taille $2k+1$, à coder les motifs et à compter le nombre d'occurrences de chacun des motifs. Pour cela, on range dans un tableau tous les motifs rencontrés dans l'image, en associant à chaque motif sa fréquence d'apparition. Le tableau des motifs est initialement vide. Pour chaque nouveau masque étudié, on compare successivement le motif courant à chacun des motifs déjà présents dans le tableau. Si le motif courant est déjà présent dans le tableau, on incrémente sa fréquence d'apparition d'une unité. Si le motif n'est pas dans le tableau, c'est à dire qu'il n'a jamais été encore rencontré dans l'image, on l'ajoute en fin de tableau et on met sa fréquence d'apparition à 1. Lorsque l'image a été entièrement balayée, on trie le tableau par ordre décroissant de la fréquence d'apparition des motifs, en utilisant par exemple un algorithme de tri rapide. Ensuite, on trace la courbe donnant la fréquence d'apparition en fonction du rang dans un repère bi-logarithmique.

On va maintenant évaluer la complexité de cet algorithme. On travaille sur une image de largeur l et de hauteur h , qui contient au total $l \times h = n$ pixels. On compare successivement les motifs aux k motifs déjà présents dans le tableau. Dans le pire des cas, c'est à dire si le motif n'est pas déjà dans le tableau, on aura donc à effectuer k comparaisons. La valeur de k varie pendant l'exécution du programme, on va s'intéresser à sa valeur maximale. Le cas le plus défavorable est celui où on doit rechercher le dernier motif d'une image où les motifs sont tous différents. Dans ce cas, le tableau contient n motifs. On doit donc effectuer au maximum n itérations pour rechercher un motif. La complexité de la phase de codage du motif est variable selon le motif utilisé, cependant le temps d'exécution de l'algorithme de codage peut

être considéré comme négligeable, pour des grandes valeurs de k , devant le temps nécessaire pour la recherche dans le tableau. La recherche dans le tableau est effectuée n fois, autant qu'il y a de motifs dans l'image. Au total, la complexité maximale de l'algorithme de comptage des motifs sera donc en $O(n^2)$. On applique ensuite un algorithme de tri sur le tableau, or on sait que la complexité moyenne d'un tel algorithme en $O(n \ln n)$ et sa complexité dans le pire des cas est en $O(n^2)$. Par conséquent, la complexité de l'algorithme de la construction de la courbe de Zipf d'une image est en $O(n^2)$. Cela signifie que le temps de calcul maximal est proportionnel au carré du nombre de pixels de l'image. De ce fait, le temps de calcul de cet algorithme est relativement long, surtout sur de grandes images.

Afin de réduire le temps de calcul, il est possible de partitionner le tableau des motifs en plusieurs sous-tableaux en utilisant une fonction de hachage. On effectue alors une pré-classification des motifs en fonction du résultat de la fonction de hachage. On diminue ainsi le nombre de motifs à comparer au motif courant, en effet on ne compare plus le motif courant qu'aux motifs présents dans un seul des sous-tableaux. Pour être efficace, la fonction de hachage doit être la plus simple possible à calculer. On peut par exemple choisir comme fonction de hachage la valeur du code du dernier pixel du motif, on aura ainsi autant de sous-tableaux qu'il y a de valeurs possibles pour ce code. A la fin du calcul, les tableaux seront concaténés et on effectuera le tri sur l'ensemble des motifs. En effet, le tri final n'étant effectué qu'une seule fois, il a une influence relativement faible sur la complexité de l'algorithme. L'utilisation de la fonction de hachage n'a aucune incidence sur l'exactitude du résultat final du calcul. Cette méthode est efficace, les essais ayant montré qu'elle permet de diviser le temps de calcul en moyenne par 4. Elle est d'autant plus intéressante que l'image est grande, sur de très petites images on constate au contraire une augmentation du temps de calcul due à l'évaluation de la fonction de hachage.

1.2 La loi de Zipf inverse

Comme la loi de Zipf, cette loi s'intéresse aux fréquences d'apparition des n -uplets d'un ensemble de symboles topologiquement ordonnés mais on s'intéresse ici au nombre de n -uplets différents ayant la même fréquence d'apparition. Ces n -uplets sont définis de la même manière qu'avec la loi de Zipf. D'après la loi de Zipf inverse, le nombre I de n -uplets différents ayant une fréquence d'apparition f est donné par la formule suivante :

$$I(f) = af^{-b} \quad (1-2-1)$$

Dans cette formule, a et b sont des constantes positives. Cette formule est valable pour les n -uplets les moins fréquents de l'image. La loi de Zipf inverse se représente graphiquement par une courbe en échelles logarithmiques, **la courbe de Zipf inverse**. Pour vérifier la loi de Zipf inverse sur les images, les motifs sont définis de la même manière que dans le cas de la loi de Zipf, à savoir par des matrices carrées de taille $2k+1$ centrées sur un pixel. Les mêmes codages seront appliqués pour les motifs de l'image et l'on appliquera également un balayage séquentiel de l'image.

L'algorithme de construction de la courbe de Zipf inverse consiste donc à balayer l'image avec le masque de capture et à compter les motifs, puis pour chaque valeur trouvée de la fréquence d'apparition, à dénombrer les motifs différents ayant cette fréquence d'apparition. La première phase de l'algorithme est identique à celle utilisée dans le cas de la courbe de

Zipf, à savoir que l'on balaie l'image et que l'on compare le motif courant aux motifs déjà présents dans le tableau, et que l'on rajoute le motif dans le tableau s'il n'y est pas.

La seconde phase de l'algorithme consiste à compter les motifs ayant une même fréquence d'apparition. Pour cela, on initialise la fréquence cherchée à 1, puis on parcourt séquentiellement le tableau des motifs afin de compter les motifs ayant la même fréquence que la fréquence courante. On réitère ensuite l'algorithme en incrémentant à chaque fois la fréquence d'une unité jusqu'à ce que l'on ait atteint la fréquence maximale. On obtient donc la distribution du nombre de motifs en fonction de leur fréquence d'apparition. On trace ensuite la représentation graphique dans un repère bi-logarithmique.

La complexité de l'algorithme de construction de la courbe de Zipf inverse est également en $O(n^2)$. La phase de comptage des motifs est la même que celle employée pour la courbe de Zipf, sa complexité est donc en $O(n^2)$. La recherche du maximum dans le tableau a une complexité en $O(n)$. Le comptage des motifs différents selon les fréquences a une complexité au plus en $O(n^2)$, en effet, il y a moins de n valeurs de la fréquence des motifs dans l'image, et pour chaque valeur de la fréquence on parcourt une fois le tableau des motifs, dont la taille maximale est n . Au total, la complexité de l'algorithme est donc en $O(n^2)$.

Les outils d'analyse statistique par des modèles en loi puissance, qui ont été définis au départ pour des phénomènes unidimensionnels, peuvent donc être facilement adaptés au cas des phénomènes bidimensionnels tels que les images.

2 Codage de l'image

Afin d'appliquer les modèles de loi puissance en analyse d'image, il est nécessaire de définir précisément le codage des images que l'on va utiliser. Les lois de Zipf ayant été définies à l'origine pour décrire la distribution des fréquences des mots d'un texte, pour pouvoir appliquer ce modèle aux images nous allons devoir définir un équivalent de la notion de mot dans le domaine des images. Un mot est associé à un ensemble de symboles adjacents. Nous envisageons ici 4 définitions. Elles ont aussi un rôle comparable à celui des flexions dans la langue naturelle. C'est le motif ici qui tient lieu de mot.

2.1 Utilisation directe des niveaux de gris de l'image

La manière la plus simple de coder l'image est d'utiliser directement les pixels de l'image, c'est à dire que l'on calcule simplement la fréquence d'apparition des différents niveaux de gris. On décompte individuellement les pixels, le n -uplet est ici réduit à un symbole unique. Dans ce cas, la modélisation par une loi de Zipf revient à effectuer le calcul de l'histogramme de l'image, ordonné suivant l'ordre décroissant des fréquences des niveaux de gris. L'analogie avec l'analyse textuelle serait ici de décompter les fréquences des lettres du texte. La figure ci-dessous montre un exemple d'application de cette méthode sur une image :

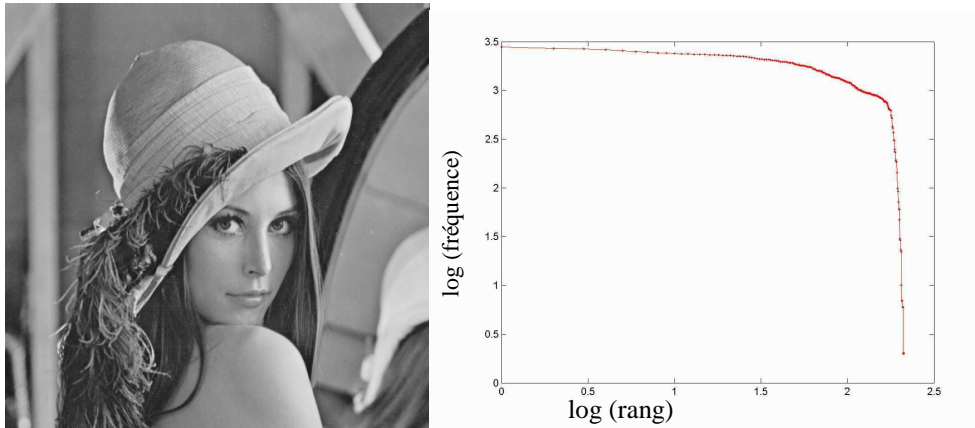


Fig . 3 : Courbe de Zipf des niveaux de gris des pixels d'une image

Sur cet exemple, on voit que la répartition des fréquences d'apparition des niveaux de gris des pixels ne suit pas globalement une loi puissance. On a une distribution linéaire pour les fréquences élevées, mais la distribution est très fortement incurvée vers le bas pour les niveaux de gris les moins fréquents. En fait, cette distribution est très similaire à celle que l'on obtient pour les fréquences des lettres de l'alphabet dans des textes écrits en français ou en anglais, à savoir que la distribution est « tirée vers le bas » par les lettres les plus rares. Cela montre que la taille de l'alphabet des niveaux de gris, bien qu'étant environ 10 fois plus importante que celle de l'alphabet latin, reste insuffisante pour que l'on puisse observer une distribution rang-fréquence qui suive une loi puissance. Il est donc nécessaire de définir des n-uplets de longueur supérieure à l'unité pour appliquer les modèles de lois puissance.

Soit $M_{2k+1}(i,j)$ le motif de taille $2k+1$ centré sur le pixel $v(i,j)$. Dans le cas où $k=1$, le motif est une matrice de taille 3×3 qui est constituée du 8-voisinage centré en (i,j) . Si l'on travaille avec des niveaux de gris codé 8 bits comme c'est le plus souvent le cas en imagerie numérique, le cardinal de l'ensemble des motifs possibles sera alors égal à 256^{2k+1} . Par exemple, pour $k=1$ on aura 256^9 , soit $4,7 \times 10^{21}$ motifs différents possibles. La figure ci-dessous montre un exemple de courbe de Zipf obtenue sur une image en 256 niveaux de gris avec des motifs de taille 3×3 .

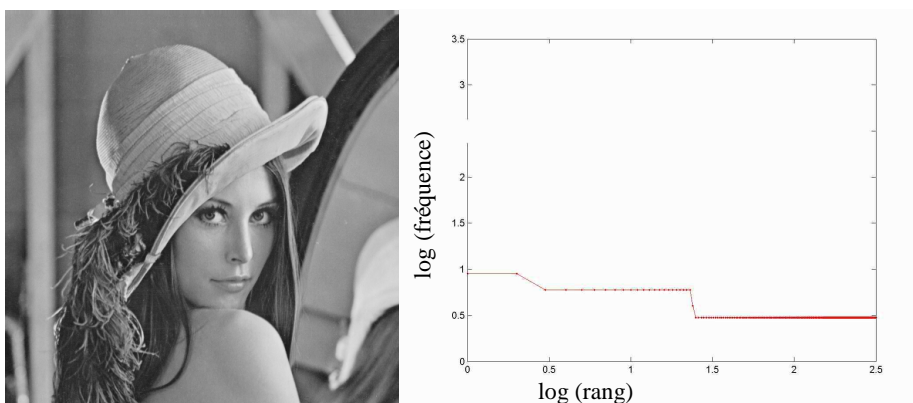


Fig. 4 : Courbe de Zipf obtenue en utilisant les niveaux de gris des pixels sur des motifs 3×3 .

On remarque que l'on observe pas non plus dans ce cas une répartition en loi puissance. En fait, la plupart des motifs n'apparaissent qu'une seule fois dans l'image, les motifs ayant une fréquence d'apparition supérieure à 1 sont très peu nombreux, et de plus leur fréquence reste faible. On retrouve ce même résultat sur toutes les images photographiques numériques. Ce résultat s'explique par le trop grand nombre de motifs différents présents dans l'image. En effet, une image photographique comporte en général de nombreux détails fins, ce qui se traduit par autant de motifs différents dans le calcul de la loi de Zipf. De plus, les zones homogènes de l'image ne sont pas totalement uniformes, elles comportent toujours de faibles variations de niveaux de gris qui sont très peu perceptibles à l'œil nu, mais qui constituent pourtant des motifs distincts. On peut se rendre compte de la non-uniformité des surfaces d'une image naturelle en examinant des vues agrandies de l'image comme celles montrées sur la figure 5.

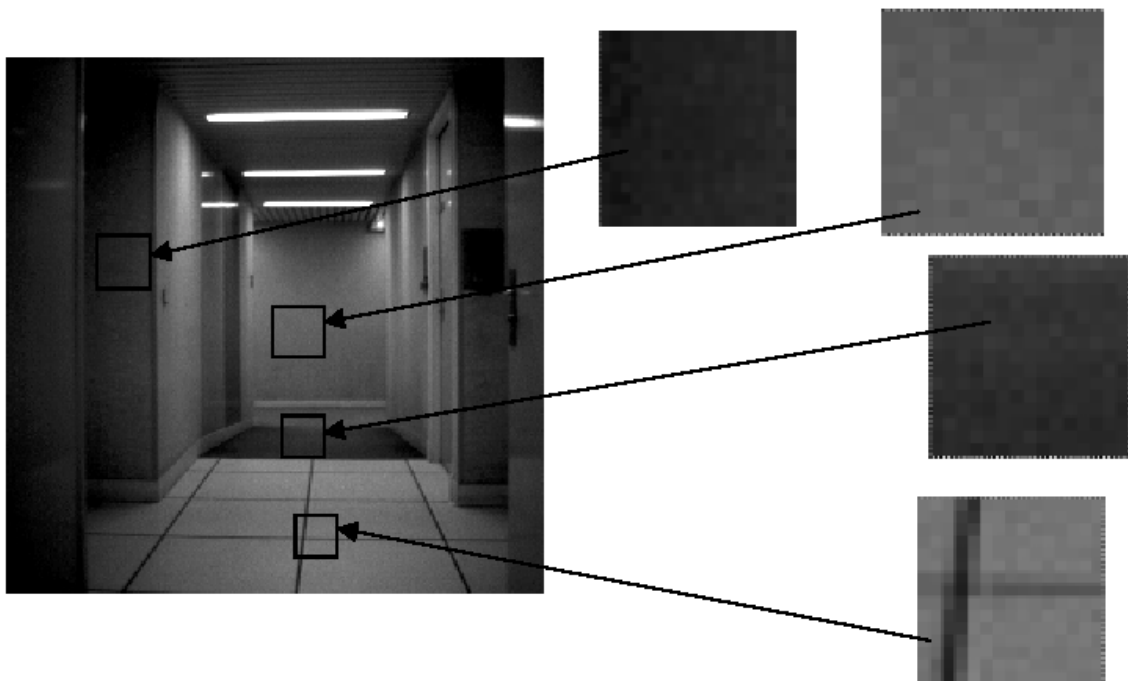


Fig. 5 : Détails des textures d'une image photographique non compressée : les zones homogènes comportent des variations de niveaux de gris.

La répartition statistique obtenue en utilisant directement les niveaux de gris de l'image n'est donc absolument pas significative quant au contenu de l'image et n'est donc absolument pas exploitable. On peut également remarquer qu'il n'est pas utile de chercher à utiliser des motifs d'une taille supérieure à 3x3, on ne ferait qu'augmenter le nombre de motifs distincts présents dans l'image, qui est déjà excessif avec des motifs de taille 3x3, et de plus, on augmenterait considérablement le temps de calcul.

C'est pourquoi il est nécessaire d'utiliser un codage des motifs qui permette de diminuer le nombre de motifs présents dans l'image.

2.2 Rééchantillonnage de l'échelle des niveaux de gris : le codage des n classes

Une méthode simple pour réduire le nombre de motifs différents susceptibles d'être présents dans une image est de rééchantillonner l'échelle des niveaux de gris de l'image afin de limiter le nombre de valeurs possibles de ceux-ci. Pour cela, on va affecter une même valeur aux pixels ayant des niveaux de gris contenus dans un même intervalle de valeurs. On partitionne l'échelle [0, 255] des niveaux de gris en un nombre n d'intervalles, ou *classes* de largeurs égales et numérotés suivant l'ordre croissant de 0 à n , d'où le nom de méthode des n classes donné à ce codage. On attribue à chaque pixel du motif la valeur $c(x,y)$ de sa classe en fonction de la valeur $g(x,y)$ de son niveau de gris, en utilisant la formule suivante :

$$c(x,y) = \text{int} \left[\frac{n/g(x,y)}{255} \right] \quad (2-2-1)$$

Concrètement, cela revient à coder l'information des niveaux de gris sur n valeurs au lieu de 256. Le nombre de motifs différents possibles est alors de n^9 pour un motif de taille 3x3. L'utilisation de ce codage, s'il permet de réduire le nombre de motifs possibles et d'obtenir une distribution des fréquences des motifs qui ait une signification, introduit évidemment une perte d'information, puisque l'on a réduit le nombre de valeurs possibles de l'échelle des niveaux de gris. Le problème que pose l'implémentation de cette méthode est le choix du nombre de classes, il faut en effet choisir une valeur de n qui soit la plus petite possible afin de réduire le nombre de motifs différents possibles, tout en étant suffisamment grande pour ne pas trop altérer la structure de l'image. Pour cela, des essais ont été effectués avec différentes valeurs de n afin de déterminer une valeur optimale qui satisfasse au mieux à ces deux exigences contradictoires. La figure suivante montre une même image codée avec différentes valeurs de n :






$n=4$	$n=6$	$n=9$	$n=12$	$n=16$
				
2.621×10^5 motifs possibles	1.008×10^7 motifs possibles	3.874×10^8 motifs possibles	5.159×10^9 motifs possibles	6.872×10^{10} motifs possibles

Fig. 6 : Exemple d'image codée avec différentes valeurs de n

Les résultats expérimentaux ont montré que le choix d'une valeur de $n=9$ permet de préserver le contenu structurel de l'image sans que la dégradation ne soit trop visible, tout en maintenant à une valeur raisonnable le nombre de motifs différents présents dans l'image. C'est donc cette valeur qui sera retenue par la suite, sauf mention contraire. Par conséquent, le codage par partitionnement de l'échelle des niveaux de gris sera désormais désigné sous le nom de méthode des 9 classes.

Au lieu d'utiliser une partition de l'échelle des niveaux de gris en classes de largeur fixe, on peut également utiliser des classes de largeur variable, afin d'adapter le codage des motifs à la distribution des niveaux de gris de l'image testée. Pour cela, on utilise la méthode des k-means afin de déterminer les centres des classes, afin de rééquilibrer la largeur des classes en fonction de la répartition des niveaux de gris de l'image. Cependant, cette méthode a pour inconvénient d'augmenter considérablement le temps de calcul par rapport à un partitionnement en classes de largeur fixe. De plus, si l'image est suffisamment contrastée, les résultats sont peu différents de ceux obtenus avec la méthode des n classes. L'utilisation de la méthode des k-means est donc peu intéressante dans la pratique, sauf dans certains cas particuliers d'images très claires ou très sombres. Et encore, même dans ce type d'image il peut être plus judicieux car plus économique en temps de calcul d'utiliser un prétraitement pour augmenter le contraste de l'image avant d'utiliser la méthode des 9 classes.

2.3 Codage des rangs généraux

Si le codage des 9 classes permet de mettre en évidence les principales structures de l'image, en revanche il ne permet pas d'étudier la texturation fine de l'image, car il fait perdre l'information sur les faibles variations de luminance, ce qui uniformise les textures de l'image. C'est pourquoi on va également définir d'autres codages qui permettent de prendre en compte les effets de ces variations tout en utilisant un alphabet de taille réduite afin de limiter le nombre de motifs possibles. Pour cela on va utiliser un codage par rangs. Le principe de ce codage est d'utiliser non pas directement l'information de luminance, mais les variations locales de niveaux de gris, afin de mettre en évidence les différences relatives plutôt que les valeurs absolues. Le codage par rangs consiste à remplacer les niveaux de gris des pixels par leur rang dans un voisinage, il s'agit d'une transformation non linéaire qui, si elle fait perdre la notion de différence des niveaux de gris, conserve les relations ordinales entre les niveaux de gris des pixels voisins. Les relations de rangs ont d'abord été étudiées dans le cadre des statistiques non paramétriques, notamment par Siegel [Sie-56] pour décrire certains types de phénomènes naturels. En analyse d'image, les codages par rang ont été utilisés pour caractériser les textures, notamment par Harwood et al .[Har-85], L. Wang et He [Wan-90] et Patel et Stoneham [Pat-92]. La méthode que l'on va utiliser pour coder les motifs est dérivée de celle proposée par Bi [Bi-97] pour la segmentation des images en régions texturées, mais ici, au lieu de définir le motif par un 4-voisinage comme dans l'approche utilisée par Bi, le motif sera constitué d'un 8-voisinage correspondant à la matrice 3x3 qui constitue notre masque de capture.

On s'intéresse à un motif constitué d'un pixel central $p(i,j)$, dont le niveau de gris est $g(i,j)$, et à ses voisins immédiats. Le motif est constitué donc constitué des 9 pixels suivants :

$$\{p(i-1,j-1),p(i-1,j),p(i-1,j+1),p(i,j-1),p(i,j),p(i,j+1), p(i+1,j-1),p(i+1,j),p(i+1,j+1)\} \quad (2-3-1)$$

Le vecteur des niveaux de gris du motif est défini de la manière suivante :

$$V(i,j)=(g(i-1,j-1),g(i-1,j),g(i-1,j+1),g(i,j-1),g(i,j),g(i,j+1), g(i+1,j-1),g(i+1,j),g(i+1,j+1))^T \quad (2-3-2)$$

A chaque pixel est associé sa position spatiale dans le motif, les pixels du motif étant numérotés de gauche à droite et de haut en bas, comme indiqué dans le tableau ci-dessous :

0	1	2
3	4	5
6	7	8

En utilisant les indices de position spatiale des pixels, le vecteur des niveaux de gris peut s'écrire :

$$V(i,j)=(g_0, g_1, g_2, g_3, g_4, g_5, g_6, g_7, g_8)^T \quad (2-3-3)$$

Pour chaque motif de vecteur de niveau de gris $v(i,j)$, on définit le vecteur de rangs $R(i,j)$ du motif :

$$R(i,j)=(r_0, r_1, r_2, r_3, r_4, r_5, r_6, r_7, r_8)^T \quad (2-3-4)$$

Où r_i est le rang du niveaux de gris f_i dans $V(i,j)$ et $0 \leq r_i \leq 8$.

Il reste maintenant à définir de manière précise le rang r_i associé à chacun des pixels du motif. Les rangs des niveaux de gris peuvent être définis par la méthode des rangs distincts ou la méthode des rangs généraux.

La méthode des rangs distincts consiste à ordonner les niveaux de gris f_i du motif, avec i appartenant à l'intervalle $[0..8]$, dans un ordre non décroissant $f_{i_0} \leq f_{i_1} \leq f_{i_2} \dots f_{i_8}$, où i_k , k appartient à l'intervalle $[0..8]$ et représente la position. Le rang du pixel est donné par : $r_{i_0}=0$, $r_{i_1}=1$, $r_{i_2}=2$, $r_{i_8}=8$. L'ordre est secondairement réalisé en fonction de la numérotation choisie pour les pixels du masque. En d'autres termes, cela revient à attribuer la valeur 0 au pixel de niveau de gris le plus bas du motif et une valeur croissante aux autres pixels, en donnant à chaque pixel du motif une valeur différente. Le nombre de motifs différents possibles, pour un motif de taille 3×3 n'est plus que de $9! = 362880$. La figure ci-dessous montre un exemple de motif codé avec la méthode des rangs distincts.

255	210	210
25	2	34
40	2	40

8	6	7
2	0	3
4	1	5

Fig. 7 : Exemple de motif codé avec les rangs distincts

Le codage des rangs distincts, si il permet de réduire de façon significative le nombre de motifs, présente un grave défaut : on affecte arbitrairement des valeurs de rang différentes à des pixels de niveaux de gris identiques. Le codage du motif dépend du sens de parcours de celui-ci : le codage n'est donc pas invariant par rotation. De plus, on introduit artificiellement une information ne figurant pas dans l'image originale et n'ayant aucune réalité physique, cette information constitue un bruit. C'est pourquoi on préférera utiliser dans la pratique non pas le codage des rangs distincts mais le codage des rangs généraux.

Le codage des rangs généraux, contrairement au précédent, tient compte des égalités de niveaux de gris à l'intérieur du motif. On affecte ici la même valeur quand les pixels du motif ont le même niveau de gris. Le codage d'un motif se fait de la manière suivante : soient f_i , les niveaux de gris des pixels du motif, avec i compris entre 0 et 8. On ordonne les f_i dans un ordre non décroissant $f_{i_0} \leq f_{i_1} \leq f_{i_2} \dots f_{i_8}$, où i_k représente la position spatiale des pixels. Le rang du motif dont le niveau de gris est le plus faible est $r_{i_0}=0$. Ensuite, on affecte les rangs des autres pixels f_{i_k} du motif sous la condition suivante : si $f_{i_k} = f_{i_{k'}}$ alors $r_{i_k} = r_{i_{k'}}$. Les pixels ayant un même niveau de gris recevront donc une même valeur de rang. La figure ci-dessous montre un exemple de motif codé avec la méthode des rangs généraux :

255	210	210
25	2	34
40	2	40

5	4	4
1	0	2
3	0	3

Fig. 8 : Exemple de motif codé avec les rangs généraux

Le codage des rangs généraux présente l'avantage d'être plus fidèle au motif d'origine, de ne pas introduire d'information artificielle et de ne pas dépendre du sens de parcours du motif. Par contre, il est plus complexe que le précédent et nécessite une combinatoire plus lourde. Contrairement aux cas précédents, le calcul du nombre de motifs possibles n'est pas trivial, et pour dénombrer les motifs nous allons devoir étudier les différents cas possibles. On se place dans le cas d'un motif constitué de K pixels.

Dans le cas où tous les pixels sont identiques, il n'y a alors qu'un seul vecteur de rangs possible, le vecteur $(0,0,0\dots 0)^T$. Si l'on note $N_1(K)$ le nombre de possibilités, alors $N_1(K)=1$

Dans le cas où seuls deux rangs sont différents, chaque rang r_q , $1 \leq q \leq K$ ne peut prendre que deux valeurs, 0 ou 1. Il y a donc 2^K possibilités, mais il faut retrancher le cas du vecteur $(0,0,0\dots 0)^T$ déjà traité et celui de $(1,1,1\dots 1)^T$ qui est impossible parce que par définition la plus petite valeur de rang d'un motif est toujours égale à 0. Donc on a $N_2(K)=2^K-2$.

Si l'on généralise le résultat précédent au cas où le vecteur de rangs du motif comporte i rangs différents tels que $2 \leq i \leq K$, chaque rang r_q du vecteur de rangs peut donc prendre une valeur de rang parmi $0,1,\dots,i-1$. Le nombre de possibilités est donc au total de i^K . De ce nombre, il faut maintenant retrancher les cas déjà comptabilisés et les cas impossibles, à savoir :

- Les cas avec tous les rangs égaux : $(0,0,0\dots 0)$, $(1,1,1\dots 1)^T$, ..., $(i-1,i-1,i-1\dots i-1)^T$, soit $C_1^i * N_1(K)$ possibilités.
- Les cas avec deux rangs différents, soit au total $C_2^i * N_2(K)$ possibilités.
- D'une manière générale, il faut décompter les motifs qui comportent j rangs, $2 \leq j \leq K$ soit au total $C_j^i * N_j(K)$

Par conséquent, le nombre de motifs distincts comportant i rangs différents est donc :

$$N_i(K) = i^K - \sum_{j=1}^{i-1} (C_j^i * N_j(K)) \quad \forall i \in [2..K] \quad \text{avec } N_1(K) = 1.$$

Au total, le nombre total de vecteurs de rangs différents codés à l'aide de la méthode des rangs généraux pour un motif de taille K est donc de :

$$N = \sum_{i=1}^K N_i(K)$$

Dans le cas d'un motif de taille 9, on aura donc $4,1179 \times 10^8$ motifs possibles. Ce nombre est supérieur à celui que l'on aurait en utilisant les rangs distincts, mais reste très inférieur au nombre de motifs différents codés sur 256 niveaux de gris.

2.4 Codage des rangs généraux avec seuil

Le codage des rangs généraux, tel qu'il a été défini, permet de mettre en évidence les variations de niveaux de gris à l'intérieur du motif. Cependant, ce codage présente le défaut de ne pas tenir compte de l'amplitude des variations des niveaux de gris, ainsi une variation de luminance très faible et pratiquement invisible à l'œil nu sera traitée de la même façon qu'une variation beaucoup plus importante. De ce fait, le résultat du codage des rangs généraux n'est pas conforme à la perception visuelle de l'œil humain.

Il est cependant possible de modifier ce codage afin de limiter l'influence des faibles variations d'intensité des niveaux de gris. Pour cela, on définit un seuil S au-dessous duquel la variation de luminance sera considérée comme négligeable. On effectue le codage des rangs généraux suivant le même principe que précédemment, mais on affecte la même valeur de rang à deux pixels si la différence entre leurs niveaux de gris est strictement inférieure à S .

Le seuil doit être choisi de manière à être conforme à la perception visuelle humaine. Dans la pratique, une valeur de 10 peut convenir, en effet les variations de niveaux de gris inférieures à 10 unités sont peu perceptibles à l'œil nu.

2.5 Bilan

Dans le cas d'une application des modèles de lois puissances aux images, il n'est pas possible d'utiliser directement les niveaux de gris de l'image, car la taille de l'alphabet des n -uplets des symboles utilisés est inadaptée. C'est pourquoi il est nécessaire de définir des codages de l'image qui permettent de réduire le nombre maximal de motifs possibles à une quantité raisonnable. Cependant, un tel codage introduit une perte d'information, c'est pourquoi il faut pour choisir le codage se demander quelle est l'information pertinente à conserver en fonction de l'application. Le codage des n classes permet de conserver les structures principales d'une image, mais fait perdre l'information sur les faibles variations de niveaux de gris et donc détails fins de l'image. Le codage des rangs généraux a les effets inverses du précédent, à savoir qu'il met en évidence l'existence de variations même légères de luminance, mais on perd l'information sur l'amplitude relative des variations. Une solution intermédiaire est le codage des rangs généraux avec seuil, qui ne prend en compte que les variations de niveaux de gris dépassant une certaine amplitude. Les différents codages proposés sont complémentaires les uns des autres, chacun mettant en évidence une caractéristique particulière de l'image.

3 Application des modèles de lois puissance

Nous allons maintenant appliquer les modèles de lois puissance que l'on a précédemment définis, avec les différents codages, sur différentes images de test afin de vérifier quelles sont les caractéristiques de l'image qui peuvent être mises en évidence par l'étude de la

distribution des fréquences d'apparition des différents motifs de l'image. Nous étudierons également l'influence de la résolution de l'image sur la répartition statistique des motifs.

3.1 Utilisation du codage des 9 classes

Pour commencer, nous allons appliquer la loi de Zipf avec le codage des 9 classes sur des images photographiques en niveaux de gris non compressées.

La figure ci-dessous montre un exemple de courbe de Zipf obtenue à l'aide de cette méthode sur une image de test. Les résultats qui sont montrés ici sont représentatifs de l'ensemble des images testées, en effet on obtient des courbes de Zipf d'allure tout à fait semblable sur la quasi-totalité des images photographiques naturelles.

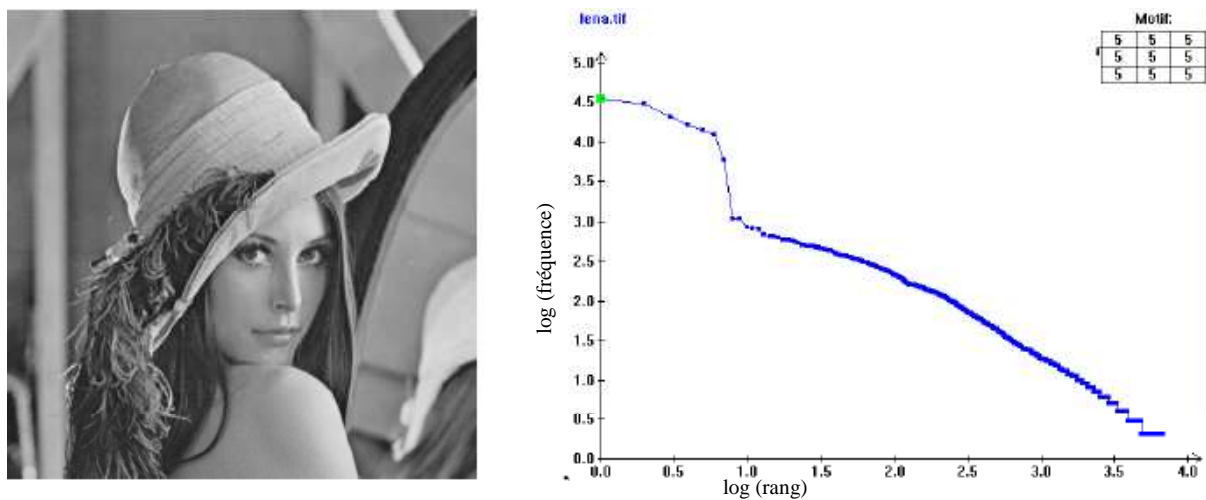


Fig. 9 : Courbe de Zipf obtenue sur l'image Lena avec le codage des 9 classes

En examinant la courbe, on remarque tout d'abord qu'elle n'est pas linéaire, mais qu'elle présente une discontinuité importante entre les fréquences d'apparition des premiers motifs et de celle des suivants. En toute rigueur, on pourrait dire que la répartition des motifs de l'image n'obéit pas à une loi puissance. Cependant, on peut constater que la courbe est en fait linéaire par morceaux et qu'elle peut donc être approximée par deux segments de droite.

Maintenant, si l'on s'intéresse à la nature des motifs les plus fréquents de l'image, on remarque qu'il s'agit toujours de motifs homogènes, c'est à dire dont tous les pixels appartiennent à une même classe de niveaux de gris. Ces motifs appartiennent aux principales zones homogènes de l'image. On peut facilement s'en rendre compte grâce à la figure ci-dessous qui montre la position spatiale des occurrences des 10 motifs les plus fréquents de l'image, classés de gauche à droite et de haut en bas par fréquence décroissante. Sur cette figure, les points colorés indiquent les positions des différentes instances du motif dont les classes des niveaux de gris sont indiquées sous l'image.



Fig. 10 : Positions des 10 motifs les plus fréquents de l'image Lena.

On observe un résultat semblable sur toutes les images photographiques naturelles. Les motifs les plus fréquents de la séquence sont toujours des motifs uniformes qui correspondent aux principales régions homogènes de l'image. Les autres motifs correspondent à des détails fins, des zones non uniformes et des contours.

On distingue donc deux classes de motifs sur l'image : les zones homogènes, représentées sur cet exemple par les 7 motifs les plus fréquents de l'image, et les détails fins représentés par tous les autres motifs. Si l'on prend séparément les deux classes de motifs, on remarque que dans le cas des motifs homogènes, la courbe rang-fréquence est linéaire pour les motifs les plus fréquents de cette classe, et s'incurve ensuite fortement vers le bas. La répartition des fréquences des motifs homogènes obéit donc à une loi puissance sauf pour les motifs les plus rares de cette classe. Ce résultat est conforme à ce que l'on obtient habituellement dans les cas où le cardinal de l'ensemble des n-uplets est petit, comme par exemple dans le cas des lettres de l'alphabet.

En ce qui concerne les motifs non homogènes, la courbe rang-fréquence est linéaire : la répartition des fréquences des motifs obéit donc à la loi de Zipf, comme c'est le cas pour les mots d'un texte.

Le codage des 9 classes sépare donc les motifs en deux classes, les motifs homogènes et les autres. La répartition statistique des fréquences des motifs en utilisant ce codage n'obéit pas à une loi puissance mais à deux, l'une pour les principales zones homogènes de l'image et l'autre pour les détails fins de l'image.

La répartition statistique des motifs, et donc l'allure de la courbe de Zipf obtenue avec le codage des 9 classes, varie suivant le contenu structurel de l'image. Sur une image simple comportant peu de détails, on observe une prédominance très nette des zones uniformes. L'ordonnée à l'origine est donc très élevée, la pente de la première partie de la courbe est presque horizontale et l'on observe une discontinuité très nette entre les deux parties de la courbe.

En revanche, sur une image plus complexe, les zones uniformes ne sont qu'en quantité plus réduite, et la courbe suit une pente nettement plus régulière, comme on peut le voir sur la figure ci-dessous, où les points colorés indiquent la position du motif le plus fréquent dans l'image.

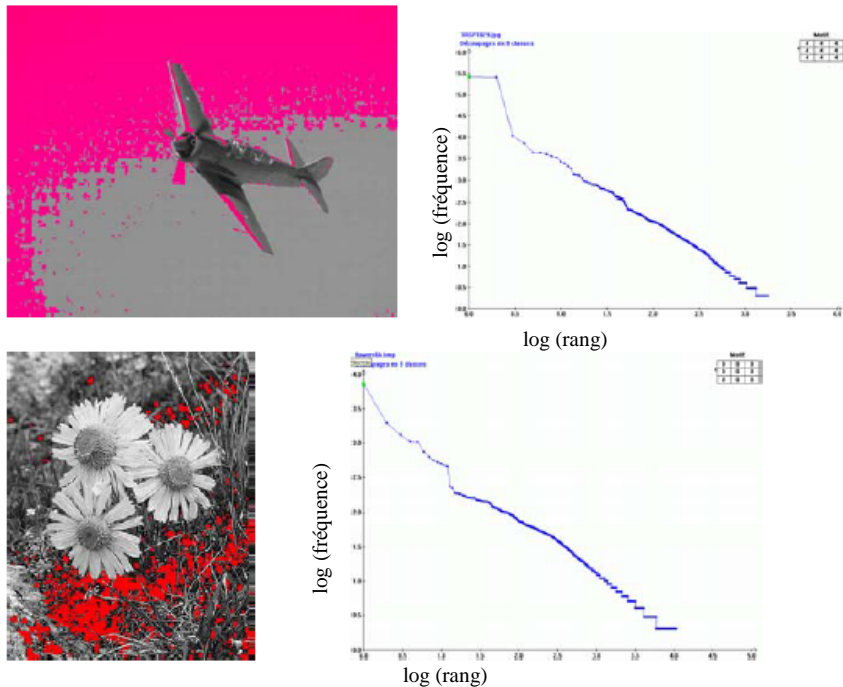


Fig. 11 : Comparaison des courbes en fonction du contenu structurel de l'image.

Nous allons voir maintenant ce qu'il en est pour des images artificielles non photographiques, comme des dessins et des images de synthèse. La figure suivante montre un exemple de résultat obtenu avec ce type d'images et le codage des 9 classes. Les résultats obtenus ne sont pas fondamentalement différents de ceux obtenus sur des images photographiques. En effet, le codage des 9 classes, en réduisant le nombre de valeurs différentes de l'échelle des niveaux de gris, lui donne une apparence proche de celle d'une image artificielle.

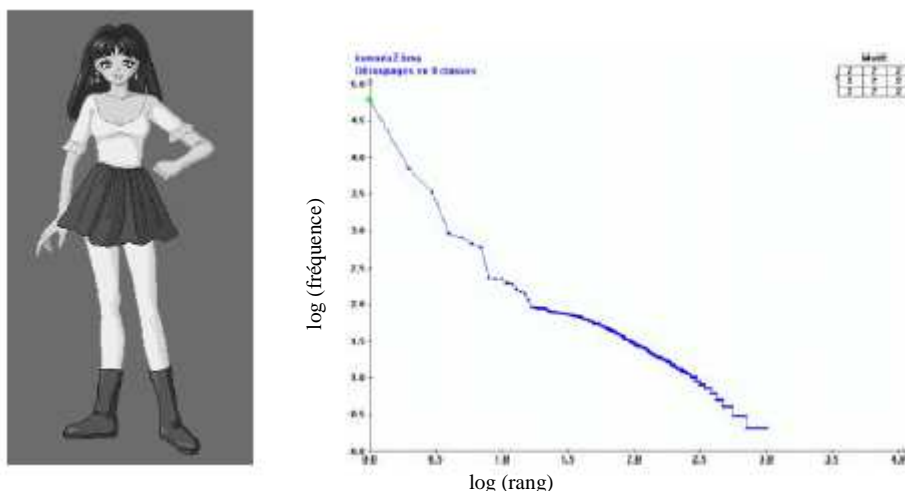


Fig. 12 : Utilisation du codage des 9 classes sur une image artificielle.

3.2 Utilisation du codage des rangs généraux

Comme le codage des 9 classes, le codage des rangs généraux va maintenant être appliqué à différents types d'images, naturelles ou artificielles, représentant des sujets divers et de complexité structurelle plus ou moins grande, afin de déterminer quelles caractéristiques de l'image l'étude de la courbe de distribution rangs/fréquence des motifs permet de mettre en évidence. Des essais effectués avec des images photographiques non compressées avec une taille des images au moins égale à 256x256 pixels ont montré que la courbe de Zipf obtenue avec une telle image est en général linéaire, contrairement à ce que l'on obtient pour le codage des 9 classes et donc que la répartition des motifs de l'image peut effectivement être modélisée par une loi puissance comme le montre la figure ci-dessous.

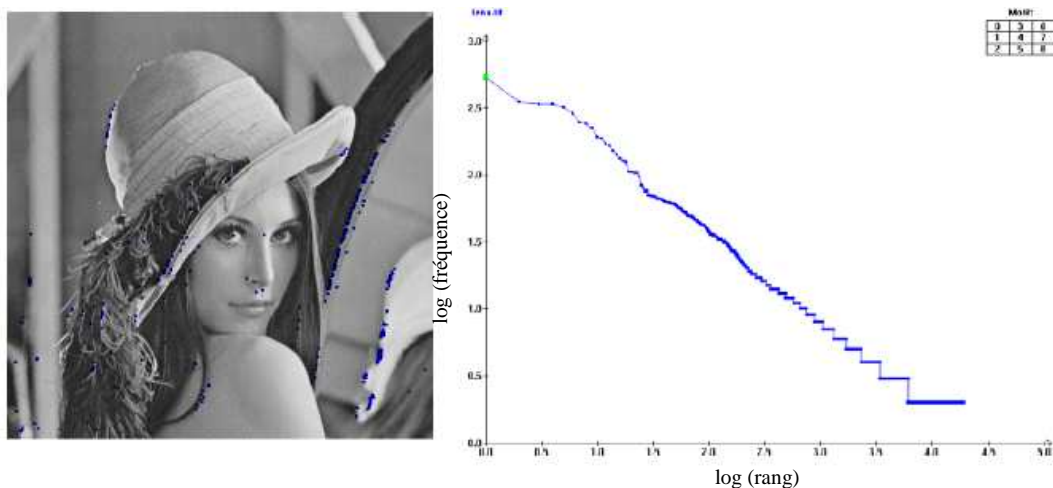


Fig.13 : Courbe de Zipf obtenue avec les rangs généraux

Nous allons maintenant nous attacher à montrer ce que modélise effectivement la courbe de Zipf obtenue à l'aide du codage des rangs généraux. Pour cela, nous allons voir à quoi correspondent les motifs les plus fréquents de l'image.



Fig. 14 : Position des 10 motifs les plus fréquents de l'image Lena.

On remarque que les motifs les plus fréquents, sur la plupart des images photographiques correspondent ici non à des zones homogènes comme dans le cas du codage des 9 classes, mais que, mis à part quelques motifs isolés, ils se regroupent essentiellement sur les contours de l'image, principalement sur les contours rectilignes ou ayant une courbure faible. Lorsque l'on utilise le codage des rangs généraux, les zones homogènes de l'image apparaissent comme formées de nombreux motifs différents sans structuration particulière.

Il arrive cependant, sur quelques images possédant des surfaces très uniformes comme l'image de la figure ci-dessous que le motif le plus fréquent soit le motif uniforme pour lequel tous les niveaux de gris sont égaux, mais le cas est rare sur les images naturelles. De plus, sur une telle image les différentes occurrences du motif uniforme sont dispersées à travers la région homogène et non pas regroupées dans des zones précises.

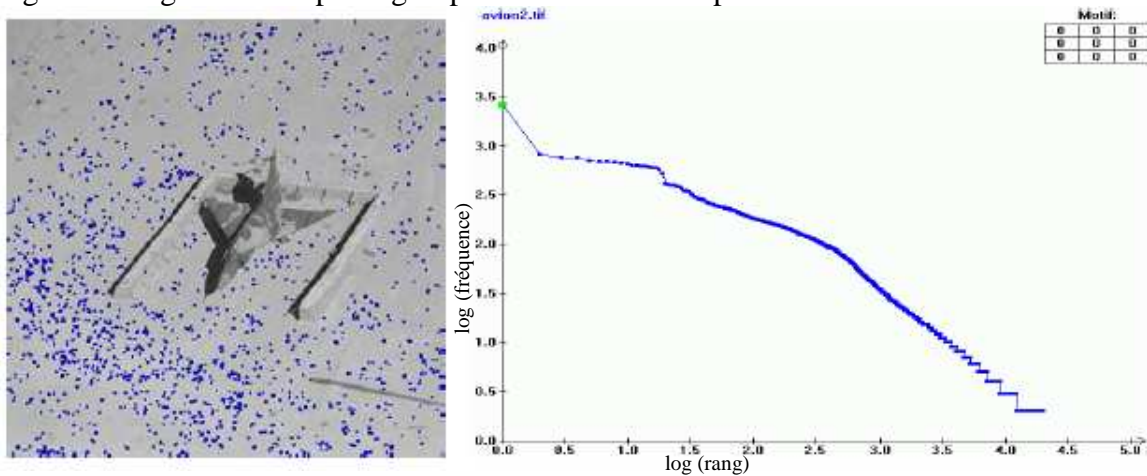


Fig.15 : Cas d'une image naturelle avec un fond très uniforme

Le plus souvent, lorsqu'on observe des concentrations de motifs uniformes sur l'image, ils correspondent en fait à des zones saturées de l'image, qui apparaissent uniformément noires ou blanches sur l'image en niveaux de gris. Dans une telle zone, il est impossible de distinguer les détails à l'intérieur de la zone, qui apparaît alors uniforme.

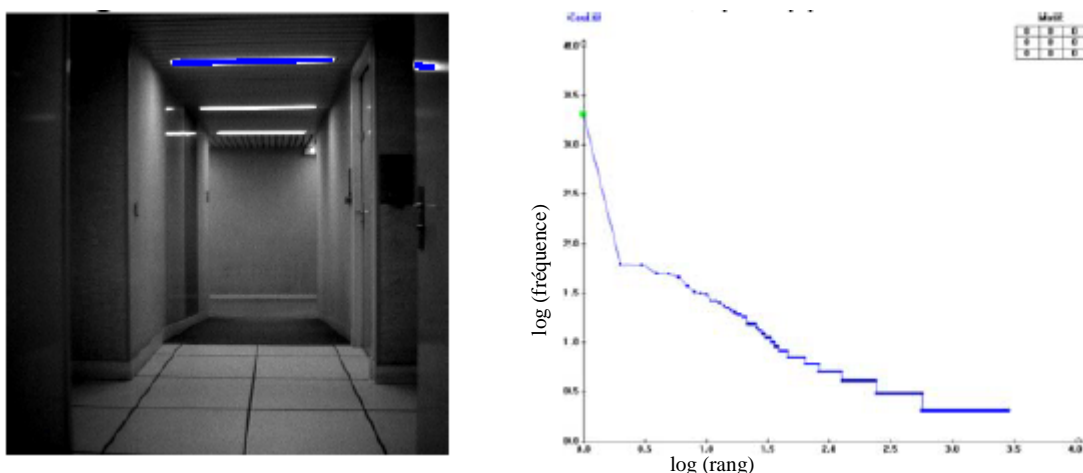


Fig. 16 : Cas d'une image comportant des zones saturées

Dans le cas particulier de l'image Lena, on remarque également des motifs qui se répètent au niveau du bord supérieur de l'image., que l'on retrouve pas sur les autres bords de l'image. Il s'agit ici très vraisemblablement d'un artefact lié à la numérisation de l'image, le bord supérieur de l'image correspondant en fait au bord de la page du magazine dont l'image à été tirée. D'une manière plus générale, les résultats obtenus avec les rangs généraux peuvent être interprétés par le fait que le codage des rangs généraux ne prend pas en compte la luminance des niveaux de gris, mais les différences locales de celle-ci. Ce codage donne en fait une information sur la direction locale du gradient des niveaux de gris. Dans une image naturelle, les zones homogènes ne sont en général pas uniformes, mais présentent en fait de faibles variations locales de luminance, sans direction privilégiée. Ceci se traduit, sur la courbe rang-fréquence des motifs codés avec les rangs généraux, par l'apparition de nombreux motifs différents dont aucun n'est plus fréquent que les autres. Au contraire, sur les contours nets, il existe une variation de luminance importante qui suit une direction privilégiée, celle du contour. Ceci se traduit par l'apparition de motifs qui se répètent et que l'on retrouve alignés le long du contour. Dans le cas des zones saturées négativement ou positivement, la variation de luminance à l'intérieur de la zone n'est plus perceptible du fait de la saturation, la zone apparaît donc totalement uniforme.

Dans le cas des images naturelles, les motifs les plus fréquents ne correspondent donc pas à des zones uniformes, mais à des contours. Nous allons maintenant voir ce qu'il en est dans le cas des images artificielles. Dans le cas d'images simples comme des dessins, ou d'images de synthèse peu texturées, on observe une très nette prédominance du motif uniforme, comme on peut le voir sur la figure ci-dessous. Cela s'explique par le fait que sur une telle image, les surfaces uniformes le sont réellement, c'est à dire qu'elles sont composés de pixels ayant exactement le même niveau de gris, à l'inverse de ce qui se produit pour des images photographiques. Les autres motifs, qui correspondent essentiellement aux contours, apparaissent nettement moins fréquents.

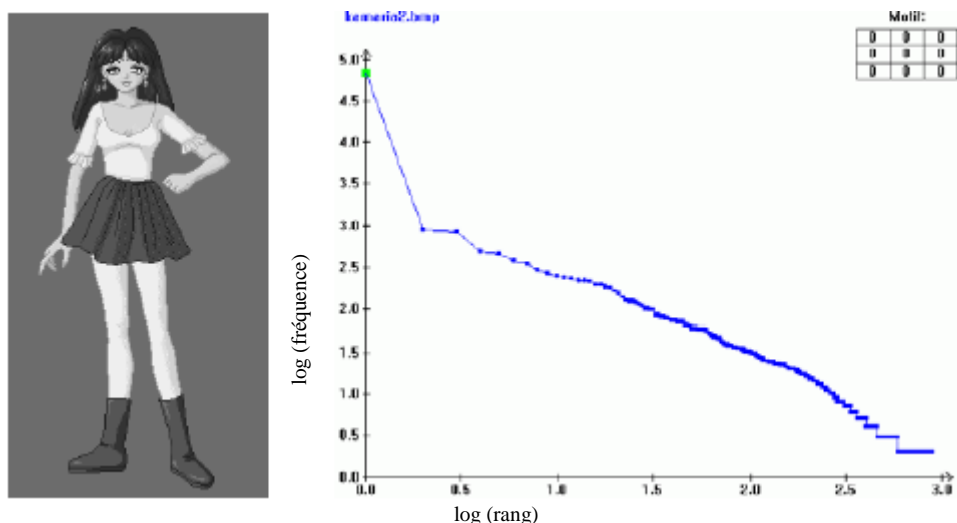


Fig. 17 : Application du codage des rangs généraux sur une image artificielle simple

En revanche, sur des images de synthèse plus complexes comprenant de nombreuses textures différentes, les résultats apparaissent davantage proches de ceux obtenus sur des images

naturelles, en effet les zones homogènes sont peu nombreuses et ce sont essentiellement les contours et les détails fins qui apparaissent sur la courbe.

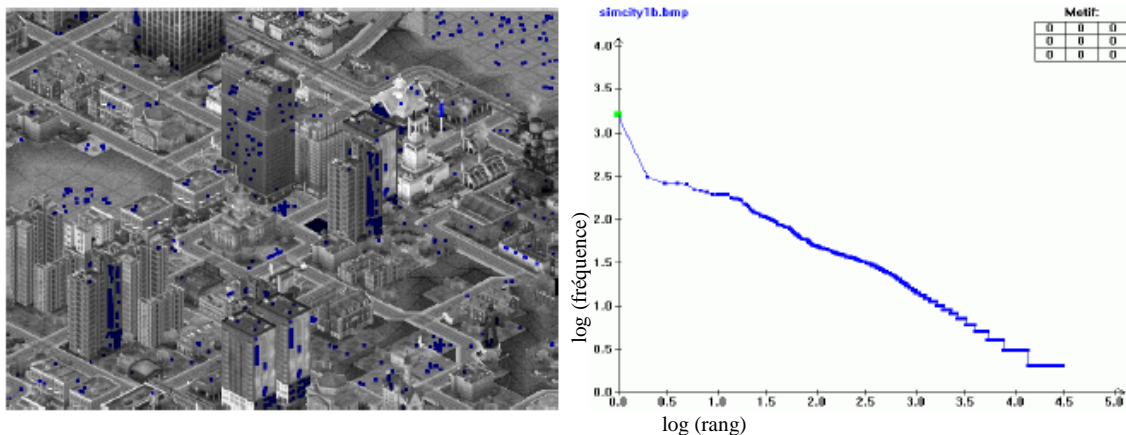


Fig. 18 : Application du codage des rangs généraux sur une image artificielle complexe

On voit donc que les allures des courbes de Zipf sont différentes sur les images naturelles et les images artificielles. Cela tient essentiellement à la structure des zones homogènes qui sont différentes dans les deux types d'images, sur les images artificielles les zones homogènes sont en général totalement uniformes, alors que sur les images naturelles elles comportent de faibles variations de niveaux de gris.

De ce fait, le codage des rangs généraux peut être utilisé pour la détection de retouches d'image. En effet, une retouche d'image consiste à introduire des éléments artificiels dans une image, et l'étude de la répartition spatiale des motifs uniformes doit permettre de détecter ces structures artificielles. La figure ci-dessous montre un exemple d'utilisation de cette méthode pour la détection de retouche d'image.



Fig. 19 : Détection de retouche d'images à l'aide des rangs généraux. Répartition spatiale du motif le plus fréquent sur l'image originale (à gauche) et sur l'image retouchée (à droite)

Sur cette image, la plaque d'immatriculation de la voiture a été changée. On observe alors une concentration de motifs uniformes dans la zone retouchée, ce qui montre le caractère artificiel de cette zone de l'image, contrairement à ce qui se passe pour l'image originale. Cependant, cette méthode de détection des retouches d'images n'a qu'une efficacité relativement limitée, elle ne fonctionne que si la retouche d'image consiste en l'introduction de zones réellement uniformes dans l'image. Il est de ce fait assez facile de tromper cette méthode de détection en

appliquant un bruit ou un effet de texture à la zone retouchée, de ce fait la zone artificielle introduite dans l'image n'est plus uniforme, et sa texture s'apparente alors à une texture naturelle.

3.3 Utilisation de la loi de Zipf inverse

Comme on l'a fait pour la loi de Zipf, on va maintenant appliquer la loi de Zipf inverse sur différentes images afin de déterminer ce que caractérise l'étude de la répartition des motifs les moins fréquents de l'image, tant avec le codage des 9 classes qu'avec le codage des rangs généraux. La figure ci-dessous montre les résultats obtenus avec les deux codages, les points colorés indiquant les motifs n'apparaissant qu'une seule fois dans l'image.

On constate que pour les deux codages, la courbe de Zipf inverse d'une image naturelle est linéaire pour les motifs les moins fréquents de l'image. La distribution du nombre de motifs distincts ayant une fréquence d'apparition donnée suit donc une loi puissance, la loi de Zipf inverse est donc vérifiée en ce qui concerne les images naturelles. On remarque que les motifs les moins fréquents de l'image, avec le codage des 9 classes, se concentrent essentiellement dans les zones riches en détails. Par contre, si l'on utilise le codage des rangs généraux, on constate que sur une image naturelle, la quasi-totalité des motifs de l'image n'apparaissent qu'une seule fois dans l'image, on confirme donc le fait que les zones homogènes des images naturelles ne sont pas totalement uniformes.

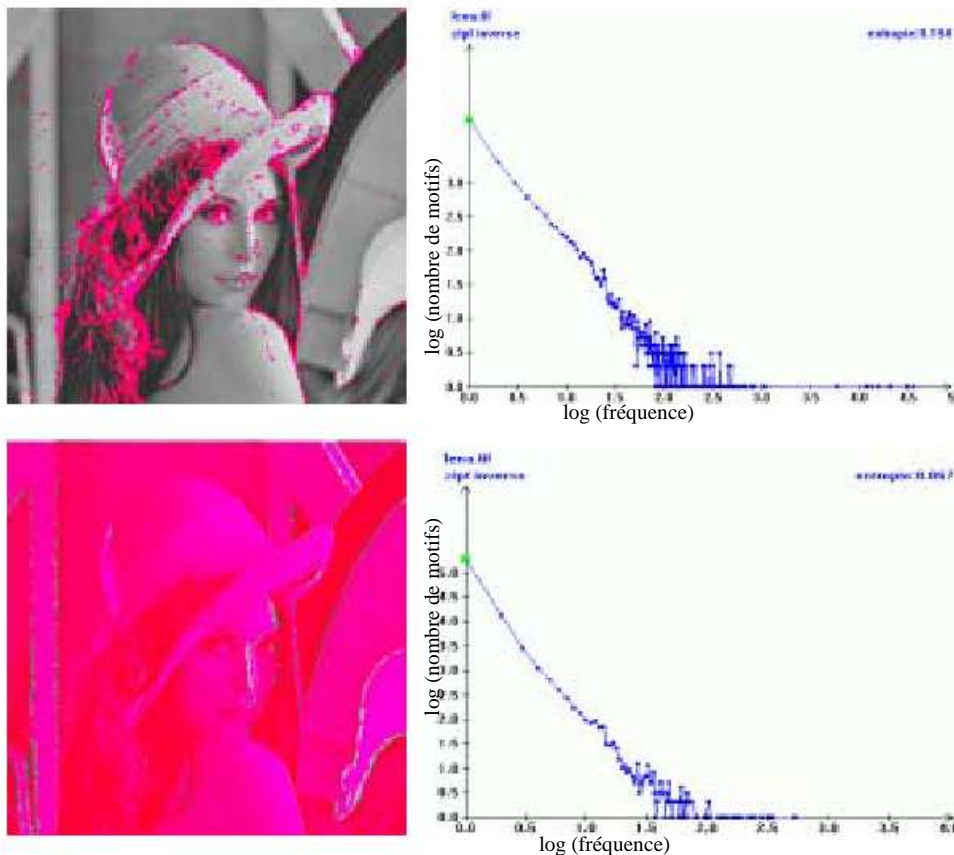


Fig. 20 : Courbe de Zipf inverse obtenue avec le codage des 9 classes (en haut) et celui des rangs généraux (en bas)

La courbe de Zipf inverse a des caractéristiques différentes suivant la complexité de l'image, comme on peut le voir sur la Fig. 21. Dans le cas d'une image comportant beaucoup de zones uniformes, avec le codage des 9 classes, le nombre des motifs n'apparaissant qu'une seule fois dans l'image est relativement petit, en conséquence l'ordonnée à l'origine et la pente de la courbe sont faibles. Dans le cas d'une image plus complexe comportant beaucoup de détails, le nombre de motifs uniques augmente, et par conséquent la pente et l'ordonnée à l'origine de la courbe sont plus élevées.

En revanche, la courbe de Zipf inverse obtenue avec les rangs généraux n'est pas un bon indicateur de la complexité de l'image, les deux courbes obtenues ayant des caractéristiques assez voisines, comme le montre la Fig. 22.

Sur des images artificielles, on constate des résultats analogues à ceux obtenus sur les images naturelles, à savoir que la pente et l'ordonnée de la courbe augmentent avec la complexité de l'image, mais on constate en plus un phénomène très particulier sur les images les plus simples : à savoir que les caractéristiques des courbes ainsi que la répartition spatiale des motifs les moins fréquents de l'image sont pratiquement identiques pour les deux codages. Cela est dû ici encore au fait que les images artificielles comportent des zones homogènes qui soient totalement uniformes, au contraire des images naturelles, et que les motifs réellement uniques sont relativement rares dans les images artificielles, ils se concentrent essentiellement au niveau des angles des contours. Sur des images de synthèse plus complexes, ce phénomène tend à s'atténuer, mais la différence entre les répartition des motifs avec les deux codages reste malgré tout relativement faible

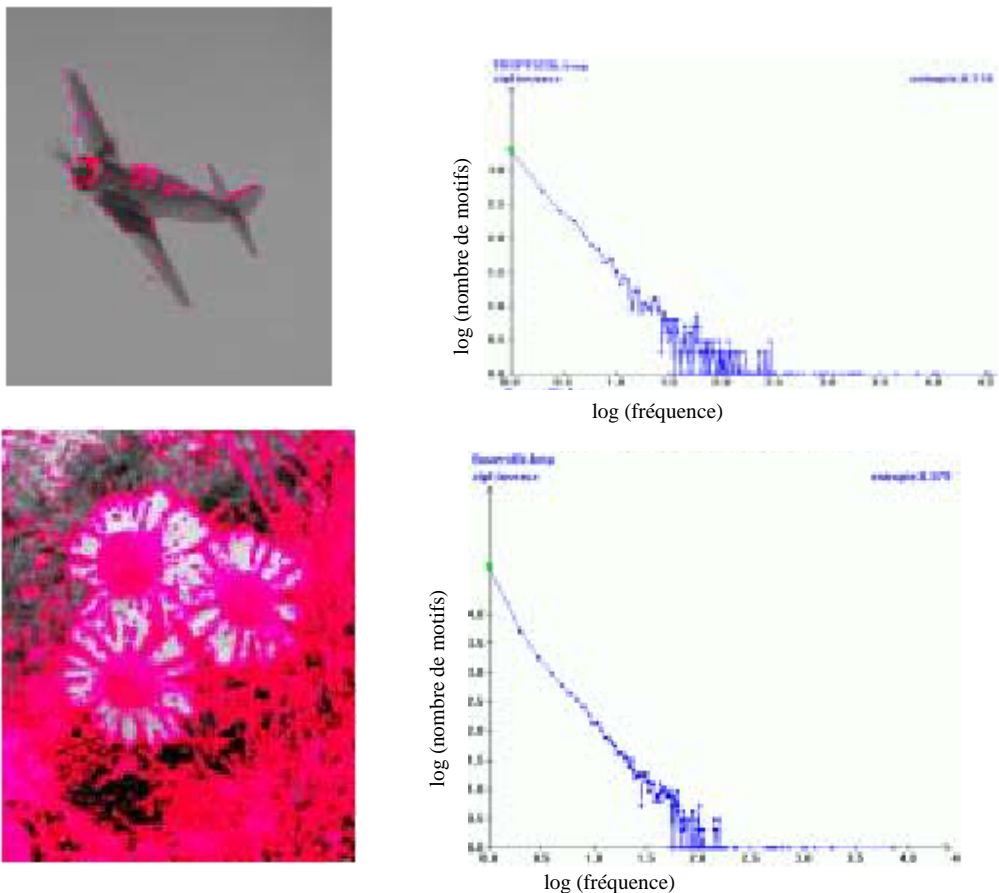


Fig. 21 : Comparaison des courbes obtenues avec le codage des 9 classes suivant la complexité de l'image

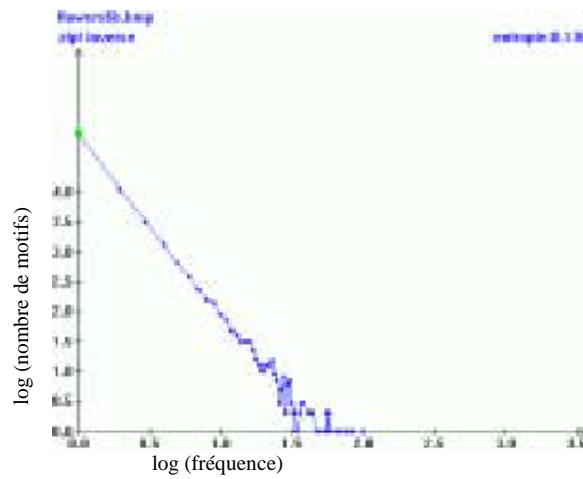
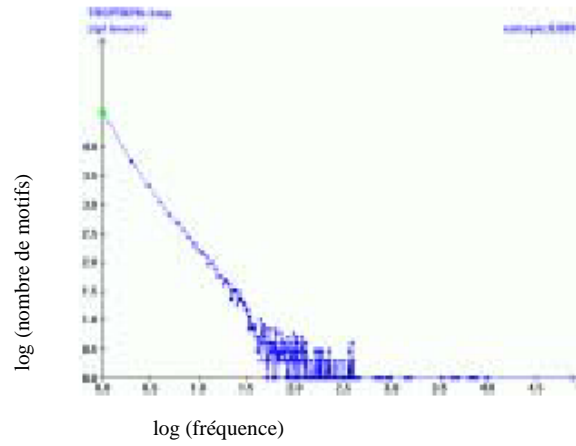
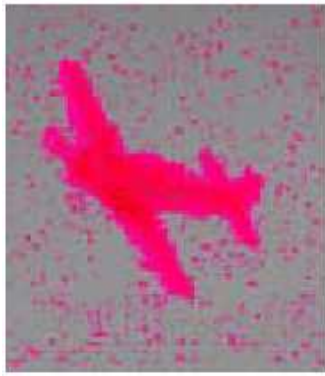


Fig .22 : Comparaison des courbes obtenues avec le codage des rangs généraux suivant la complexité de l'image

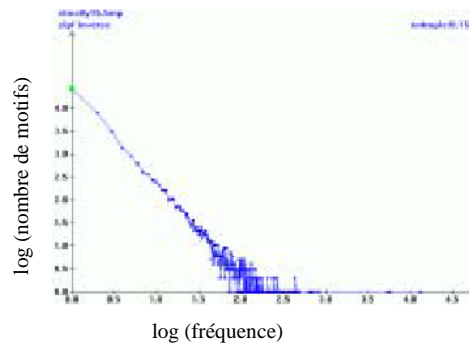
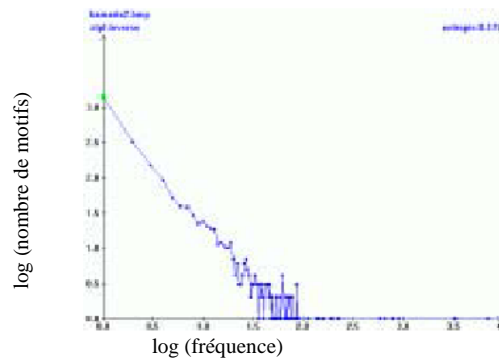


Fig. 23 : Utilisation du codage des 9 classes sur des images de synthèse

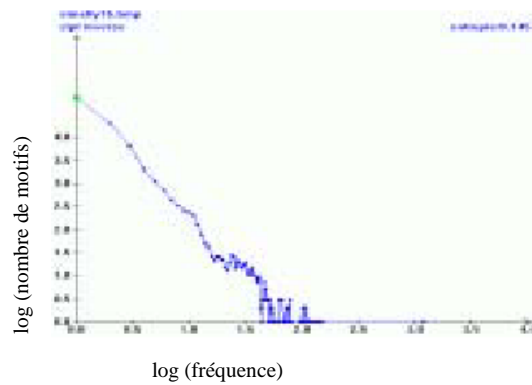
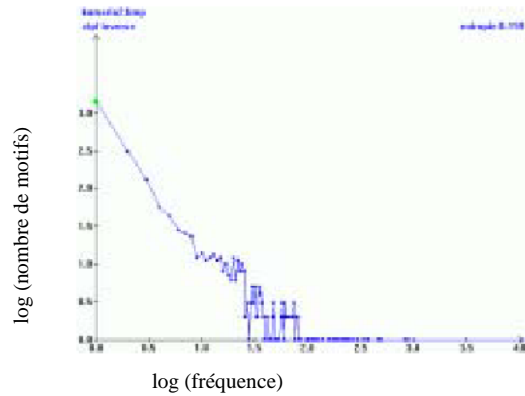


Fig. 24 : Utilisation du codage des rangs généraux sur des images de synthèse

3.4 Bilan

D'une manière générale, on peut dire que les différents codages utilisés pour les lois puissance ne mettent pas en évidence les mêmes caractéristiques de l'image, et leur utilisation ne sera donc pas la même. L'utilisation du codage des 9 classes pour la construction de la courbe de Zipf permet de mettre en évidence les zones homogènes de l'image, elle donne un aperçu de la complexité d'ensemble de l'image. En revanche elle ne permet pas d'étudier les détails fins de l'image. La méthode des rangs généraux, quand à elle, met en évidence la structuration fine de l'image. C'est pourquoi elle permet d'opérer une discrimination entre les images naturelles, c'est à dire les photographies, et les images artificielles, du fait de la structure différente des zones homogènes sur les deux types d'images. Les résultats déduits des courbes de Zipf inverse confirment l'interprétation des faits donnée par la courbe de Zipf, la courbe de Zipf inverse permet également d'évaluer la complexité de l'image selon les caractéristiques de la courbe et de distinguer les images artificielles des images naturelles.

4 Propriétés du modèle

Nous allons maintenant étudier différentes propriétés de la loi de Zipf appliquée aux images. Pour commencer, nous allons étudier l'influence de la résolution de l'image sur la courbe de Zipf associée à cette image, en fonction du codage utilisé pour les motifs. Ensuite, nous allons démontrer dans le cas théorique, l'invariance de la loi de Zipf par rapport à une translation ou à une rotation de l'image. Enfin, nous allons essayer d'établir une relation entre la loi de Zipf et la géométrie fractale.

4.1 Influence de la résolution de l'image

Afin d'évaluer l'influence de la résolution de l'image sur la courbe de Zipf, en fonction du codage utilisé, des essais ont été effectués sur des images dont la taille a été réduite par rééchantillonnage des pixels. Les courbes de Zipf ont ensuite été calculées sur les images avec différents rapports de réduction, en utilisant le codage des 9 classes et le codage des rangs généraux. La figure ci-dessous montre un exemple de résultat obtenu avec l'image Lena réduite à 75%, 50%, 25% et 10% de sa taille initiale de 512 x 512 pixels.

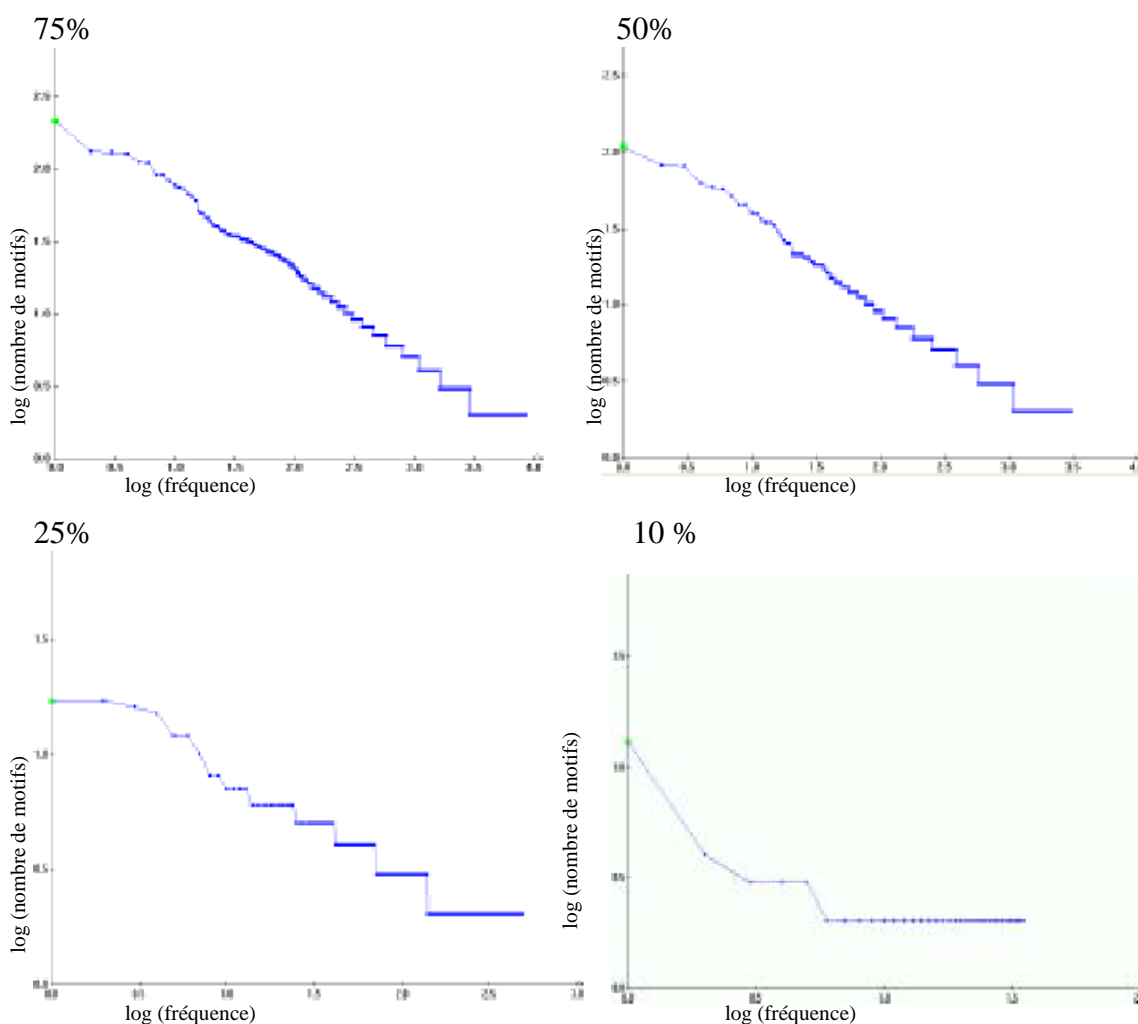


Fig. 25 : Courbes de Zipf obtenues avec le codage des rangs généraux sur plusieurs versions réduites de la même image de référence, Lena.

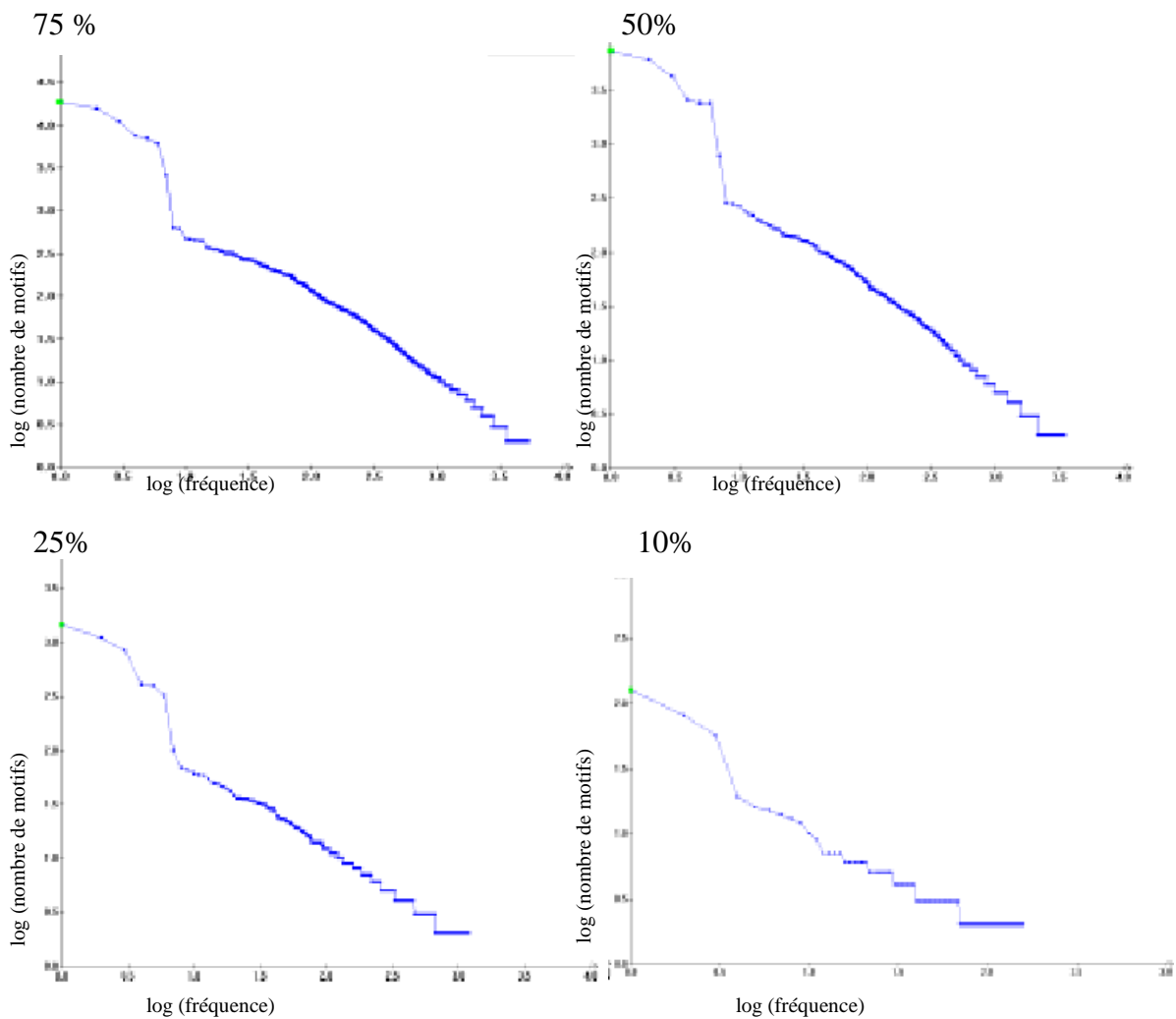


Fig .26 : Courbes de Zipf obtenues avec le codage des 9 classes sur plusieurs versions réduites de la même image de référence.

On peut ici constater que l'allure de la courbe de Zipf d'une image est sensible au changement de résolution, et que cette sensibilité est différente selon les codages. Pour le codage des rangs généraux, une altération de la courbe est perceptible même pour une réduction relativement modérée de l'image, et la dégradation augmente avec la réduction de l'image. Plus spécifiquement, on observe un aplatissement de la courbe lié à la diminution du nombre de motifs. En effet, le codage des rangs généraux s'intéresse principalement aux détails fins de l'image, or la réduction de la taille de celle-ci fait perdre de l'information au niveau des détails, ce qui fait que le codage des rangs généraux n'est pas robuste au changement de résolution de l'image.

Par contre, si l'on utilise le codage des 9 classes, l'allure de la courbe est peu modifiée jusqu'à une réduction de 25% de la taille originale de l'image. En effet, le codage des 9 classes s'intéresse aux zones homogènes de l'image qui sont davantage préservées lors du changement de résolution de l'image.

La méthode des rangs généraux, étant donné qu'elle s'intéresse aux détails fins, est très sensible à la résolution de l'image et nécessite donc de travailler sur des images de haute résolution. Par comparaison, la méthode des 9 classes est beaucoup plus robuste à la résolution de l'image, elle peut donc être utilisée sur des images de plus faible résolution.

4.2 Comportement de la loi de Zipf face à une translation

Dans le cadre de cette étude, on considère une image continue I de taille infinie constituée d'une infinité de points. A chaque point de coordonnées cartésiennes (x,y) du plan de l'image est associé un niveau de gris k représenté sous forme d'un nombre réel. On considère donc l'image comme une fonction continue I de \mathfrak{R}^2 définie par :

$$I: \mathfrak{R}^2 \rightarrow \mathfrak{R}, (x,y) \rightarrow k \quad (3-2-1)$$

On suppose que la répartition des fréquences d'apparition des motifs de l'image I peut être modélisée par une loi de Zipf, c'est à dire que la fréquence d'apparition $N_{\sigma(i)}$ du motif $\delta(i)$ de rang i et dont l'indice dans la suite des fréquences est $\sigma(i)$ est donnée, pour toute valeur de i , par la loi puissance :

$$N_{\sigma(i)} = k * i^a \quad k, a \in \mathfrak{R} \quad (3-2-2)$$

Ce motif $\delta(i)$ centré en (x,y) est défini comme la restriction de l'image I au voisinage d'un point (x,y) du plan de l'image :

$$\delta(i) = I_{V(x,y)} \text{ tel que } I_{V(x,y)} = \{ I(p,q) \mid \forall (p,q) \in V(x,y) \} \quad (3-2-3)$$

Ce voisinage est un sous-ensemble de l'ensemble des parties $P_{\mathfrak{R}^2}$. On peut définir ce voisinage comme étant la boule $B_{(x,y)}$ de centrée en (x,y) et de rayon 1 :

$$B_{(x,y)} = \{ (p,q) \mid \text{t.q. } d[(x,y),(p,q)] \leq 1 \} \text{ avec } (x,y) \in \mathfrak{R}^2 \quad (3-2-4)$$

On va maintenant montrer que cette boule est invariante par translation. On applique à cette boule une translation de vecteur (u,v) . Soit $B_{(x',y')}$ l'image de cette boule par la translation.

$$B_{(x',y')} = \{ (p',q') \mid \text{t.q. } d[(x',y'),(p',q')] \leq 1 \} \text{ avec } (x,y) \in \mathfrak{R}^2 \quad (3-2-5)$$

La translation étant une isométrie, on sait que l'on aura, pour tout point (p,q) de la boule centrée en (x,y) :

$$d[(x',y'),(p',q')] = d[(x,y),(p,q)] \quad (3-2-6)$$

Par conséquent, l'image de tout point de la boule $B(x,y)$ sera contenue dans la boule image $B(x',y')$, et la position relative des points de la boule sera inchangée. La boule est donc invariante par translation.

De ce fait, il est possible de translater toutes les boules afin de les centrer sur l'origine, de manière à pouvoir définir une distance entre les motifs associés. On définit ainsi un espace

des fonctions des restrictions de la fonction initiale de surface I de l'image de base. On définit pour tout point (x,y) une translation $\mathfrak{S}_{(x,y)}$ qui permet de translater la boule $B_{(x,y)}$ au voisinage du point origine (0,0). Pour chaque point $(x,y) \in \mathfrak{R}^2$ on effectue une translation par un vecteur $(-x,-y)$ afin d'amener la boule centrée sur (x,y) à la boule centrée sur (0,0). Cette translation est définie de la manière suivante, pour chaque point :

$$\mathfrak{S}_{(x,y)} : B_{(x,y)} \rightarrow B_{(0,0)}, \text{ tel que } \forall (p,q) \in B_{(x,y)} \mathfrak{S}_{(x,y)}(p,q) = (p-x, q-y) \in B_{(0,0)} \quad (3-2-7)$$

Le motif $\delta(i)$ peut alors se définir comme le translaté $I_{V(x,y)}$ de la boule $B_{(x,y)}$ autour de l'origine. Chaque élément translaté de la boule centrée sur (0,0) aura alors une image définie par la fonction $I_{V(x,y)}$ égale à celle donnée par I pour l'élément correspondant dans la boule centrée sur (x,y).

$$I_{V(x,y)} : \mathfrak{S}_{(x,y)}[B_{(x,y)}] \rightarrow \mathfrak{R}^+ \text{ et } \forall (p,q) \in \mathfrak{S}_{(x,y)}[B_{(x,y)}] \Rightarrow I_{V(x,y)}(p,q) = I(m,n) \quad (3-2-8)$$

où $(m,n) \in B_{(x,y)}, \mathfrak{S}_{(x,y)}(m,n) = (p,q), m=p-x$ et $n=q-y$

En utilisant la notion d'espace de fonctions continues, on peut définir une distance entre motifs, car cet espace est un espace métrique. Dans cet espace de fonctions continues, le motif $I_{V(x,y)}$ peut se définir par la formule suivante :

$$I_{V(x,y)} : B_{(0,0)} \rightarrow \mathfrak{R}^+, B_{(0,0)} = \mathfrak{S}_{(x,y)}[B_{(x,y)}] \quad (3-2-9)$$

Il existe une relation d'ordre entre les points d'un motif $\delta(i)$, tel que pour tout couple de points différents appartenant à un des voisinages qui déterminent le motif, ces deux points ont des valeurs égales par la fonction de surface de l'image et correspondent en même temps à des emplacements relativement égaux sur la surface. Du fait de l'existence de cette relation d'ordre, on peut définir un critère de comparaison entre motifs. Ce critère de comparaison peut être défini par :

$$\forall (x,y), (k,l) \in \mathfrak{R}^2, \text{ avec} \\ I_{V(x,y)} = I_{V(k,l)} \Leftrightarrow d[I_{V(x,y)}, I_{V(k,l)}] = 0 \Rightarrow \sup_{(p,q) \in B_{(0,0)}} (|I_{V(x,y)}(p,q) - I_{V(k,l)}(p,q)|) = 0 \quad (3-2-10)$$

On définit maintenant la fréquence d'apparition du motif $\delta(i)$. Soit $P_{\sigma(i)}$ l'ensemble des occurrences d'un même motif. Entre deux occurrences distinctes d'un même motif, seule la position spatiale change, c'est à dire que $P_{\sigma(i)}$ est un ensemble de boules distinctes $B_{(x_j,y_j)}$ du plan dont les translattées à l'origine sont toutes identiques entre elles. $P_{\sigma(i)}$ est donc défini par :

$$P_{\sigma(i)} = \{B_{(x_j,y_j)} \mid j \in [1.. N_{\sigma(i)}], \forall (k,l) \neq (x_i,y_i), d^2[I_{V(x,y)}, I_{V(k,l)}] = 0, I_{V(x,y)} = I_{V(k,l)} = \delta(i)\} \quad (3-2-11)$$

Par définition la fréquence d'apparition $N_{\sigma(i)}$ est le nombre d'occurrences spatialement distinctes d'un même motif donc la fréquence d'apparition du motif de rang i sera :

$$N_{\sigma(i)} = \text{Card} (P_{\sigma(i)}) \quad (3-2-12)$$

On applique maintenant à l'image I une translation $\mathfrak{S}_{(u,v)}$ de vecteur (u,v). Soit I' l'image transformée. Les fréquences d'apparition des motifs de l'image en fonction du rang sont données par la loi puissance :

$$N'_{\sigma(i)} = k' * i^{-a'} \quad (3-2-13)$$

Pour que la courbe de Zipf soit invariante par translation, il faut et il suffit que :

- 1) La transformée d'un motif contienne exactement tous les images de tous les points du motif d'origine
- 2) Les positions relatives des points d'un même motif soient inchangées
- 3) Deux motifs distincts dans l'image d'origine restent distincts dans l'image transformée
- 4) La fréquence d'apparition d'un motif est la même dans l'image originale et dans l'image transformée
- 5) La restriction de la fonction surface I de l'image initiale $I_{V(x,y)}$, associée à une boule centrée sur un point (x,y) qui va donner un motif $\delta(j)$ donne après transformation, un nouveau motif $\delta'(j)$ de même fréquence d'apparition $N_{\sigma(j)} = N_{\sigma'(j)}$.

Les propositions 1) et 2) se déduisent facilement du fait que la translation soit une isométrie.

Pour toute boule $B_{(x,y)}$ de centre (x,y) , un point (p,q) appartient à la boule si et seulement si la distance $d[(p,q),(x,y)]$ de ce point au centre de la boule est inférieure à 1. Soient (x',y') et (p',q') les images des points (x,y) et (p,q) par la translation $\mathfrak{T}_{(u,v)}$ de vecteur (u,v) . La translation étant une isométrie, on aura donc :

$$d[(p',q'),(x',y')] = d[(p,q),(x,y)] \quad \forall (x,y), \forall (p,q) \in B_{(x,y)}, \quad (3-2-14)$$

Par conséquent, on aura :

$$d[(p',q'),(x',y')] \leq 1 \Leftrightarrow d[(p,q),(x,y)] \leq 1 \quad (3-2-15)$$

L'image de tout point de la boule originale $B_{(x,y)}$ sera contenue dans la boule image $B_{(x',y')}$. De même, si le point (p,q) n'est pas contenu dans la boule $B_{(x,y)}$, on aura alors :

$$d[(p,q),(x,y)] > 1 \Leftrightarrow d[(p',q'),(x',y')] > 1 \quad (3-2-16)$$

Donc, la transformée $B_{(x',y')}$ de la boule $B_{(x,y)}$ contient les images de tous les points de la boule originale.

De même, la distance relative de deux points quelconques (m,n) et (k,l) de $B_{(x,y)}$ reste inchangée, en effet on a :

$$d[(m,n),(k,l)] = d[(m',n'),(k',l')] \quad \forall (m,n) \in B_{(x,y)}, \forall (k,l) \in B_{(x,y)} \quad (3-2-17)$$

La proposition 3 signifie que deux motifs distincts de l'image d'origine auront des images distinctes dans l'image transformée, c'est à dire que l'on aura :

$$\forall (x,y) \neq (p,q) \in \mathfrak{R}^2 \text{ avec } I_{V(x,y)} \neq I_{V(p,q)} \Leftrightarrow I'_{V(x',y')} \neq I'_{V(p',q')} \quad (3-2-18)$$

Or, deux motifs sont différents si et seulement si la distance entre eux est non nulle, c'est à dire que l'on a :

$$I_{V(x,y)} \neq I_{V(p,q)} \Leftrightarrow d[(I_{V(x,y)}, I_{V(p,q)})] \neq 0 \quad (3-2-19)$$

Or, d'après les propriétés de la translation, on sait que l'on a :

$$d[(I'_{V(x',y')}, I'_{V(p',q')})] = d[(I_{V(x,y)}, I_{V(p,q)})] \quad (3-2-20)$$

et donc :

$$d[(I'_{V(x',y')}, I'_{V(p',q')})] \neq 0 \Leftrightarrow d[(I_{V(x,y)}, I_{V(p,q)})] \neq 0 \quad (3-2-21)$$

Par conséquent, on a :

$$I_{V(x,y)} \neq I_{V(p,q)} \Leftrightarrow I'_{V(x',y')} \neq I'_{V(p',q')} \quad (3-2-22)$$

La proposition 3 est donc vérifiée.

D'après la proposition 4, si un motif $\delta(j)$ a une fréquence d'apparition $N_{\sigma(j)}$, son image $\delta'(j)$ par une translation aura la même fréquence d'apparition $N_{\sigma(j)}$, c'est à dire que l'on aura :

$$N'_{\sigma'(j)} = \text{Card}(P'_{\sigma'(j)}) = \text{Card}(P_{\sigma(j)}) = N_{\sigma(j)} \quad (3-2-23)$$

Avec :

$$P_{\sigma(j)} = \{ (x_i, y_i) \mid \forall (k, l) \neq (x_i, y_i) \in \mathfrak{R}^2, \forall i \in N, I_{V(x_i, y_i)}, d^2[I_{V(x_i, y_i)}, I_{V(k, l)}] = 0 \text{ et } I_{V(k, l)} = \delta(j) \} \quad (3-2-24)$$

$$P'_{\sigma'(j)} = \{ (x'_i, y'_i) \mid \forall (k', l') \neq (x'_i, y'_i) \in \mathfrak{R}^2, \forall i \in N, I_{V(x'_i, y'_i)}, d^2[I_{V(x'_i, y'_i)}, I_{V(k', l')}] = 0 \text{ et } I_{V(k', l')} = \delta'(j) \} \quad (3-2-25)$$

$$(k', l') = \mathfrak{S}_{(u, v)}(k, l), (x_i, y_i) = \mathfrak{S}_{(u, v)}(x_i, y_i), \forall i \in N \quad (3-2-26)$$

$$\forall (p, q) \in P_{\sigma(j)} \Leftrightarrow (p', q') \in P'_{\sigma'(j)}, \text{ avec } (p', q') = \mathfrak{S}_{(u, v)}(p, q) \quad (3-2-27)$$

On va d'abord démontrer que l'image (p', q') de toute instance (p, q) d'un motif de l'ensemble $P_{\sigma(j)}$ par la translation de vecteur (u, v) appartient à l'ensemble des instances des motifs transformés $P'_{\sigma'(j)}$, c'est à dire que l'on a :

$$\forall (p, q) \in P_{\sigma(j)} \Rightarrow (p', q') \in P'_{\sigma'(j)} \quad (3-2-28)$$

Si l'on suppose le contraire, c'est à dire que l'on a $(p', q') \notin P'_{\sigma'(j)}$, il existe alors un motif (m', n') de l'ensemble $P'_{\sigma'(j)}$ tel que l'on ait :

$$d^2[I'_{V(m', n')}, I'_{V(p', q')}] \neq 0 \text{ avec } (m', n') = \mathfrak{S}_{(u, v)}(m, n) \text{ et } (m, n) \in P_{\sigma(j)} \quad (3-2-29)$$

D'après la proposition 3, si $d^2[I'_{V(m',n')}, I'_{V(p',q')}] \neq 0$ alors $d^2[I_{V(m,n)}, I_{V(p,q)}] \neq 0$, donc les motifs $I_{V(m,n)}$ et $I_{V(p,q)}$ sont différents, or par définition l'ensemble $P_{\sigma(j)}$ est constitué de toutes les instances d'un unique motif. Dans ce cas, on aurait alors $(p,q) \notin P_{\sigma(j)}$, ce qui est faux car cela contredit l'hypothèse

Par conséquent, on aurait $(p,q) \notin P_{\sigma(j)}$, ce qui est faux car cela contredit l'hypothèse.

Donc, $(p',q') \in P'_{\sigma'(j)}$, c'est à dire que l'ensemble des instances de l'image d'un motif par la rotation contient au moins toutes les images des instances du motif initial.

On va maintenant démontrer la proposition réciproque :

$$\forall (p,q) \in P_{\sigma(j)} \Leftrightarrow (p',q') \in P'_{\sigma'(j)}, \text{ avec } (p,q) = \mathfrak{S}_{(u,v)}(p',q') \quad (3-2-30)$$

Si l'on suppose le contraire, à savoir que (p,q) n'appartient pas à l'ensemble $P_{\sigma(j)}$, cela veut dire qu'il existe un point (m,n) tel que :

$$d^2[I_{V(m,n)}, I_{V(p,q)}] \neq 0, \text{ avec } (m,n) = \mathfrak{S}_{(u,v)}(m',n') \text{ et } (m',n') \in P'_{\sigma'(j)} \quad (3-2-31)$$

D'après la proposition 3, si la distance $d^2[I_{V(m,n)}, I_{V(p,q)}]$ est non nulle alors la distance $d^2[I'_{V(m',n')}, I'_{V(p',q')}]$ est également non nulle, par conséquent les motifs $I_{V(m',n')}$ et $I_{V(p',q')}$ sont différents, par conséquent on aurait $(p,q) \notin P_{\sigma(j)}$ ce qui est faux car cela contredit l'hypothèse.

Par conséquent, $(p,q) \in P_{\sigma(j)}$, c'est à dire que les antécédents des instances d'un même motif dans l'image transformée sont les instances d'un même motif dans l'image initiale.

Par conséquent, les ensembles des instances d'un motif dans l'image originale et dans l'image transformée contiennent le même nombre de motifs points, donc les motifs ont la même fréquence d'apparition $N_{\sigma(j)}$ dans les deux images. On a donc démontré la proposition 4.

D'après la proposition 5, la restriction de la fonction surface I de l'image initiale $I_{V(x,y)}$, associée à une boule centrée sur un point (x,y) qui va donner un motif $\delta(j)$ donne un motif $\delta'(j)$ de même fréquence d'apparition $N_{\sigma(j)}$ par une translation de vecteur (u,v) .

Soit un point $(x,y) \in \mathfrak{R}^2$ et une de boule $B_{(x,y)}$ de rayon 1 est centrée sur (x,y) .

On définit la fonction de translation $\mathfrak{S}_{(x,y)} : B_{(x,y)} \rightarrow B_{(0,0)}$ tel que, pour tout point (p,q) de la boule $B_{(x,y)}$, on ait :

$$\forall (p,q) \in B_{(x,y)} \text{ on a } \mathfrak{S}_{(x,y)}(p,q) = (p-x, q-y) = (m,n) \in B_{(0,0)} \quad (3-2-32)$$

La fonction motif $I_{V(x,y)}$ associée au point (x,y) est définie par :

$$I_{V(x,y)} : B_{(0,0)} \rightarrow \mathfrak{R}^2 \text{ tel que } \forall (m,n) \in B_{(0,0)}, \text{ avec } (m,n) = \mathfrak{S}_{(x,y)}(p,q) \text{ et } (p,q) \in B_{(x,y)} \quad (3-2-33)$$

$$I_{V(x,y)}(m,n) = I(p,q) = I(\mathfrak{S}_{(x,y)}^{-1}(m,n)) \quad (3-2-34)$$

La position des points (p',q') et (x',y') , images de (p,q) et (x,y) par la translation $\mathfrak{S}_{(u,v)}$ est donnée par :

$$(p',q') = \mathfrak{S}_{(u,v)}(p,q) = (p+u, q+v) \text{ avec } I'(p',q') = I(p,q) \quad (3-3-35)$$

$$(x',y') = \mathfrak{S}_{(u,v)}(x,y) = (x+u, y+v) \text{ avec } I'(x',y') = I(x,y) \quad (3-3-36)$$

On a, à travers la fonction de translation $\mathfrak{S}_{(x',y')} : B_{(x',y')} \rightarrow B_{(0,0)}$

$$\mathfrak{S}_{(x',y')}(x',y') = (0,0) \quad (3-3-37)$$

$$\text{et } \mathfrak{S}_{(x',y')}(p',q') = (p'-x', q'-y') = (m',n') \in B_{(0,0)} \quad (3-3-38)$$

On définit la fonction motif $I_{V(x',y')}$ associée au point (x',y') :

$$I_{V(x',y')} : B_{(0,0)} \rightarrow \mathfrak{R}^2 \text{ tel que } \forall (m',n') \in B_{(0,0)} \quad (3-3-39)$$

$$\text{avec } (m',n') = \mathfrak{S}_{(x',y')}(p',q'), \text{ où } (p',q') \in B_{(x',y')} \quad (3-3-40)$$

$$I_{V(x',y')}(m',n') = I(p',q') = I(\mathfrak{S}_{(x',y')}^{-1}(m',n')) \quad (3-3-41)$$

Comme $I(p',q') = I(p,q)$, on en déduit que :

$$I_{V(x',y')}(m',n') = I(p,q), \quad \forall (p,q) \in B_{(x,y)} \quad (3-3-42)$$

Donc, tout point qui appartient à la boule $B_{(x,y)}$ centrée sur (x,y) qui par translation donne un point (m,n) dans la boule $B_{(0,0)}$ centrée sur $(0,0)$ de taille 1, donne le même point si l'on applique à l'image une translation $\mathfrak{S}_{(u,v)}$ de vecteur (u,v) . La proposition 5 est donc démontrée.

Par conséquent, la courbe de Zipf associée à une image est invariante par translation.

4.3 Comportement de la loi de Zipf face à une rotation

Après avoir démontré l'invariance de la courbe de Zipf dans le cas d'une translation, nous allons maintenant étudier son comportement si on applique une rotation. On utilisera la même modélisation mathématique que dans le cas de la translation. On va maintenant appliquer à une rotation à notre image. On définit donc une rotation plane \mathfrak{R}_α d'angle $\alpha \in [-\pi, \pi]$ définie comme suit :

$$\mathfrak{R}_\alpha : \mathfrak{R}^2 \rightarrow \mathfrak{R}^2, \text{ tel que } \forall (x,y) \in \mathfrak{R}^2 \rightarrow \mathfrak{R}_\alpha(x,y) \in \mathfrak{R}^2 \text{ et } I(x,y) = I'(x',y') \quad (3-3-1)$$

La matrice de rotation est définie par :

$$M(\mathfrak{R}_\alpha) = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \quad (3-3-2)$$

\mathfrak{R}_α est bijective, c'est à dire qu'il existe une rotation inverse \mathfrak{R}_α^{-1} d'angle $-\alpha$.

L'image I est soumise à la rotation R_α . Soit I' l'image transformée de I par cette rotation R_α . Nous allons maintenant démontrer que la courbe de Zipf associée à une image est invariante par rotation, c'est à dire que la loi de distribution rang-fréquence des motifs de l'image I' transformée par la rotation R_α est identique à celle de l'image originale I. Si la loi de Zipf est invariante par rotation, la distribution des fréquences des motifs de l'image transformée I' sera donc décrite par la loi puissance $N'_{\sigma(i)} = k' * i'^a$ telle que l'on ait :

$$N'_{\sigma(i)} = k' * i'^a = k * i^a = N_{\sigma(i)} \quad (3-3-3)$$

Pour démontrer l'invariance de la loi de Zipf à la rotation il suffit de démontrer les propositions suivantes :

Pour que la courbe de Zipf soit invariante par rotation, il faut et il suffit que :

- 1) La transformée d'un motif contienne exactement toutes les images de tous les points du motif d'origine
- 2) Les positions relatives des points d'un même motif soient inchangées
- 3) Deux motifs distincts dans l'image d'origine restent distincts dans l'image transformée
- 4) La fréquence d'apparition d'un motif est la même dans l'image originale et dans l'image transformée
- 5) La restriction de la fonction surface I de l'image initiale $I_{V(x,y)}$, associée à une boule centrée sur un point (x,y) qui va donner un motif $\delta(j)$ donne après transformation, un nouveau motif $\delta'(j)$ de même fréquence d'apparition $N_{\sigma(j)} = N_{\sigma'(j)}$.

Pour démontrer l'invariance par rotation de la courbe de Zipf associée à une image, il faut et il suffit de démontrer ces cinq propositions. Si la courbe de Zipf est invariante par rotation, cela signifie que la fréquence d'apparition $N_{\sigma(i)}$ d'un motif d'un certain rang i est invariable, mais que l'ordre des motifs varie. Nous allons maintenant tenter de démontrer chacune de ces cinq propositions.

Les propositions 1 et 2 se déduisent du fait que la rotation est une isométrie. Soit la boule $B_{(x,y)}$ centrée en (x,y). Un point (p,q) appartient à la boule si et seulement si la distance $d^2[(p,q),(x,y)]$ de ce point au centre de la boule est inférieure à 1. Soient (x',y') et (p',q') les images des points (x,y) et (p,q) par la rotation \mathfrak{R}_α . La rotation étant une isométrie, on aura donc :

$$d^2[(p',q'),(x',y')] = d^2[(p,q),(x,y)] \quad \forall (x,y), \forall (p,q) \in B_{(x,y)}, \quad (3-3-4)$$

Par conséquent, si le point (p,q) appartient à la boule (x,y), on aura :

$$d^2[(p',q'),(x',y')] \leq 1 \Leftrightarrow d^2[(p,q),(x,y)] \leq 1 \quad (3-3-5)$$

L'image de tout point de la boule originale $B_{(x,y)}$ sera contenue dans la boule image $B_{(x',y')}$. De même, si le point (p,q) n'est pas contenu dans la boule $B_{(x,y)}$, on aura alors :

$$d^2[(p,q),(x,y)] > 1 \Leftrightarrow d^2[(p',q'),(x',y')] > 1 \quad (3-3-6)$$

Donc, la transformée $B_{(x',y')}$ de la boule $B_{(x,y)}$ contient les images de tous les points de la boule originale. La proposition 1 est donc vérifiée. De même, la distance relative de deux points quelconques (m,n) et (k,l) de $B_{(x,y)}$ reste inchangée, en effet on a :

$$d^2[(m,n),(k,l)] = d^2[(m',n'),(k',l')] \quad \forall (m,n) \in B_{(x,y)}, \forall (k,l) \in B_{(x,y)} \quad (3-3-7)$$

On vérifie donc la proposition 2. Deux points situés à une même distance relative sur l'image originale I resteront à cette distance relative dans l'image I' après la rotation. Ces deux points qui donnent un même point dans la boule $B_{(0,0)}$, donneront le même point dans cette boule après rotation, simplement tournés d'un angle α par rapport à leur position originale.

La proposition 3 signifie que deux motifs différents $I_{V(x,y)}$ et $I_{V(p,q)}$ centrés sur deux points distincts (x,y) et (p,q) de l'image I auront des images différentes dans l'image transformée I' . Soit $(x',y') = \mathfrak{R}_\alpha(x,y)$ et $(p',q') = \mathfrak{R}_\alpha(p,q)$. et soit $I'_{V(x',y')}$ et $I'_{V(p',q')}$ les images de ces deux motifs par la rotation \mathfrak{R}_α , on aura donc :

$$\forall (x,y) \neq (p,q) \in \mathfrak{R}^2 \text{ avec } I_{V(x,y)} \neq I_{V(p,q)} \Leftrightarrow I'_{V(x',y')} \neq I'_{V(p',q')} \quad (3-3-8)$$

On va d'abord démontrer que l'on a :

$$\forall (x,y) \neq (p,q) \in \mathfrak{R}^2 \text{ avec } I_{V(x,y)} \neq I_{V(p,q)} \Rightarrow I'_{V(x',y')} \neq I'_{V(p',q')}. \quad (3-3-9)$$

Si l'on suppose le contraire, à savoir que deux motifs différents ont une image identique, la distance entre les deux motifs translétés à l'origine sera nulle, donc on aura

$$I'_{V(x',y')} = I'_{V(p',q')} \Rightarrow d^2[I'_{V(x',y')}, I'_{V(p',q')}] = 0 \quad (3-3-10)$$

Les motifs transformés sont définis par :

$$\begin{aligned} I'_{V(x',y')} : \mathfrak{S}_{(x',y')}[B_{(x',y')}] &\rightarrow \mathfrak{R}^+ \quad \mathfrak{S}_{(x',y')} : B_{(x',y')} \rightarrow B_{(0,0)}, \\ \forall (m',n') \in B_{(x',y')} &\rightarrow \mathfrak{S}_{(x',y')}(m',n') = (m' - x', n' - y') \in B_{(0,0)} \end{aligned} \quad (3-3-11)$$

et

$$\begin{aligned} I'_{V(p',q')} : \mathfrak{S}_{(p',q')}[B_{(p',q')}] &\rightarrow \mathfrak{R}^+ \quad \mathfrak{S}_{(p',q')} : B_{(p',q')} \rightarrow B_{(0,0)}, \\ \forall (k',l') \in B_{(p',q')} &\rightarrow \mathfrak{S}_{(p',q')}(k',l') = (k' - p', l' - q') \in B_{(0,0)} \end{aligned} \quad (3-3-12)$$

avec :

$$I'_{V(x',y')}(m',n') = I'(m',n') \text{ et } I'_{V(p',q')}(k',l') = I'(k',l') \quad (3-3-13)$$

Pour tout point (a',b') de la boule $B_{(0,0)}$, on peut définir un point (m',n') de la boule $B_{(x',y')}$ tel que l'on ait $(a',b') = \mathfrak{S}_{(x',y')}(m',n')$. La restriction de l'image I' au point (a',b') est alors

$$I'_{V(x',y')}(a',b') = I'(m',n') \quad (3-3-14)$$

De même, il existe un point (k,l) de la boule $B_{(p',q')}$ tel que $(a',b') = \mathfrak{S}_{(p',q')}(k',l')$. On a alors

$$I'_{V(p',q')}(a',b') = I'(k',l') \quad (3-3-15)$$

Si les deux motifs ont une image identique, on aura alors $I'(m',n')=I'(k',l')$. Par conséquent on aura :

$$I'_{V(x',y')}(a',b') = I'_{V(p',q')}(a',b') \quad (3-3-16)$$

Donc la distance entre les deux motifs sera :

$$d^2[I'_{V(x',y')}, I'_{V(p',q')}] = \sup_{(a',b') \in B(0,0)} (| I'_{V(x',y')}(a',b') - I'_{V(p',q')}(a',b') |) = 0 \quad (3-3-17)$$

Donc deux points des voisinages $B_{(x',y')}$ et $B_{(p',q')}$ qui donnent par translation un même point dans la boule $B_{(0,0)}$, vont avoir des valeurs égales sur l'image transformée I' .

On considère maintenant les points (m,n) et (k,l) qui sont les antécédents de (m',n') et (k',l') , c'est à dire que l'on a $(m,n) = \mathfrak{R}^{-1}_\alpha(m',n')$ et $(k,l) = \mathfrak{R}^{-1}_\alpha(k',l')$.

On sait que l'image de tout point d'une boule $B_{(x,y)}$ appartient à la boule image $B_{(x',y')}$, c'est à dire que l'on a : $\forall (x,y) \in \mathfrak{R}^2$, avec $(x',y') = \mathfrak{R}_\alpha(x,y) \Leftrightarrow B_{(x,y)} = B_{(x',y')}$ Par conséquent les points (m,n) et (k,l) vont appartenir respectivement à la boule $B_{(x,y)}$ et à la boule $B_{(p,q)}$, avec $(x,y) = \mathfrak{R}^{-1}_\alpha(x',y')$ et $(p,q) = \mathfrak{R}^{-1}_\alpha(p',q')$,

De plus, on a, pour tout point des deux voisinages :

$$I(m,n) = I(m',n') \text{ et } I(k,l) = I(k',l') \quad (3-3-18)$$

Donc, il résulte que pour tout point (m,n) de la boule $B_{(x,y)}$ et pour tout point (k,l) de la boule $B_{(p,q)}$ on a :

$$I(m,n)=I(k,l) \quad (3-3-19)$$

On considère maintenant les motifs $I_{V(x,y)}$ et $I_{V(p,q)}$ associés aux points (x,y) et (p,q) de l'image originale :

$$I_{V(x,y)} : \mathfrak{S}_{(x,y)}[B_{(x,y)}] \rightarrow \mathfrak{R}^+$$

$$, \mathfrak{S}_{(x,y)} : B_{(x,y)} \rightarrow B_{(0,0)}, \forall (m,n) \in B_{(x,y)} \rightarrow \mathfrak{S}_{(x,y)}(m,n) = (m-x, n-y) \in B_{(0,0)}$$

$$\text{et } I_{V(x,y)} : \mathfrak{S}_{(p,q)}[B_{(p,q)}] \rightarrow \mathfrak{R}^+$$

$$\mathfrak{S}_{(p,q)} : B_{(p,q)} \rightarrow B_{(0,0)}, \forall (k,l) \in B_{(p,q)} \rightarrow \mathfrak{S}_{(p,q)}(k,l) = (k-p, l-q) \in B_{(0,0)} \quad (3-3-20)$$

$$\text{avec } I_{V(x,y)}(m,n) = I(m,n) \text{ et } I_{V(p,q)}(k,l) = I(k,l). \quad (3-3-21)$$

$$I_{V(x,y)}(m,n) = I_{V(p,q)}(k,l), \forall (m,n) \in B_{(x,y)} \text{ et } \forall (k,l) \in B_{(p,q)}. \quad (3-3-22)$$

Dans ce cas, on aurait alors :

$$d^2[I_{V(x,y)}, I_{V(p,q)}] = \sup_{(a,b) \in B(0,0)} | I_{V(x,y)}(a,b) - I_{V(p,q)}(a,b) | = 0 \quad (3-3-23)$$

Ce qui est faux car on a fait l'hypothèse que les motifs étaient différents dans l'image originale

Par conséquent on aura donc :

$$I'_{V(x',y')} \neq I'_{V(p',q')} \quad (3-3-23)$$

C'est à dire que tous les motifs différents avant la rotation dans l'image originale I , resteront différents dans l'image transformée I' .

On va maintenant démontrer que tous les motifs différents dans l'image transformée I' ont pour antécédents des motifs différents dans l'image originale I .

$$\forall (x,y) \neq (p,q) \in \mathfrak{R}^2 \text{ avec } I_{V(x,y)} \neq I_{V(p,q)} \Leftarrow I'_{V(x',y')} \neq I'_{V(p',q')}. \quad (3-3-24)$$

On suppose le contraire, c'est à dire que les deux motifs sont identiques dans l'image originale, et donc que leur distance est nulle. Dans ce cas on a :

$$I_{V(x,y)} = I_{V(p,q)} \Rightarrow d^2[I_{V(x,y)}, I_{V(p,q)}] = 0 \Rightarrow \sup_{(a,b) \in B(0,0)} |I_{V(x,y)}(a,b) - I_{V(p,q)}(a,b)| = 0 \quad (3-3-25)$$

(3-2-19)

On en déduit donc que :

$$\forall (a,b) \in B(0,0) \quad I_{V(x,y)}(a,b) = I_{V(p,q)}(a,b) \Rightarrow \exists ! (m,n) \in B_{(x,y)} \text{ et } (k,l) \in B_{(p,q)} \quad (3-3-26)$$

avec

$$I_{V(x,y)}(a,b) = I(m,n) \text{ et } I_{V(p,q)}(a,b) = I(k,l) \text{ tel que } I(m,n) = I(k,l). \quad (3-3-27)$$

Donc, avant la rotation on a par hypothèse deux motifs identiques entre eux dans l'image originale, ce qui se traduit par l'égalité des niveaux de gris pour deux points dans chaque voisinage $B_{(x,y)}$ et $B_{(p,q)}$. Ces points correspondent en fait, dans la boule $B_{(0,0)}$ à un même point, mais obtenu à travers deux fonctions différentes de translation : $\mathfrak{S}_{(x,y)}$ et $\mathfrak{S}_{(p,q)}$.

On fait subir la rotation \mathfrak{R}_α aux boules $B_{(x,y)}$ et $B_{(p,q)}$. On sait qu'elles vont donner dans l'image transformée I' deux boules $B'_{(x',y')}$ et $B'_{(p',q')}$ de centre (x',y') et (p',q') telles que $(x',y') = \mathfrak{R}_\alpha(x,y)$ et $(p',q') = \mathfrak{R}_\alpha(p,q)$.

Soit deux points $(m,n) \in B_{(x,y)}$ et $(k,l) \in B_{(k,l)}$ tel que $I(m,n) = I(k,l)$. Si on applique la rotation \mathfrak{R}_α , on aura donc :

$$(m',n') = \mathfrak{R}_\alpha(m,n), \quad I(m,n) = I'(m',n') \text{ et } (k',l') = \mathfrak{R}_\alpha(k,l), \quad I(k,l) = I'(k',l') \quad (3-3-28)$$

Par conséquent on aura $I'(m',n') = I'(k',l')$, et ceci pour n'importe quel couple de points des voisinages $B_{(x,y)}$ et $B_{(p,q)}$, dont les images par les translations $\mathfrak{S}_{(x,y)}$ et $\mathfrak{S}_{(p,q)}$ donnent un même point dans la boule $B_{(0,0)}$.

Cela implique que pour tous points (m',n') de $B_{(x',y')}$ et (k',l') de $B_{(p',q')}$ tels que leurs antécédents $(m,n) = \mathfrak{R}_\alpha^{-1}(m',n')$ et $(k,l) = \mathfrak{R}_\alpha^{-1}(k',l')$ soient tels que $I(m,n) = I(k,l)$ on aura :

$$I'(m',n') = I'(k',l'). \quad (3-3-29)$$

Les images de ces points par les translations $\mathfrak{S}_{(x',y')}$ et $\mathfrak{S}_{(p',q')}$ sont définies par :

$$I'_{V(x',y')} : \mathfrak{S}_{(x',y')}[B_{(x',y')}] \rightarrow \mathfrak{R}^+, \quad \mathfrak{S}_{(x',y')} : B_{(x',y')} \rightarrow B_{(0,0)},$$

$$\forall (m',n') \in B_{(x',y')} \rightarrow \mathfrak{S}_{(x',y')}(m',n') = (m' - x', n' - y') \in B_{(0,0)} \quad (3-3-30)$$

et

$$I'_{V(p',q')} : \mathfrak{S}_{(p',q')}[B_{(p',q')}] \rightarrow \mathfrak{R}^+, \quad \mathfrak{S}_{(p',q')} : B_{(p',q')} \rightarrow B_{(0,0)},$$

$$\forall (k',l') \in B_{(p',q')} \rightarrow \mathfrak{S}_{(p',q')}(k',l') = (k' - p', l' - q') \in B_{(0,0)} \quad (3-3-31)$$

$$\text{avec } I'_{V(x',y')}(m',n') = I'(m',n') \text{ et } I'_{V(p',q')}(k',l') = I'(k',l'). \quad (3-3-32)$$

On définit le point $(a', b') \in B_{(0,0)}$ tel que $(a', b') = \mathfrak{Z}_{(x', y')}(m', n')$ et $(a', b') = \mathfrak{Z}_{(p', q')}(k', l')$.

Donc, par définition des fonctions de translation, $\forall (a', b') \in B_{(0,0)}$ on a

$$I'_{V(x', y')}(m', n') = I'_{V(p', q')}(k', l') \quad (3-3-33)$$

Dans ces conditions, la distance entre les deux motifs est donnée par :

$$d^2[I'_{V(x', y')}, I'_{V(p', q')}] = \sup_{(a', b') \in B_{(0,0)}} [I'_{V(x', y')}(a', b') - I'_{V(p', q')}(a', b')] = 0 \quad (3-3-34)$$

La proposition (3-3-34) est fautive car elle contredit l'hypothèse de départ. Par conséquent, si la distance entre les antécédents des deux motifs est non nulle, les deux motifs seront différents, et donc on aura :

$$d^2[I_{V(x, y)}, I_{V(p, q)}] \neq 0 \Rightarrow I_{V(x, y)} \neq I_{V(p, q)} \quad (3-3-35)$$

Donc, au moins tous les motifs différents dans l'image transformée auront des antécédents différents dans l'image initiale.

D'après (3-3-9) et (3-3-35) on en déduit que :

$$\forall (x, y) \neq (p, q) \in \mathfrak{R}^2 \text{ avec } I_{V(x, y)} \neq I_{V(p, q)} \Leftrightarrow I'_{V(x, y)} \neq I'_{V(p, q)} \quad (3-3-36)$$

avec $(x', y') = \mathfrak{R}_\alpha(x, y)$ et $(p', q') = \mathfrak{R}_\alpha(p, q)$. On a donc démontré la proposition 3.

La proposition 4 peut s'énoncer de la façon suivante. Si on a un motif $\delta(j)$ qui a une fréquence d'apparition $N_{\sigma(j)}$, après la rotation, le nouveau motif $\delta'(j)$ constitué de l'ensemble des points du motif $\delta(j)$ transformés par la rotation aura la même fréquence d'apparition $N_{\sigma(j)}$, c'est à dire que l'on aura :

$$N'_{\sigma'(j)} = \text{Card}(P'_{\sigma'(j)}) = \text{Card}(P_{\sigma(j)}) = N_{\sigma(j)} \quad (3-3-37)$$

Avec :

$$P_{\sigma(j)} = \{(x_i, y_i) \mid \forall (k, l) \neq (x_i, y_i) \in \mathfrak{R}^2, \forall i \in N, I_{V(x_i, y_i)}, d^2[I_{V(x_i, y_i)}, I_{V(k, l)}] = 0 \text{ et } I_{V(k, l)} = \delta(j)\} \quad (3-3-38)$$

$$P'_{\sigma'(j)} = \{(x'_i, y'_i) \mid \forall (k', l') \neq (x'_i, y'_i) \in \mathfrak{R}^2, \forall i \in N, I_{V(x'_i, y'_i)}, d^2[I_{V(x'_i, y'_i)}, I_{V(k', l')}] = 0 \text{ et } I_{V(k', l')} = \delta'(j)\} \quad (3-3-39)$$

$$\text{Où } (k', l') = \mathfrak{R}_\alpha(k, l), (x_i, y_i) = \mathfrak{R}_\alpha(x_i, y_i), \forall i \in N \quad (3-3-40)$$

$$\forall (p, q) \in P_{\sigma(j)} \Leftrightarrow (p', q') \in P'_{\sigma'(j)}, \text{ avec } (p', q') = \mathfrak{R}_\alpha(p, q). \quad (3-3-41)$$

On cherche d'abord à démontrer que pour toute instance (p, q) d'un motif de l'ensemble $P_{\sigma(j)}$, son image (p', q') par la rotation \mathfrak{R}_α appartient à l'ensemble des instances des motifs transformés $P'_{\sigma'(j)}$, c'est à dire que l'on a :

$$\forall (p, q) \in P_{\sigma(j)} \Rightarrow (p', q') \in P'_{\sigma'(j)}, \text{ avec } (p', q') = \mathfrak{R}_\alpha(p, q) \quad (3-3-42)$$

On suppose le contraire. Si l'on suppose le contraire, c'est à dire que l'on a $(p', q') \notin P'_{\sigma'(j)}$, il existe alors un motif (m', n') de l'ensemble $P'_{\sigma'(j)}$ tel que l'on ait

$$d^2[I'_{V(m',n')}, I'_{V(p',q')}] \neq 0 \text{ avec } (m',n') = \mathfrak{R}_\alpha(m,n) \text{ et } (m,n) \in P_{\sigma(j)} \quad (3-3-43)$$

D'après la proposition 3, si $d^2[I'_{V(m',n')}, I'_{V(p',q')}] \neq 0$ alors $d^2[I_{V(m,n)}, I_{V(p,q)}] \neq 0$, donc les motifs $I_{V(m,n)}$ et $I_{V(p,q)}$ sont différents, or par définition l'ensemble $P_{\sigma(j)}$ est constitué de toutes les instances d'un unique motif.

Par conséquent, on aurait $(p,q) \notin P_{\sigma(j)}$, ce qui est faux car cela contredit l'hypothèse.

Donc, $(p',q') \in P'_{\sigma'(j)}$, c'est à dire que l'ensemble des instances de l'image d'un motif par la rotation contient au moins toutes les images des instances du motif initial.

On va maintenant démontrer la proposition réciproque :

$$\forall (p,q) \in P_{\sigma(j)} \Leftarrow (p',q') \in P'_{\sigma'(j)}, \text{ avec } (p,q) = \mathfrak{R}^{-1}_\alpha(p',q') \quad (3-3-44)$$

Si l'on suppose le contraire, à savoir que (p,q) n'appartient pas à l'ensemble $P_{\sigma(j)}$, cela veut dire qu'il existe un point (m,n) tel que :

$$d^2[I_{V(m,n)}, I_{V(p,q)}] \neq 0, \text{ avec } (m,n) = \mathfrak{R}^{-1}_\alpha(m',n') \text{ et } (m',n') \in P'_{\sigma'(j)} \quad (3-3-45)$$

D'après la proposition 3, si la distance $d^2[I_{V(m,n)}, I_{V(p,q)}]$ est non nulle alors la distance $d^2[I'_{V(m',n')}, I'_{V(p',q')}]$ est également non nulle, par conséquent les motifs $I_{V(m',n')}$ et $I_{V(p',q')}$ sont différents, par conséquent on aurait $(p,q) \notin P_{\sigma(j)}$ ce qui est faux car cela contredit l'hypothèse.

Donc, $(p,q) \in P_{\sigma(j)}$, c'est à dire que les antécédents des instances d'un même motif dans l'image transformée sont les instances d'un même motif dans l'image initiale.

D'après les relations (3-4-6) et (3-4-10), les ensembles correspondants à un motifs et à la transformée de ce motif par la rotation ont le même nombre de points, donc la même fréquence d'apparition $N_{\sigma(j)}$.

Nous allons maintenant démontrer la proposition 5. Cette proposition nous dit que la restriction de la fonction surface I de l'image initiale $I_{V(x,y)}$, associée à une boule centrée sur un point (x,y) qui donne un motif $\delta(j)$, après une rotation d'angle α , va donner un nouveau motif $\delta'(j)$ de même fréquence d'apparition $N_{\sigma(j)}$.

Soit un point $(x,y) \in \mathfrak{R}^2$ et une de boule $B_{(x,y)}$ de rayon 1 est centrée sur (x,y) .

On définit la fonction de translation $\mathfrak{S}_{(x,y)} : B_{(x,y)} \rightarrow B_{(0,0)}$ tel que, pour tout point (p,q) de la boule $B_{(x,y)}$, on ait :

$$\forall (p,q) \in B_{(x,y)} \text{ on a } \mathfrak{S}_{(x,y)}(p,q) = (p-x, q-y) = (m,n) \in B_{(0,0)} \quad (3-3-46)$$

On peut définir la fonction motif $I_{V(x,y)}$ associée au point (x,y) :

$$I_{V(x,y)} : B_{(0,0)} \rightarrow \mathfrak{R}^2 \text{ tel que } \forall (m,n) \in B_{(0,0)}, \text{ avec } (m,n) = \mathfrak{S}_{(x,y)}(p,q), \text{ où } (p,q) \in B_{(x,y)} \quad (3-3-47)$$

$$I_{V(x,y)}(m,n) = I(p,q) = I(\mathfrak{S}_{(x,y)}^{-1}(m,n)) \quad (3-3-48)$$

On applique une rotation \mathfrak{R}_α d'angle α . La position des points (p',q') et (x',y') , images de (p,q) et (x,y) est donnée par :

$$(p',q') = \mathfrak{R}_\alpha(p,q) = (p \cdot \cos \alpha - q \cdot \sin \alpha, p \cdot \sin \alpha + q \cdot \cos \alpha) \text{ avec } I'(p',q') = I(p,q) \quad (3-3-49)$$

$$(x',y') = \mathfrak{R}_\alpha(x,y) = (x \cdot \cos \alpha - y \cdot \sin \alpha, x \cdot \sin \alpha + y \cdot \cos \alpha) \text{ avec } I(x',y') = I(x,y) \quad (3-3-50)$$

On a, à travers la fonction de translation $\mathfrak{S}_{(x',y')} : B_{(x',y')} \rightarrow B_{(0,0)}$

$$\mathfrak{S}_{(x',y')}(x',y') = (0,0) \quad (3-3-51)$$

$$\text{et } \mathfrak{S}_{(x',y')}(p',q') = (p' - x', q' - y') = (m', n') \in B_{(0,0)} \quad (3-3-52)$$

On définit la fonction motif $I'_{V(x',y')}$ associée au point (x',y') :

$$I_{V(x',y')} : B_{(0,0)} \rightarrow \mathfrak{R}^2 \text{ tel que } \forall (m',n') \in B_{(0,0)} \quad (3-3-53)$$

$$\text{avec } (m',n') = \mathfrak{S}_{(x',y')}(p',q'), \text{ où } (p',q') \in B_{(x',y')} \quad (3-3-54)$$

$$I'_{V(x',y')}(m',n') = I'(p',q') = I(\mathfrak{S}_{(x',y')}^{-1}(m',n')) \quad (3-3-55)$$

Comme $I(p',q') = I(p,q)$, on en déduit que :

$$I_{V(x',y')}(m',n') = I(p,q), \quad \forall (p,q) \in B_{(x,y)} \quad (3-3-56)$$

Donc, tout point qui appartient à la boule $B_{(x,y)}$ centrée sur (x,y) qui par translation donne un point (m,n) dans la boule $B_{(0,0)}$ centrée sur $(0,0)$ de taille 1, donne après une rotation d'angle α un point (m',n') de même valeur du niveau de gris que le point (p,q) et qui se trouve tourné d'un angle α dans l'espace de la boule $B_{(0,0)}$. En fait on retrouve dans le motif $I_{V(x',y')}$ tous les niveaux de gris qui constituent le motif original $I_{V(x,y)}$, mais associés à des points tournés d'un angle α par rapport à leur position dans l'image originale.

De ces 5 propositions que l'on vient de démontrer, on déduit que la courbe de Zipf associée à une image est invariante par rotation. La distribution des fréquences des motifs d'une image en fonction de leurs rangs est identique dans une image transformée par une rotation à ce qu'elle était dans l'image originale.

On peut remarquer que l'invariance de la courbe de Zipf par rotation ou par translation découle essentiellement du fait que la rotation, tout comme la translation, est une isométrie, ce résultat peut être généralisé à toute transformation du plan qui soit une isométrie.

Après avoir montré l'invariance de la courbe de Zipf par rotation dans le cas purement théorique d'une image continue, nous allons voir ce qu'il en est dans la pratique.

Dans la pratique, on travaille sur des images numériques qui par nature sont discrètes, et qui comptent un nombre fini de pixels. Ces pixels sont représentés sous forme d'une matrice

rectangulaire. Une telle image n'est pas invariante par rotation d'un angle α quelconque, mais seulement par des multiples de 90° . Donc, l'image sera invariante uniquement pour des rotations dont l'angle est un multiple de 90° .

Nous allons maintenant vérifier ce qu'il en est pour la courbe de Zipf associée à une image. La figure suivante montre les courbes de Zipf associées à une même image avant et après une rotation de 90° vers la droite, obtenue avec le codage par les rangs généraux. On remarque que les deux courbes de Zipf sont absolument identiques, on vérifie donc l'invariance de la courbe de Zipf dans le cas d'une rotation de 90° . On obtient le même résultat pour toute rotation d'un multiple de 90° dans un sens ou dans l'autre. Si l'on s'intéresse aux motifs de la courbe, on constate que pour tout motif de rang i de la courbe originale, le motif de même rang dans la courbe associée à l'image transformée est le motif original tourné de 90° . On vérifie donc les résultats théoriques attendus dans le cas particulier d'une rotation par un multiple de 90°

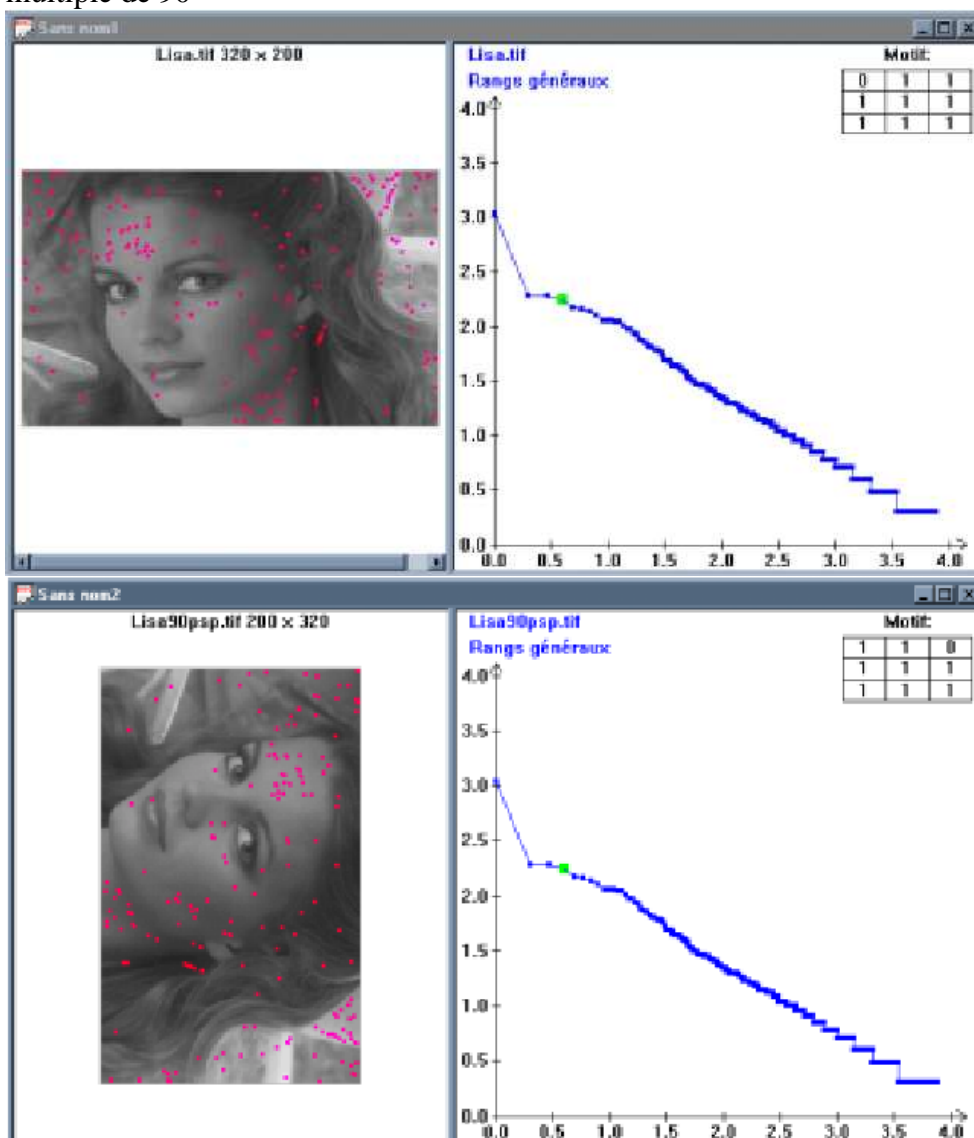


Fig. 27 : Courbes de Zipf d'une image avant et après une rotation de 90°

Nous allons maintenant voir ce qui se produit pour une rotation d'un angle qui ne soit pas un multiple de 90° . Dans ce cas, l'image n'est plus invariante par rotation. En effet, la rotation introduit des erreurs d'arrondis et des approximations sur les pixels de l'image qui se traduisent par l'apparition de trous dans l'image. Ces trous sont régulièrement espacés et forment une trame de pixels noirs qui se superposent à l'image, le motif de cette trame étant différent suivant l'angle de rotation. La figure 28 montre un exemple de résultat obtenu avec une rotation de -30° .

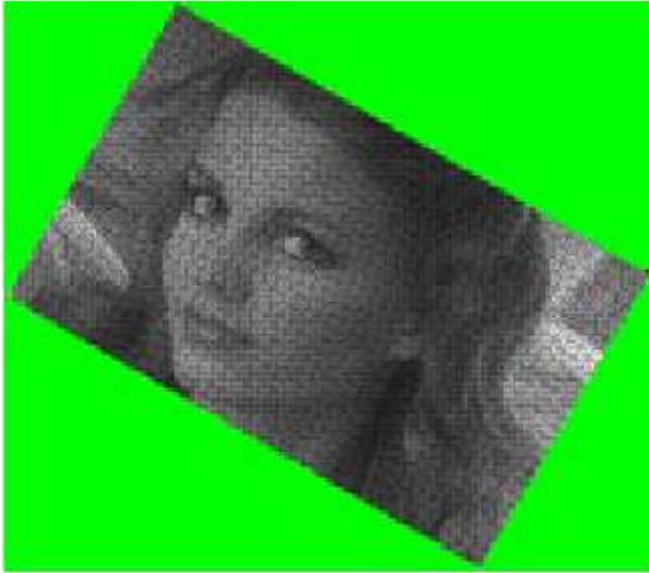


Fig 28 : Image ayant subi une rotation d'un angle de -30°

Il est possible de supprimer cet effet de trame sur l'image transformée en appliquant un lissage de l'image. On peut utiliser pour cela un filtre moyenneur ou un filtrage médian. On reconstitue ainsi l'apparence de l'image initiale. Cependant, l'image filtrée n'est pas identique à l'image initiale, il ne s'agit que d'une approximation de celle-ci. La distribution des motifs de l'image sera donc différente de celle de l'image originale. La courbe de Zipf associée à l'image transformée sera donc différente de celle de l'image initiale, car les deux images sont différentes. La figure 29 montre les courbes de Zipf obtenues sur l'image précédente avec différents angles de rotation. On remarque que la pente de la droite de régression de la courbe de Zipf diminue si l'on applique une rotation à la courbe, cette pente étant minimale pour un angle de 45° .

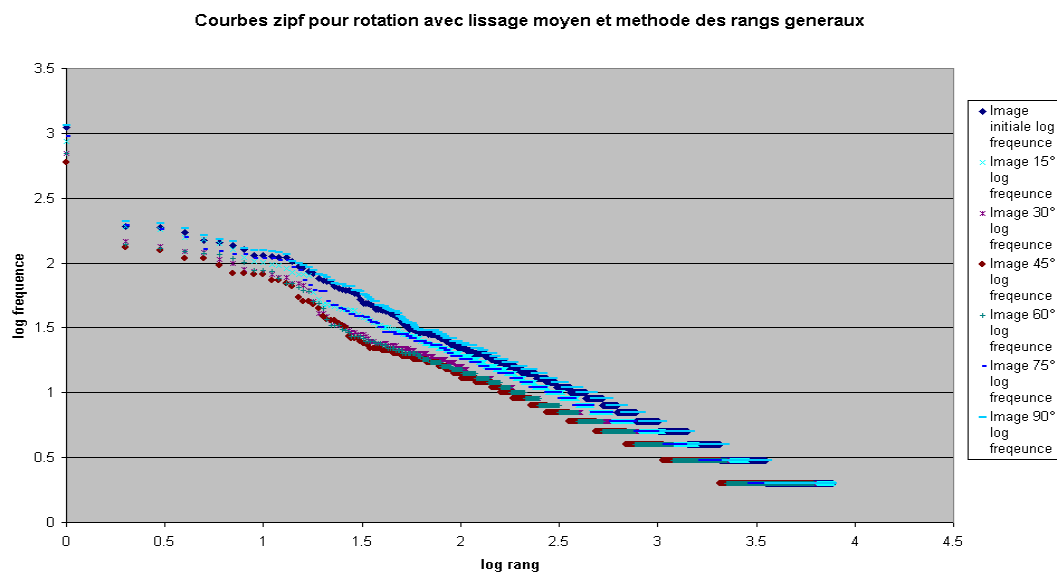


Fig. 29: Exemple de courbes de Zipf obtenues après rotation et filtrage pour différents angles de rotation.

Pour conclure, on peut dire que si l'on a montré que la courbe de Zipf est invariante par rotation dans la cas théorique, en revanche dans la pratique, sur des images numériques, cette invariance ne peut être vérifiée que pour des rotations par des multiples de 90° . Dans tous les autres cas, l'invariance par rotation n'est pas vérifiée car l'image elle-même n'est pas invariante par rotation, il est en effet impossible d'appliquer une rotation sans altérer l'image.

4.4 Relations entre la loi de Zipf et la géométrie fractale

Il est possible d'établir une relation entre la notion de dimension fractale définie par Mandelbrot [Man-83] et la loi de Zipf. Cette relation a déjà été évoquée dans le chapitre précédent en ce qui concerne l'interprétation donnée par Mandelbrot dans le cas des fréquences d'appartion des mots dans le textes. Nous allons voir ici une relation analogue dans un cas plus proche de celui de la fréquence d'appartion des motifs d'une image. Cette relation a été établie par Gaité et Manrubia [Gai-02] dans le cadre de travaux ayant pour but d'étudier la distribution de la taille des espaces vides entre les galaxies dans l'univers. Cette modélisation mathématique est basée sur les ensembles de Cantor.

On se place dans un cas le plus général possible. On considère un ensemble de quantités $\{\Lambda_k\}$ qui correspondent à k mesures d'instances distinctes d'un même phénomène (population d'une ville, fréquence d'un mot, taille des vides intergalactiques dans l'univers...). On définit $p(\Lambda)$ comme la probabilité d'occurrence de la valeur Λ de la quantité mesurée. Une autre manière de caractériser le phénomène est de déterminer sa distribution rang-fréquence, ou loi de Zipf. Les valeurs de Λ sont classées par ordre décroissant et à chaque valeur est associée son rang R dans la séquence. De ce fait, la fonction $\Lambda(R)$ donne une information équivalente à celle donnée par la distribution de probabilité $p(\Lambda)$. En particulier, si $\Lambda(R)$ est proportionnel à

R^{-z} , alors $p(\Lambda)$ est proportionnel à $\Lambda^{-\alpha}$ avec $\alpha=1+1/z$. En effet, $p(\Lambda)$ est la proportion d'éléments dont la valeur est Λ . De ce fait, le nombre total d'éléments de valeur supérieure à Λ est proportionnelle à la distribution accumulée de $p(\Lambda)$.

Si $p(\Lambda) \propto \Lambda^{-\alpha}$, alors $R(\Lambda) \propto \Lambda^{1-\alpha}$. On peut alors en déduire la relation entre z et α .

Pour illustrer la relation entre la loi de Zipf et la géométrie fractale, on considère des ensembles de Cantor définis dans un intervalle unitaire. Si l'on se restreint aux fractales déterministes, on peut facilement déterminer des relations qui permettent de lier les deux phénomènes. Un ensemble de Cantor déterministe peut être généré par une procédure itérative. Son générateur est caractérisé par trois paramètres. L'un de ces paramètres est le facteur d'échelle r , tel que $r < 1$. L'intervalle unité est partitionné en $1/r$ intervalles de largeur égale. On conserve un nombre N de ces intervalles, sur lesquels on répète la procédure et les $(1/r-N)$ intervalles restants sont éliminés. Les quantités N et r permettent de définir la dimension fractale D selon la formule définie par Mandelbrot [Man-83] :

$$D = -\log N / \log r \quad (4-4-1)$$

Les N intervalles pouvant être choisis de manière différente, en particulier ils peuvent être adjacents ou non. Cela peut être modélisé par un paramètre m qui représente le nombre de trous dans le générateur. Plus on a d'intervalles adjacents, moins on a de trous et plus la valeur de m est petite. Le paramètre m permet ainsi de définir une mesure de lacunarité, qui exprime la taille des vides dans l'ensemble considéré. On peut ainsi avoir plusieurs ensembles de Cantor de même dimension fractale qui diffèrent par la valeur de m . Les trois paramètres N , r et m définissent complètement le générateur d'un ensemble de Cantor. Par exemple l'ensemble de Cantor triadique classique est défini par $N=2$, $r=1/3$ et $m=1$. La largeur moyenne c des m trous initiaux, exprimée en unités de r , peut être obtenue à partir de la relation :

$$cm + N = 1/r \quad (4-4-2)$$

Après avoir défini l'ensemble de Cantor, on va maintenant calculer la loi de Zipf qui y est associée. Les trous sont rangés par ordre décroissant de largeur et à chaque trou est associé son rang R . Soit R_i le rang du trou le plus grand généré à l'itération i . A chaque itération, on génère donc les trous de rang R_i à $R_{i+1}-1$. A la première itération de la construction de l'ensemble de Cantor, on a m trous de rang variant de 0 à m et de largeur moyenne $\Lambda_1 = cr$. A la seconde itération, on aura mN trous de rang compris entre $1+m$ et $m+mN$ et de largeur moyenne $\Lambda_2 = cr^2$. Plus généralement, à l'itération i on génère mN^{i-1} trous dont la largeur moyenne est :

$$\Lambda_i = cr^i \quad (4-4-3)$$

Le rang des trous générés à l'itération i varie de

$$R_i = 1 + m \frac{N^{i-1} - 1}{N - 1} \quad (4-4-4)$$

$$\text{à } R_{i+1} - 1 = mN^{i-1} - 1 \quad (4-4-5)$$

On peut vérifier que les rangs du premier et du dernier trou de chaque classe ont la même relation d'échelle que les paramètres du système. A partir des relations (4-4-3) et (4-4-4), il est possible de calculer explicitement la loi de Zipf. En éliminant le paramètre i et en arrangeant différemment les termes, on obtient la relation suivante, qui est valable pour des grandes valeurs de R :

$$\Lambda(R) \approx f(r, N, m) R^{-1/D} \text{ avec } f(r, N, m) = \frac{1-rN}{m} \left(\frac{N-1}{m} \right)^{-1/D} \quad (4-4-6)$$

L'expression ci-dessus est vérifiée pour les grandes valeurs de R. La distribution des largeurs des trous N_r a été définie par Mandelbrot comme étant le nombre de trous de largeur supérieure à une échelle Λ . Selon Mandelbrot, cette distribution suit une loi puissance d'exposant $-D(N, \alpha \Lambda^{-D})$. On peut remarquer que le rang R_i est défini en additionnant le nombre total de trous de largeur supérieure ou égale à celle du $i^{\text{ème}}$ trou. De ce fait, la distribution de la largeur des trous correspond à $R(\Lambda)$ qui peut être définie par l'équation suivante :

$$R(\Lambda) = F(r, N, m) \Lambda^{-D} \quad (4-4-7)$$

où la valeur de $F(r, N, m)$ est définie par la formule :

$$F(r, N, m) = f(r, N, m)^D = \frac{m}{N-1} \left(\frac{1-rN}{m} \right)^D \quad (4-4-8)$$

$F(r, N, m)$ est ici une mesure de la lacunarité de l'ensemble fractal. F est proportionnel à m^{1-D} et croît avec m , du fait que plus F est petit, plus l'ensemble devient lacunaire. F varie de $F=(1-rN)^D/(N-1) < 1$ pour $m=1$ à $F=1$ pour $m=N-1$.

Un cas particulier est celui des ensembles fractals à trous de largeur minimale. Dans ce cas la valeur minimale de m est $m=N+1$ et $1/r=2N-1$. Dans ce cas, l'expression de Λ donnée par l'équation (12) est exacte pour toute valeur de R et la lacunarité F^{-1} est égale à 1 et donc minimale. Un exemple de ce cas particulier est l'ensemble de Cantor triadique.

Nous allons maintenant étudier le cas des ensembles de dimension supérieure à 1. En dimension 1, le nombre de trous est facile à définir car un point divise un segment en deux parties disjointes. Par contre, dans un espace de dimension supérieure à 1, les zones vides de l'espace sont souvent connectées entre elles et un trou ne peut pas être défini de façon unique. On peut cependant généraliser les résultats obtenus en dimension 1 si l'on considère que l'on ne peut fusionner ensemble que les trous de dimension équivalente, c'est à dire qui sont générés lors d'une même itération. Dans ce cas, on peut exprimer la valeur de Λ :

$$\Lambda(R) \approx \frac{1-r^{dN}}{m} \left(\frac{N-1}{m} \right)^{-d/D} R^{-d/D} \quad (4-4-9)$$

Un cas particulier est le générateur qui est obtenu en enlevant un certain nombre des $1/r^2$ parties d'un carré symétriquement par rapport à la diagonale. Ce résultat est simplement le produit cartésien de deux ensembles de Cantor unidimensionnels. Le complémentaire d'un tel ensemble est connexe, mais si l'on suppose que les trous obtenus à chaque itération sont indépendants, on vérifie la relation (4-4-9).

Si l'on définit les trous d'une manière différente, l'expression de $\Lambda(R)$ sera différente, et donc la lacunarité de l'ensemble fractal va changer. Cependant, la relation d'échelle entre la taille des vides et la loi de Zipf reste la même, car elle est indépendante de m et de c .

Les phénomènes naturels ne pouvant pas toujours être modélisés par des modèles déterministes, on est amené à considérer des modèles aléatoires. On considère un ensemble de Cantor bidimensionnel dans lequel les parties que l'on enlève à chaque itération de la construction sont choisies aléatoirement, les résultats obtenus en dimension 2 pouvant être extrapolés à des dimensions supérieures. On enlève à chaque itération i des régions du plan d'aire r^{2i} choisies aléatoirement. Lorsque l'on veut modéliser un phénomène naturel, comme la distribution des vides intergalactiques, la distribution spatiale est irrégulière et on ne peut plus la modéliser par un ensemble de Cantor. C'est pourquoi Gaite et Manriubia ont proposé un algorithme permettant de construire un ensemble fractal où les vides sont distribués irrégulièrement. On considère un plan dans lequel des points sont distribués irrégulièrement. On découpe le plan en cellules élémentaires comportant au plus un point, et on décide de couvrir les zones vides par des cercles ou des carrés, selon une méthode qui s'apparente à

celle du box-counting. Pour chaque cellule du plan, on construit le carré (ou le cercle) vide le plus grand possible centré sur cette cellule. Parmi les carrés ainsi construits, on choisit le plus grand, d'aire Λ_1 et on lui attribue le rang 1. On remplit toutes les cellules de ce carré et on recommence la procédure jusqu'à ce que l'on ait recouvert toutes les zones vides du plan. On obtient donc une suite de carrés vides classés par aire décroissante. La répartition de l'aire des zones vides suit une loi de Zipf d'exposant $-2/D$, où D représente la dimension fractale de l'ensemble des points.

Si en théorie la répartition des tailles des vides intergalactiques suit une loi puissance, ce n'est pas exactement le cas dans la pratique : la courbe représentant la taille des vides galactiques en fonction du rang, représentée dans un repère bi-logarithmique n'est pas linéaire, mais présente une forte incurvation vers le bas. D'après Gaite et Manrubia, ces résultats ne permettent pas de conclure quant à la validité du modèle, du fait de l'imprécision des données dont on dispose, il est en effet difficile de dénombrer les vides intergalactiques et d'évaluer précisément leur taille.

Une analogie peut être faite entre la répartition des tailles des vides galactiques et la fréquence des motifs d'une image. En effet, la fréquence d'un motif d'une image est la nombre total d'unités élémentaires de surface occupées par ce motif, cette quantité est donc homogène à une aire. On peut donc appliquer un raisonnement analogue à celui décrit précédemment à des zones homogènes d'une image. Cependant, à la différence du modèle décrit précédemment pour lequel les points de l'espace n'avaient que deux valeurs possibles, vide et non-vidé, on est ici amené à considérer la valeur du niveau de gris, ou de la classe à laquelle appartient le pixel de l'image. Sur le modèle du générateur de Cantor, cela revient à associer à chaque trou une valeur β qui correspond à la valeur du code du motif. La fréquence d'apparition du motif est alors la somme des largeurs des trous ayant la même valeur de β .

Cependant, si la valeur de β associée à chaque trou est choisie aléatoirement, il n'existe pas a priori de relation simple liant la somme des largeurs des trous ayant la même valeur de β à la dimension fractale de l'ensemble de Cantor. C'est pourquoi il n'est pas possible de montrer de manière exacte que la répartition des fréquences des motifs suivra une loi puissance. D'ailleurs, dans la pratique, la loi de distribution rang-fréquence des motifs que l'on observe sur des images réelles n'est en général pas linéaire dans le repère bi-logarithmique.

Conclusion

Les modèles de lois puissance comme la loi de Zipf et de Zipf inverse peuvent être appliqués à des images numériques. Pour cela, on définit des motifs qui sont les équivalents des mots dans le cas de l'analyse de texte. Les résultats obtenus en utilisant directement les niveaux de gris des pixels étant inexploitable du fait du trop grand nombre de motifs, il est nécessaire de définir d'autres codages qui permettent de réduire le nombre de motifs différents présents dans l'image, comme le codage des n classes ou le codage des rangs généraux. Suivant le codage utilisé, les propriétés de l'image mises en évidence seront différentes : le codage des n classes permet de caractériser les principales régions homogènes de l'image et le codage des rangs généraux met en évidence les contours et détails fins. L'application des modèles de lois puissance permet notamment de caractériser la complexité du contenu structurel de l'image et de distinguer entre les images naturelles ou artificielles. Cela permet d'envisager des applications de ces modèles telle que la mesure de la distorsion d'une image compressée ou encore la détection d'un objet artificiel ou d'une zone d'intérêt dans une image naturelle.

Il est à noter que la répartition des fréquences d'apparition des motifs de l'image ne suit pas exactement la loi de Zipf, en effet sa représentation graphique dans un repère bi-logarithmique n'est pas une droite. En revanche, la loi de Zipf inverse est vérifiée pour les motifs les moins fréquents de l'image.

La répartition des fréquences d'apparition modélisée par la loi de Zipf est invariante par une transformation isométrique du plan comme la translation ou la rotation, cependant, dans le cas de la rotation, cela ne se vérifie dans la pratique que pour des rotations d'un multiple de 90° du fait des déformations de l'image introduites lors d'une rotation par un angle quelconque dans le cas des images discrètes.

Il est possible de définir un lien entre la géométrie fractale et la loi de Zipf, cependant ce modèle fractal est difficilement applicable au cas des images en niveaux de gris.

Maintenant que nous avons vu que l'on peut appliquer les modèles de lois puissance pour analyser des images, nous allons les utiliser dans des applications pratiques, à savoir l'évaluation de la qualité de compression des images et la détection d'objets et de zones d'intérêt.

Chapitre 3 : Evaluation de la qualité de la compression des images

Introduction

Après avoir défini les modèles de loi puissance utilisables aux images, nous allons maintenant les utiliser dans le cadre d'une application concrète, celle de l'évaluation de la qualité des images compressées. La compression d'une image à l'aide méthode de compression avec pertes comme JPEG ou JPEG2000 introduit une dégradation de l'image qui se traduit par des modifications des détails de celle-ci. La répartition des fréquences d'apparition des motifs modélisée par la loi de Zipf doit permettre de caractériser cette dégradation, ce qui fait que l'on peut définir des critères de mesure de la dégradation basés sur la loi de Zipf.

Dans ce chapitre, après avoir présenté les principes généraux de la compression d'images, nous allons passer en revue les différents critères de mesure de la dégradation d'une image qui existent dans la littérature. Ensuite, nous définirons différents critères de mesure de la qualité de l'image compressée basés sur la loi de Zipf et adaptés à différents types de compression. Enfin, nous montrerons comment utiliser ces critères pour optimiser de façon automatique la compression d'une image, c'est à dire pour obtenir une image de la meilleure qualité possible pour un fichier le plus petit possible.

1 Généralités sur la compression d'images

1.1 Nécessité de la compression des images

Les images numériques comptent parmi les données les plus volumineuses qui soient. Si l'on code une image numérique sous forme d'une matrice de pixels, comme c'est le cas pour un fichier de type « bitmap », dans le cas d'une image codée en 256 niveaux de gris chaque pixel de l'image est codé par un octet qui représente le niveau de gris de l'image, et dans le cas d'une image couleur, par trois octets qui représentent les trois composantes de l'espace couleur utilisé (RGB, TSL, HSV, YUV, YCbCr...). Or, une image numérique contient une densité d'information considérable, nettement plus importante que celle contenue par exemple dans un fichier texte. En effet, un pixel d'une image occupe un espace mémoire équivalent à celui d'un caractère dans le cas d'une image en niveaux de gris, et de trois caractères dans le cas d'une image couleur.

Pour bien se rendre compte du volume de données que représente une image même petite, prenons un exemple. Si l'on considère une image en niveaux de gris de taille 320x200, cette image contient 64000 pixels. Un tirage papier de cette image à une résolution de 150 dpi (résolution minimale pour obtenir un rendu visuellement acceptable) aura pour dimensions 5,42 x 3,38 cm, soit une aire de 18,32cm². Chaque pixel de cette image occupera une surface de 0,003 mm², à comparer avec les 5 à 10 mm² qu'occupe en moyenne un caractère d'imprimerie. Si l'on considère qu'une page de texte comporte environ 3000 caractères, notre petite image de faible résolution occupera un espace mémoire équivalent à plus de 21 pages de texte si l'image est en niveaux de gris, et à 64 pages si l'image est en couleur.

On remarque bien ici la quantité énorme de données que représentent les images numériques, ce qui fait que le coût du stockage des images sur disque dur, et a fortiori de leur transmission

sur Internet devient très rapidement prohibitif. C'est pourquoi l'on utilise des méthodes de compression afin de réduire le volume des données à stocker ou à transmettre.

1.2 La compression sans perte

La compression sans pertes est basée sur la théorie de l'information développée par Shannon [Sha-48]. Le principe de cette compression est d'éliminer la redondance présente dans les données à transmettre afin de maximiser la quantité moyenne d'information par bit de données transmis, sans perdre aucune information par rapport aux données originales. Plusieurs algorithmes de compression sans pertes ont été mis au point, ces algorithmes ne sont pas spécifiques aux images et peuvent être employés pour tout type de données. Nous en présentons ici quelques uns.

Le **codage de Huffman** [Huf-52] consiste à coder les symboles à transmettre, par exemple les octets d'un fichier, par des mots binaires de longueur variable suivant la probabilité d'apparition de chacun des symboles, les symboles les plus fréquents seront codés par les mots les plus courts. On code les symboles en utilisant un arbre binaire. Il a été démontré que le codage de Huffman est optimal en ce sens qu'il minimise la taille du code par longueur entière de bits.

Le **codage arithmétique** consiste à coder une séquence de symboles par un nombre à virgule flottante compris entre 0 et 1. Le principe de l'algorithme est d'associer chaque symbole du fichier à un intervalle en fonction de sa probabilité d'apparition. On commence par initialiser l'intervalle de travail à $[0,1]$, le premier symbole est représenté par la plage qui lui est affectée. Ensuite, chaque symbole se restreint lui-même de la façon suivante : la borne inférieure est augmentée du produit de la largeur de l'intervalle précédent par la borne inférieure de l'intervalle du symbole lu, et la borne supérieure est diminuée du produit de la largeur de l'intervalle précédent par la borne supérieure de l'intervalle du symbole lu. On choisit ensuite un nombre dans l'intervalle final pour obtenir le code du fichier. Cette méthode est plus lente que le codage de Huffman mais permet un meilleur taux de compression.

Le codage RLE (**Run-Length Encoding**) consiste à coder les séquences de symboles identiques consécutifs par le nombre de symboles dans la séquence. Ce codage est utilisé notamment dans les formats d'image GIF et PCX, il est particulièrement bien adapté pour coder des images qui comportent de grandes zones parfaitement uniformes, comme des dessins ou des schémas. En revanche, cette méthode est nettement moins efficace dans le cas d'images photographiques qui ne contiennent pas de telles zones uniformes.

Le **codage de Ziv et Lempel** [Ziv-77], [Ziv-78] est un codage par dictionnaire. Le dictionnaire est construit de manière dynamique pendant la lecture du fichier. Lors de la compression, on remplace chaque nouvelle occurrence d'un symbole par un pointeur vers la position du même symbole dans le dictionnaire. Il existe plusieurs versions de cet algorithme, notamment l'algorithme LZ77 et l'algorithme LZW.

Les différents algorithmes de compression sans pertes permettent de compresser un fichier sans aucune perte d'information, ils sont efficaces sur certains types de données comme les textes mais ils ne permettent qu'un taux de compression relativement faible dans le cas des images, en effet l'efficacité de la compression dépend de la redondance présente dans les données à coder, or les images ont en général une redondance très faible. Ceci justifie le développement d'autres méthodes de compression.

1.3 La compression avec perte

L'efficacité limitée de la compression sans pertes dans le cas des images a conduit à développer des algorithmes de compression avec perte qui permettent des taux de compression beaucoup plus élevés, mais au prix d'une perte d'information. En effet, l'œil humain est incapable de distinguer facilement toutes les nuances d'une image. Ainsi, dans le cas des images, une distorsion par rapport à l'original peut être admise alors que ce ne serait pas acceptable par exemple dans un texte ou dans un fichier exécutable. Différentes méthodes de compression avec pertes, ou compression destructrice, spécifiques aux images ont été proposées, elles utilisent des transformations du signal de l'image dans le domaine fréquentiel qui permettent de représenter ce signal sous une forme pouvant être codée de manière plus compacte, mais avec une précision moindre qui fait perdre de l'information. Les méthodes de compression d'images avec pertes comprennent notamment JPEG, SPIHT et JPEG2000.

Le **codage JPEG** est basé sur la transformée en cosinus discret (DCT). Il a été développé par le Joint Photographic Experts Group [Wal-91], [Pen-93] afin de normaliser la compression d'image. La norme JPEG est le standard de codage des images de loin le plus employé à l'heure actuelle, notamment sur Internet. Le principe de la compression JPEG est de découper l'image en blocs de 8x8 pixels et de remplacer chaque bloc par une combinaison linéaire de 64 motifs de base prédéfinis représentés sur la figure ci-dessous

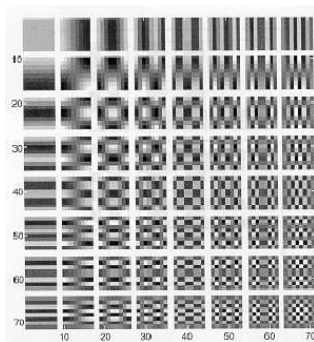


Fig.1 : Motifs de base du codage JPEG

Chaque bloc est représenté par une matrice 8x8 dont les coefficients représentent la pondération de chaque motif de base. La transformation JPEG peut s'appliquer sur des images en niveaux de gris ou des images couleur. Dans le dernier cas, on effectue un changement d'espace couleur pour représenter l'image dans l'espace YCbCr.

Les coefficients de la matrice sont obtenus par la transformation en cosinus discret, chaque coefficient d'indice (u,v) étant donné par la formule suivante :

$$DCT(u,v) = \sum_{x=0}^7 \sum_{y=0}^7 g(x,y) f(x,y,u,v) \quad (1-3-1)$$

Dans cette formule, g(x,y) représente la valeur du pixel situé à la position (x,y) du bloc et f(x,y,u,v) le coefficient de la transformée en cosinus discret a la forme suivante :

$$f(x,y,u,v) = C(u)C(v) \cos\left(\frac{(2x-1)u\pi}{16}\right) \cos\left(\frac{(2y+1)v\pi}{16}\right) \quad (1-3-2)$$

$$\text{avec } C(u) = \begin{cases} \frac{1}{2\sqrt{2}} & \text{si } u=0 \\ \frac{1}{2} & \text{sinon} \end{cases} \quad (1-3-3)$$

Le calcul est suivi d'une phase de quantification qui consiste à diminuer la précision de la représentation des coefficients de la DCT afin de réduire le nombre de bits nécessaires à leur stockage. C'est à ce niveau-là que l'on perd de l'information. La quantification consiste à diviser terme à terme les coefficients de la matrice DCT par ceux d'une matrice de quantification Q. Le résultat de la division est arrondi à l'entier inférieur. Les coefficients ayant une valeur inférieure à 1 seront donc arrondis à 0 et l'information qu'ils représentent sera perdue. On voit ici que l'on privilégie les coefficients de basse fréquence, c'est à dire les motifs les plus uniformes, et que l'on atténue plus fortement les coefficients qui correspondent aux détails fins de l'image. C'est le choix de la matrice de quantification qui permet de paramétrer la compression de l'image. Les coefficients $q(u,v)$ de la matrice de quantification Q sont calculés en fonction d'un facteur de qualité Q selon la formule suivante :

$$q(u,v) = \frac{1}{1 + K(1 + u + v)} \quad (1-3-4)$$

Selon la valeur de K, on perdra plus ou moins d'information et l'image compressée sera d'une qualité plus ou moins grande.

Afin de réduire la taille du fichier, on applique aux données issues de la quantification une compression conservatrice. Comme après la quantification les coefficients de plus haute fréquence sont le plus souvent nuls, on parcourt la matrice quantifiée en zigzag afin de regrouper ensemble les zéros, comme le montre la figure ci-dessous.

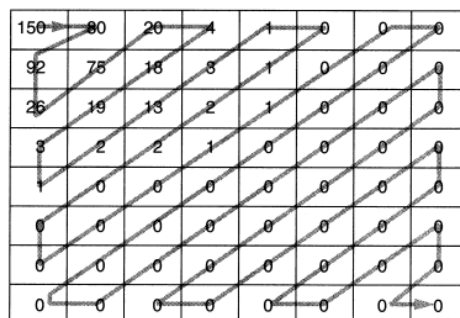


Fig.2 : Séquence de parcours de la matrice quantifiée

On utilise un codage RLE pour coder les zéros consécutifs de la matrice, et on compresse ensuite la séquence entière en utilisant un codage de Huffman ou un codage arithmétique.

La compression JPEG permet de réduire sensiblement la taille du fichier image, mais introduit une distorsion du fait de la décomposition en blocs et de la quantification, si le taux de compression est important, les discontinuités entre les blocs deviennent visibles et l'on observe un effet de mosaïque qui correspond aux motifs de base de basse fréquence.

La **compression SPIHT** (Set Partitioning In Hierarchical Trees), mise au point par Said et Pearlman [Sai-96] est une méthode de compression d'images basée sur une transformée en ondelettes et un partitionnement en arbres hiérarchiques. Le principe de base de la méthode est de classer l'information afin de séparer l'information importante de l'information non

importante. Dans un premier temps, on applique une décomposition en ondelettes sur l'image. Les coefficients $c(i,j)$ de la transformée en ondelettes sont ensuite représentés sous forme d'un arbre hiérarchique. Cet arbre est ensuite divisé en sous-arbres. Ces sous-arbres sont organisés de manière à regrouper ensemble les coefficients les plus importants, c'est à dire ceux dont l'amplitude est la plus importante.

Les coefficients des ondelettes sont donc rangés dans un arbre en commençant par la bande de résolution la plus basse. Pour les bandes de résolution plus élevées, on code uniquement les coefficients importants, c'est à dire ceux qui sont supérieurs à un seuil. La valeur initiale du seuil est $n = \log_2 c_{\max}$ où c_{\max} représente ici l'amplitude maximale des coefficients d'ondelettes à encoder. On construit alors une liste de bits significatifs et une liste de bits insignifiants. A chaque passe on encode les bits significatifs. Les bits insignifiants deviennent les bits de départ de la passe suivante sur lesquels un nouveau calcul de seuil est appliqué. On effectue ainsi un certain nombre de passes, plus le nombre de passes est important et plus la précision du codage sera grande. Un algorithme de tri identique pour le codeur et pour le décodeur permet de retrouver l'ordre des coefficients. Un des avantages de cette méthode est l'utilisation d'un codage encastré pour les fichiers compressés. Cela signifie que même si l'on ne reçoit qu'une petite partie du fichier compressé, on peut reconstituer l'image entière, avec cependant une qualité plus faible. Ce type de codage permet d'accroître la robustesse aux erreurs de transfert car même si la connexion est coupée, quelle que soit la taille du fichier reçu, on pourra reconstituer l'image. Cela fonctionne même en cas d'arrêt du codeur en plein milieu de la compression.

La **compression JPEG2000** [ISO-00] est un nouveau standard de compression à base d'ondelettes qui permet, par rapport à une compression JPEG, d'obtenir des images de meilleure qualité à taille de fichier égale. La norme JPEG2000 permet de réaliser une compression avec ou sans pertes, le codage de l'image est conçu de manière à faciliter une transmission rapide de l'information sur un réseau et à permettre un décodage progressif de l'image. La norme JPEG2000 offre également la possibilité de définir une *région d'intérêt* qui sera moins compressée que le reste de l'image et aura donc une meilleure qualité visuelle. Pour une présentation plus détaillée de la norme JPEG2000, on se reportera à l'ouvrage de Taubman et Marcellin [Tau-01]. La compression JPEG2000 utilise les ondelettes de Daubechies [Dau-92]. Ces ondelettes sont définies par les deux équations suivantes :

$$\varphi(t) = \sum_{k=-\infty}^{+\infty} h_k \varphi(2t-k) \quad (1-3-5)$$

$$w(t) = \sum_{k=-\infty}^{+\infty} (-1)^k h_{n-k} \varphi(2t-k) \quad (1-3-6)$$

La suite $(h_k)_k$ possède un nombre fini de valeurs différentes de 0. Pour déterminer une ondelette de Daubechies, il suffit de déterminer les coefficients h_k puis la fonction φ . Suivant le nombre de coefficients h_k non nul, on obtient une ondelette différente. On notera db_k l'ondelette de Daubechies ayant k coefficients. La méthode pour le calcul des coefficients h_k pour chaque ondelette db_k a été décrite par Meyer [Mey-93]. On se reportera à cet ouvrage pour le détail des calculs. Dans le cas de la compression JPEG2000, on utilise la paire (9,7) irréversible de Daubechies, qui définit un filtre passe-bas à 9 coefficients et un filtre passe-haut à 7 coefficients, tous deux à coefficients irrationnels. Cette transformation en ondelettes est irréversible et ne permet donc pas de compresser sans perte. Pour réaliser une compression sans pertes, il faut utiliser une autre base d'ondelettes, à savoir la paire (5,3) réversible de Le Gall qui est à coefficients rationnels, mais qui ne permet pas d'obtenir des taux de compression aussi élevés que la précédente.

L'image est décomposée en différentes sous-bandes de fréquence horizontales et verticales à l'aide de l'algorithme de Mallat. Chacune des sous-bandes représente les caractéristiques horizontales et verticales de l'image à différents niveaux de détail. La norme JPEG2000 n'impose pas un nombre précis de niveaux de décomposition de l'image. Avant d'appliquer cette décomposition, on effectue un décalage de niveau qui consiste à soustraire un même nombre à chacune des composantes couleurs de l'image.

Après l'application de la transformée en ondelettes, on effectue une quantification scalaire uniforme sur la matrice des coefficients d'ondelettes afin de réduire la précision de la représentation, ceci dans le but de réduire la taille des données à transmettre. On élimine ainsi les coefficients de faible amplitude, qui comportent peu d'information et ont donc une influence faible sur la qualité finale de l'image. La quantification fixe le taux de compression minimal de l'image, c'est à dire quand toute l'information est transmise. On définit un pas de quantification Δ_b pour chaque sous-bande de fréquence b , chacune des sous-bandes pouvant avoir (ou non) un pas de quantification différent. Le coefficient de la matrice quantifiée $q_b(u,v)$ de fréquence b et d'indice (u,v) est obtenu à partir du coefficient d'ondelette $a_b(u,v)$ par la formule :

$$q_b(u,v) = \text{sgn}(a_b(u,v)) \left(\frac{a_b(u,v)}{\Delta_b} \right) \quad (1-3-7)$$

Après quantification, chacune des sous-bandes de la décomposition en ondelettes est divisée en sous-blocs (*code blocks*). Chaque *code-block* est codé indépendamment en utilisant un codage arithmétique adaptatif avec contexte. Chaque *code-block* est codé plan de bits par plan de bits, en commençant par les bits de poids le plus fort. Dans chaque plan de bits, les bits sont séparés en trois groupes, en fonction de leur voisinage, puis codés en trois passes successives. Afin de réduire le volume de données à transmettre et/ou de permettre un décodage progressif de l'image, le *bit-stream* obtenu après codage peut être tronqué à différents niveaux qui correspondent à différents niveaux de qualité d'image.

La phase d'allocation de débit consiste à organiser les données en paquets afin de les stocker et/ou de les transmettre. Chaque paquet correspond à un *layer* d'un niveau de résolution donné d'une composante de l'image. Il est constitué d'un en-tête identifiant son contenu et permettant un accès rapide dans le *code-stream*, ainsi que les données compressées obtenues par concaténation d'un certain nombre de passes de codage des *code-blocks* d'un même niveau de résolution. Afin d'augmenter le taux de compression, les dernières passes d'un *code-block* sont souvent sautées, ce qui a pour conséquence de diminuer la précision des coefficients dont les bits de poids faibles ont été éliminés.

Chaque paquet est ajouté au *code-stream* à la suite d'un en-tête regroupant tous les paramètres de codage et suivant un ordre d'inclusion dépendant du type de progression souhaité. Il est possible de modifier la progression au moment de la transmission à partir d'un même *code-stream* côté serveur. Ainsi, suivant l'ordre dans lequel le décodeur reçoit les paquets, il est capable de reconstruire une image progressivement par résolution ou par *layers*. Le codage progressif offre pour avantage de pouvoir décoder partiellement l'image, cela est particulièrement utile dans le cas où le canal de transmission a un faible débit, l'utilisateur peut ainsi choisir entre une image de faible qualité décodée plus rapidement ou une image de plus grande qualité décodée plus lentement.

Afin de réduire l'espace mémoire nécessaire au codage, il est possible de découper l'image en imagettes (*tiles*) et d'appliquer séparément la transformée en ondelettes sur chacune des imagettes. Cependant, on introduit une distorsion supplémentaire liée aux discontinuités entre les bords des imagettes.

Une fonctionnalité particulièrement intéressante de la norme JPEG 2000 est la possibilité de définir une zone d'intérêt qui sera moins compressée que le reste de l'image. Le principe est

d'encoder en priorité dans le *code-stream* toutes les données qui correspondent à cette zone d'intérêt, elle sera alors décompressée en premier par rapport au reste de l'image. On traite ensuite séparément la zone d'intérêt et le reste de l'image, l'opération revenant à utiliser des pas de quantification différents pour les deux régions.

Par rapport à la norme JPEG, JPEG2000 permet d'obtenir des images de meilleure qualité visuelle à taille de fichier égale, ou des fichiers plus petits pour une qualité visuelle comparable. La dégradation causée par les deux types de compression n'est pas de même nature, la compression JPEG2000 ayant tendance à rendre l'image floue.

2 Evaluation de la qualité de la compression

La compression d'images avec perte, quelle que soit sa nature, par exemple de type JPEG, SPIHT ou JPEG2000 introduit une perte de qualité de l'image. Cette perte de qualité est plus ou moins visible selon les paramètres de la compression utilisés. La manière la plus évidente d'évaluer cette perte de qualité est de faire appel à un ou plusieurs observateurs humains pour examiner les images afin d'en évaluer la qualité. Mais cette manière de procéder donne des résultats subjectifs et souvent imprécis, instables au cours du temps, l'observateur finissant par se fatiguer. Or, on aimerait pouvoir obtenir une mesure chiffrée et objective de la qualité de la compression, notamment afin de comparer entre eux différents algorithmes de compression. C'est pourquoi diverses méthodes ont été proposées pour mesurer automatiquement la qualité d'une image compressée. Ces méthodes sont de plusieurs types. Elles se basent néanmoins sur les approches d'évaluation humaines que nous étudierons dans un premier temps. Les méthodes numériques consistent à attribuer à chaque image une mesure de qualité sous forme d'un nombre. Les méthodes perceptuelles sont fondées sur une modélisation du système visuel humain mais elles aboutissent à un nombre. Les méthodes graphiques représentent des caractéristiques sous forme d'un graphe ou d'une image reconstituée. Elles ne répondraient donc pas à l'objectif fixé.

2.1 L'évaluation subjective humaine

Une image est créée par un auteur qui a pour objectif de transmettre une information. Chaque observateur, en fonction de sa culture ne verra pas le même intérêt dans l'image. Ainsi, la notion même de qualité d'image est liée à la perception humaine, et les images étant en principe destinées à être regardées par un être humain, la meilleure référence dans ce domaine est celle donnée par l'évaluation humaine. Cette évaluation étant subjective par nature, afin d'obtenir des résultats fiables il est nécessaire de faire appel à plusieurs observateurs et de définir un protocole expérimental bien précis de manière à ce que les résultats soient comparables.

La composition du groupe d'observateurs doit être convenablement choisie. Idéalement, la constitution du groupe d'observateurs devrait comprendre à la fois des non-experts et des experts. Les non-experts sont des individus n'ayant pas de connaissances techniques dans le domaine de l'analyse d'image, ils représentent l'observateur lambda, tandis que les experts ont une connaissance du domaine auquel se rattache l'image ou de la compression d'images et savent les points particuliers à vérifier et les dégradations attendues en fonction de la méthode de compression.

L'évaluation de la qualité de l'image peut être absolue ou comparative. L'évaluation absolue consiste pour l'observateur à attribuer à chaque image une catégorie en fonction d'une échelle prédéfinie. L'évaluation comparative consiste à comparer plusieurs images et à les classer par ordre de qualité. Différentes procédures d'évaluation et des échelles de qualité ont été définies, notamment par Jain [Jai-89], Mannos et Sakrison [Man-74] et C.F. Hall [Hal-81]. Différentes échelles de qualité utilisables ont été listées par Eskicioglu [Esk-00] et sont résumées dans le tableau suivant. Nous avons traduit les termes utilisés dans [Esk-00] en essayant de se rapprocher le plus possible de l'original

Echelle MOS :	10,9 Très bon	
5 Excellent	8,7 Bon	
4 Bon	6,5,4 Correct	
3 Correct	3,2 Mauvais	
2 Mauvais	1,0 Très mauvais	
1 Insuffisant		
(échelle absolue de qualité)	(échelle absolue de qualité)	
1 Non remarquable	7 Le meilleur	1 Beaucoup mieux
2 A peine remarquable	6 Bien au dessus de la moyenne	2 Mieux
3 Définitivement remarquable	5 Légèrement au dessus de la moyenne	1 Légèrement mieux
4 Défaut non contestable	4 Moyen	0 Pareil
5 Quelque peu contestable	3 Légèrement au dessous de la moyenne	-1 Légèrement pire
6 Définitivement contestable	2 Bien au dessous de la moyenne	-2 Pire
7 Extrêmement contestable	1 Le pire	-3 Plus pire
(échelle absolue de dégradation)	(échelle absolue de qualité)	(échelle relative de différence de qualité entre deux images)

Tableau 1 : Différentes échelles de mesure subjective de la qualité d'une image. Adapté de [Esk-00].

Les résultats obtenus par une évaluation humaine sont variables suivant les observateurs et ne sont pas toujours reproductibles d'un groupe d'observateurs à un autre. De nombreux facteurs peuvent influencer les résultats, notamment le type ou la taille de l'image, l'origine sociale ou culturelle, le niveau d'études et la motivation des observateurs, les centres d'intérêt des personnes, les conditions expérimentales comme l'éclairage ou la qualité des images, le comportement de l'expérimentateur, etc.

Tous ces facteurs qui caractérisent le comportement d'un être humain face à une image, s'ils sont intéressants à étudier par eux-mêmes pour les psychologues et les sociologues, ne facilitent certes pas l'utilisation d'une évaluation subjective humaine pour évaluer avec précision les performances d'un algorithme de compression d'image.

Le plus utilisé de ces critères subjectifs est le « Mean Opinion Score » (MOS) défini en 1982 par le CCIR [CCI-82]. Le MOS définit une échelle de 5 niveaux de dégradation. Ce critère est initialement destiné à la télévision, depuis lors son utilisation s'est étendue à l'image numérique fixe ou mobile. L'utilisation du MOS consiste à faire attribuer par chacun des observateurs une note de 1 à 5 et le résultat final est la moyenne des résultats donnés par les

observateurs. Les conditions d'observation doivent obligatoirement être précisées, notamment le nombre d'observateurs, l'environnement d'observation, le temps d'observation, la distance entre l'observateur et l'image et les dimensions de l'image.

Les critères subjectifs donnent des résultats variables suivant les observateurs et les conditions d'expérimentation, de plus ils nécessitent la présence de plusieurs observateurs humains, ce qui les rend particulièrement lourds et difficiles à mettre en oeuvre dans la pratique et nécessite une organisation considérable. Par leur nature même, ils ne sont évidemment pas automatisables mais, par apprentissage et sur des bases d'images comparables, on peut construire des réseaux de neurones fournissant des indices de qualité. Leur utilisation est limitée à la famille d'images considérée. C'est pourquoi ont été développées des méthodes permettant d'obtenir par des moyens informatiques une évaluation objective unique de la qualité d'une image.

2.2 Les critères numériques univariés

Les critères numériques donnent une mesure de la qualité de l'image sous forme d'une unique valeur scalaire associée à l'image. Ils fournissent donc une évaluation purement quantitative de la dégradation de l'image, sans donner d'information sur le type de distorsion ou sur la localisation spatiale des défauts de l'image. Selon W.K Pratt [Pra-78], ces critères numériques de qualité de l'image peuvent être classés en deux grandes catégories : les critères univariés et les critères bivariés. Les critères univariés estiment la qualité de l'image dégradée dans l'absolu, sans utiliser l'image originale, tandis que les critères bivariés comparent l'image dégradée avec l'image originale qui est utilisée comme référence, par conséquent ils ont besoin des deux images. Nous allons tout d'abord nous intéresser aux critères univariés.

D'une manière générale, un critère univarié U est calculé à partir d'une image $F(a(j,k))_{j,k}$ de dimensions $M \times N$ en utilisant une formule du type :

$$U = \sum_{j=1}^M \sum_{k=1}^N O[a(j,k)] \quad (2-2-1)$$

où O représente un opérateur. Les arguments de cet opérateur peuvent ne pas se limiter à la seule valeur du niveau de gris en un point. Les critères univariés sont assez rarement employés pour évaluer la qualité des algorithmes de compression d'image. En effet, la dégradation de la qualité d'une image compressée s'exprime par la dégradation par rapport à l'image originale, il est donc difficile de mesurer cette dégradation si l'on ne dispose pas de l'image de référence. En règle générale, ce type de critère ne peut être utilisé que si l'on connaît à l'avance le type de distorsion que l'image est amenée à subir.

L'IQM (Image Quality Measure) proposé par Nill et Bouzas [Nil-92] est un critère univarié basé sur le spectre de puissance bidimensionnel de l'image. Cette méthode a été développée pour mesurer la qualité des images aériennes. Le spectre de puissance est défini comme le carré de l'amplitude de la transformée de Fourier de l'image, il donne des informations sur le contraste et le niveau de détails de l'image. L'utilisation du spectre de puissance d'une image pour en estimer la qualité est basée sur l'hypothèse que la forme de l'équation de ce spectre est invariante d'une image à l'autre. Bien que toutes les scènes soient différentes dans le domaine spatial, quand ces scènes sont analysées dans le domaine fréquentiel, sans tenir compte de la phase, on peut montrer que toutes les scènes ont en fait théoriquement le même spectre de puissance. Plus spécifiquement, la puissance à une fréquence non nulle f est donnée par une équation de la forme $P=1/f^k$, avec $k=2$ dans le cas d'une analyse

unidimensionnelle. Un spectre de cette forme a été observé dans la plupart des scènes naturelles, ce fait est probablement lié à la notion de dimension fractale.

Une scène naturelle peut être modélisée par un processus de Poisson doublement stochastique, c'est à dire un processus où l'on a non seulement une variation stochastique des changements d'état du processus, mais aussi de l'amplitude de ces changements. On considère que la luminosité d'un point donné de la scène $I(x,y)$ est une variable aléatoire gaussienne de moyenne \bar{I} et de variance σ_s^2 . Le fait qu'il existe ou non un changement de luminosité en un point peut être modélisé par une variable aléatoire à deux états dont la loi de probabilité est une loi de Poisson de la forme :

$$\frac{(\alpha\Delta x)^\lambda}{\lambda!} \exp(-\alpha\Delta x) \quad (2-2-2)$$

où Δx est l'intervalle de distance sur lequel on a une probabilité λ de changement de luminosité et où α est la réciproque de la largeur moyenne de l'impulsion. La transformée de Fourier bidimensionnelle F de la fonction d'autocorrélation de ce processus, pour une fréquence spatiale radiale ρ peut être donnée par la formule :

$$F = \frac{2\pi\alpha\sigma_s^2}{(\alpha^2 + \rho^2)^{3/2}} + K\bar{I}^2\delta(\rho) \quad (2-2-3)$$

où K est une constante. Dans le cas d'une image en niveau de gris $a(x,y)$, de taille $M \times M$ pixels, la transformée de Fourier discrète bidimensionnelle de l'image est donnée par la formule :

$$H(u,v) = \sum_{x=0}^{M-1} \sum_{y=0}^{M-1} \exp(-2\pi iy \frac{v}{M}) \exp(-2\pi ix \frac{u}{M}) a(x,y) \quad \text{avec } (u,v) = [-M/2 \dots M/2]^2 \quad (2-2-4)$$

Pour que les variables u et v de la transformée de Fourier correspondent aux composantes de fréquence spatiale de l'image d'entrée, u et v doivent être normalisés par le nombre de pixels dans les direction x et y soit u/M et v/M cycles par largeur de pixel.

Le spectre de puissance est défini par $|H(u,v)|^2$. Comme l'image d'entrée est une fonction asymétrique réelle, sa transformée de Fourier est hermitienne et son spectre de puissance est bilatéralement symétrique. De ce fait, on peut ignorer les valeurs négatives de u et de v . Le spectre est normalisé par rapport à la luminosité moyenne μ et le nombre de pixels M^2 de l'image. On obtient le spectre de puissance normalisé :

$$P(u,v) = \frac{|H(u,v)|^2}{\mu^2 M^2} \quad (2-2-5)$$

Comme plusieurs des quantités qui seront calculées sont symétriques par rotation, on convertira le spectre en coordonnées polaires et on utilisera :

$$P(\rho,\theta) = \frac{|H(\rho,\theta)|^2}{\mu^2 M^2} \quad (2-2-6)$$

La sensibilité du système visuel humain à la fréquence spatiale est modélisée par une fonction de transfert de modulation $A(Tp)$ qui dépend du module de la fréquence spatiale. Afin de compenser la sensibilité au bruit du spectre de puissance, un filtre de Wiener de réponse $W(\rho)$ est appliqué sur le spectre au-dessus d'un certain niveau de bruit dans l'image. Le spectre peut être normalisé par rapport à la taille de l'image et dans le cas d'images aériennes en vue oblique, ce spectre peut être pondéré par un facteur d'échelle $S(\theta_1)$ dépendant de l'angle de prise de vue. On calcule un unique paramètre de qualité IQM qui est la somme des valeurs pondérées du spectre de puissance. Le spectre de puissance contenant des informations sur la netteté, le contraste et le niveau de détails de l'image,

caractérise la qualité de celle-ci. Le facteur de qualité IQM est la somme des valeurs pondérées du spectre de puissance, il est calculé selon la formule suivante :

$$IQM = \frac{1}{M^2} \sum_{\theta=-180^{\circ}}^{+180^{\circ}} \sum_{\rho=0,01}^{0,5} S(\theta)W(\rho)A^2(T\rho)P(p,\theta) \quad (2-2-7)$$

La pondération visuelle aligne le facteur IQM par rapport à la qualité perçue par un observateur humain. Le filtrage permet de compenser la sensibilité au bruit du spectre de puissance. La normalisation par rapport à la luminosité et à la taille de l'image permet des comparaisons entre des images de taille et de luminosité moyenne différentes. Le facteur de qualité IQM peut être converti en une valeur du critère NIIRS (National Imagery Interpretability Rating Scale). Le NIIRS est une échelle de qualité visuelle utilisée pour les images aériennes, ce critère est dépendant de l'échelle de l'image. Pour convertir le critère IQM qui est indépendant de l'échelle en NIIRS, on utilise un facteur d'échelle et une courbe de régression NIIRS-IQM déterminée préalablement.

Le critère proposé par Z. Wang et al. [Wan-02a] est destiné à mesurer de manière objective la qualité d'une image compressée en JPEG. Avec ce type de compression, la perte de qualité a deux origines principales. Elle vient d'une part de l'introduction des blocs 8x8 sur lesquels est calculée la transformation DCT et d'autre part du flou qui provient de la non prise en compte des coefficients de haute fréquence. La méthode proposée permet d'estimer les deux types de dégradation à partir des différences entre les blocs de l'image. Contrairement au paragraphe précédent où l'on avait utilisé une image carrée, on considère ici une image rectangulaire $a(m,n)$ de taille $M \times N$. On calcule un signal de différentiation le long de chaque ligne horizontale de l'image, en utilisant la formule :

$$d_h(m,n) = a(m,n+1) - a(m,n) \quad , \quad n \in [0, N-1] \quad (2-2-8)$$

L'influence des blocs est évaluée en mesurant la différence moyenne entre les blocs

$$B_h = \frac{1}{M([N/8]-1)} \sum_{i=1}^{M} \sum_{j=1}^{[N/8]-1} |d_h(i,8j)| \quad (2-2-9)$$

Le flou est évalué en mesurant l'activité du signal, à l'aide de deux paramètres. Le premier paramètre A_h est la moyenne des différences absolues du signal entre les échantillons d'image à l'intérieur d'un bloc, calculé selon la formule :

$$A_h = \frac{1}{7} \left[\frac{8}{M(N-1)} \sum_{i=1}^M \sum_{j=1}^{N-1} |d_h(i,j)| - B_h \right] \quad (2-2-10)$$

Le second paramètre d'activité Z_h est le taux de passage à zéro du signal $d_h(m,n)$ qui est défini par la formule :

$$Z_h = \frac{1}{M(N-2)} \sum_{i=1}^M \sum_{j=1}^{N-2} z_h(m,n) \quad (2-2-11)$$

où $z_h=1$ si le signal $d_h(m,n)$ passe par 0 au point de coordonnées (m,n) et $z_h=0$ dans le cas contraire. De la même manière, on calcule les paramètres verticaux A_v, B_v et Z_v à partir de la différence d_v entre les lignes verticales. On obtient alors les paramètres de qualité A, B et Z qui sont définis par

$$A = \frac{A_h + A_v}{2}, \quad B = \frac{B_h + B_v}{2} \quad \text{et} \quad Z = \frac{Z_h + Z_v}{2} \quad (2-2-12)$$

Le critère de qualité S est défini par la formule :

$$S = \alpha + \beta A^{\gamma_1} B^{\gamma_2} Z^{\gamma_3} \quad (2-2-13)$$

où $\alpha, \beta, \gamma_1, \gamma_2$ et γ_3 sont des coefficients qui sont déterminés empiriquement à partir de l'évaluation subjective d'une base d'images d'apprentissage.

Cette méthode présente l'avantage d'être économique en ressources machines, ce qui permet son utilisation non seulement pour des images fixes, mais aussi pour des séquences vidéos.

Cette méthode a été construite pour être appliquée uniquement dans le cas d'une compression JPEG. Cependant, une méthode similaire a été mise au point par Sheikh et al. [She-02] pour la compression JPEG 2000. Cette méthode utilise alors un modèle statistique des coefficients d'ondelettes. On travaille ici sur les valeurs normalisées de la luminance de l'image. On considère l'histogramme normalisé $h(c)$ du logarithme de l'amplitude des coefficients d'ondelettes à une échelle et à une orientation données. On estime la probabilité p_s que l'amplitude d'un coefficient soit supérieure à un seuil T_h à partir de l'histogramme normalisé à l'aide de la formule suivante :

$$p_s = \sum_{c > T_h} h(c) \quad (2-2-14)$$

On calcule un vecteur des probabilités $\{p_{si}\}$ où l'indice i représente la bande de fréquence sur laquelle on travaille. On utilise les deux bandes d'échelles les plus fines pour les trois directions horizontale, verticale et diagonale. On réduit la dimension de ce vecteur à une seule composante p_w à l'aide d'une analyse en composantes principales, en utilisant la formule suivante :

$$p_w = \sum_{i=1}^6 c_i (p_{si} - \mu_{si}) \quad (2-2-15)$$

où c_i représente les coefficients du premier vecteur de base de l'analyse en composantes principales et μ_{si} les coefficients du vecteur des moyennes. Les auteurs de l'article ne précisent pas comment ils ont fait pour appliquer une analyse en composante principale sur le vecteur $\{p_{si}\}$.

On utilise alors l'équation ci-dessous pour calculer le facteur de qualité Q_P , selon la formule suivante :

$$Q_P(p_w) = K \left(1 - \exp \left(- \frac{(p_w - u)}{T} \right) \right) \quad (2-2-16)$$

K est la valeur maximale du facteur de qualité que l'algorithme puisse donner, u est un coefficient de réglage et T la constante de dégradation. Les différents paramètres de la formule et les vecteurs de base de l'analyse en composantes principales sont déterminés à l'aide d'une phase d'apprentissage, à partir d'images dont la qualité a été évaluée par des observateurs humains à l'aide d'une méthode subjective.

Les critères univariés ont pour principal avantage de pouvoir se passer d'image de référence, cela est particulièrement utile lorsque l'on ne dispose pas de l'image originale non compressée, ce cas est fréquent dans la pratique. Cependant, ces critères sont souvent empiriques et leurs paramètres doivent être déterminés expérimentalement à partir de l'évaluation subjective d'une base d'apprentissage. De plus, ces critères sont spécialisés en fonction d'un type d'image particulier, comme les images aériennes, ou d'un type spécifique de dégradation comme la compression JPEG. Cela limite leur champ d'application à certains cas bien précis. Nous allons donc maintenant étudier un certain nombre de propositions qui ont été faites dans le cadre d'approches utilisant l'image initiale et l'image décompressée pour mesurer la dégradation.

2.3 Les critères numériques bivariés classiques

La plupart des critères numériques utilisés pour évaluer la qualité de la compression sont donc des critères bivariés, ils prennent en entrée les deux images, originale et décompressée et les comparent directement l'une à l'autre. De nombreux critères numériques bivariés ont été proposés.

Les critères les plus anciennement utilisés pour évaluer la qualité d'une image compressée sont l'erreur quadratique moyenne MSE (Mean Squared Error) et le rapport signal/bruit, PSNR (Peak Signal Noise Ratio). L'erreur quadratique moyenne est calculée entre deux images a et b de même taille $M \times N$. L'image a est l'image originale et l'image b est l'image compressée. Elle est définie par la formule suivante :

$$MSE = \frac{1}{MN} \sum_{x=1}^M \sum_{y=1}^N (a(x,y) - b(x,y))^2 \quad (2-3-1)$$

On peut également utiliser l'erreur notée RMSE (Root Mean Square Error) qui est tout simplement la racine carrée de l'erreur MSE. Un autre de ces critères est l'erreur absolue moyenne EAM, calculée selon la formule suivante :

$$EAM = \frac{1}{MN} \sum_{x=1}^M \sum_{y=1}^N |a(x,y) - b(x,y)| \quad (2-3-2)$$

Plus généralement, une L_p -norme est définie par la formule suivante :

$$L_p(a,b) = \left(\frac{1}{MN} \sum_{x=1}^M \sum_{y=1}^N |a(x,y) - b(x,y)|^p \right)^{1/p} \quad (2-3-3)$$

L'EAM et le RMSE sont des cas particuliers de L_p -norme, avec respectivement $p=1$ et $p=2$. Il est également possible de normaliser l'erreur par rapport à la somme des niveaux de gris de l'image originale, on obtient alors l'erreur absolue normalisée ou NAE, utilisée notamment par Eskicioglu et Fisher [Esk-95], où elle est donnée par la formule

$$NAE(a,b) = \frac{\sum_{x=1}^M \sum_{y=1}^N |a(x,y) - b(x,y)|}{\sum_{x=1}^M \sum_{y=1}^N |a(x,y)|} \quad (2-3-4)$$

Une image peut être considérée comme un signal, et la dégradation de cette image due à la compression peut être assimilée à un bruit, c'est pourquoi on peut estimer la qualité d'une image compressée en calculant le rapport signal/bruit. Le PSNR peut se calculer à partir de l'erreur MSE en utilisant la formule suivante, où H désigne le nombre de niveaux de gris de l'échelle sur laquelle est définie l'image :

$$PSNR = 10 \cdot \log_{10} \left(\frac{H^2}{MSE} \right) \quad (2-3-5)$$

En plus des L_p -normes et du PSNR, qui sont de loin les plus employés, d'autres critères numériques bivariés ont été proposés, notamment par Eskicioglu [Esk-95]. Parmi ces critères, on peut notamment citer la fidélité spatiale, le contenu structurel ou la cross-corrélation normalisée.

Plus récemment, un critère basé sur les propriétés statistiques de l'image a été proposé par Wang et Bovik [Wan-02b]. Il s'agit de l'indice universel de qualité d'image, ainsi appelé car

il peut être appliqué pour n'importe quel couple d'images, quelles que soient les conditions de visualisation et le type de distorsion. Cet indice de qualité est défini comme le produit de trois facteurs qui dépendent de la moyenne, de l'écart type et de la covariance de l'image originale a et de l'image dégradée b . Ces trois facteurs représentent la perte de corrélation, la distorsion de luminance et la distorsion de contraste. L'indice de qualité Q est défini par la formule suivante :

$$Q = \frac{\sigma_{ab}}{\sigma_a \sigma_b} \times \frac{2\bar{a}\bar{b}}{(\bar{a})^2 + (\bar{b})^2} \times \frac{2\sigma_a \sigma_b}{\sigma_a^2 + \sigma_b^2} \quad (2-3-6)$$

Dans cette formule, \bar{a} et \bar{b} sont les moyennes des niveaux de gris des deux images, σ_a et σ_b les écarts type et σ_{ab} la covariance des deux images. Ce critère a une valeur comprise entre -1 et $+1$, la valeur maximale 1 étant obtenue lorsque les deux images sont identiques. Ce critère donne une meilleure évaluation de la dégradation de la qualité de l'image que des critères comme le MSE, car il modélise mieux la distorsion structurelle de l'image compressée.

Le principal avantage de ces critères classiques est d'être très simples à calculer. Cependant, les résultats numériques ne sont pas toujours satisfaisants car ils ne sont pas en correspondance avec la perception visuelle humaine, notamment parce qu'ils ne prennent pas en compte le type de distorsion ni la position des erreurs dans l'image. Ainsi, des images présentant des dégradations très différentes peuvent avoir une même valeur de leur critère de qualité. C'est pourquoi d'autres mesures plus élaborées de la qualité de la compression ont été proposées, en particulier des critères se basant sur une modélisation de la vision humaine. Néanmoins, les critères classiques de type PSNR ou MSE sont toujours très utilisés, notamment à des fins d'évaluation comparative des différentes mesures de qualité de l'image.

2.4 Les critères pondérés

Les critères pondérés sont une adaptation des mesures numériques classiques, que l'on modifie en appliquant une pondération afin de prendre en compte la sensibilité du système visuel humain et le type de distorsion. Ils permettent ainsi de mieux modéliser la perception différente de l'œil humain selon la nature de la distorsion subie par l'image.

Le modèle de Daly simplifié [Dal-93] consiste à pondérer les mesures classiques, PSNR et EAM par une fonction de sensibilité au contraste appelée également fonction de transfert de modulation (MTF) qui dépend de la fréquence spatiale radiale du signal. Ce modèle est destiné à évaluer la qualité des images compressées en JPEG. On construit une matrice de pondération qui affecte un poids à chacun des blocs de la DCT. La fonction de pondération $W(u,v)$ est donnée par la formule suivante :

$$W(u,v) = \begin{cases} 2,2(0,192 + 0,114\tilde{f}(u,v))\exp(-0,114\tilde{f}(u,v))^{1,1} & \text{si } \tilde{f}(u,v) > f_{\max} \\ 1,0 & \text{sinon} \end{cases} \quad (2-3-1)$$

Dans cette formule, $\tilde{f}(u,v)$ désigne la fréquence spatiale radiale en cycles/degrés et f_{\max} est la fréquence pour laquelle l'exponentielle est maximale. La fréquence spatiale radiale dépend de l'orientation, de la résolution de l'image et de la distance d'observation. Les éléments de la matrice de pondération W pondèrent les coefficients de la DCT des blocs de taille 8×8 de l'image et indiquent l'importance des fréquences spatiales.

Cette fonction de pondération va être utilisée pour calculer des mesures pondérées du PSNR et de EAM. La pondération permet de diminuer les valeurs des erreurs dans les hautes

fréquences qui sont supposées être moins visibles par le système visuel humain. Les mesures pondérées de dégradation sont des valeurs moyennes sur l'image de mesures locales de dégradation calculées sur des blocs de l'image. La mesure de PSNR pondérée est donnée par la formule suivante :

$$\text{PSNRP} = 10\log_{10}(\text{Max}^2/\text{EQMP}) \quad (2-3-2)$$

où Max est l'amplitude de l'échelle des niveaux de gris de l'image, et EQMP l'erreur quadratique moyenne pondérée de toute l'image. L'EQMP est définie comme étant la moyenne sur toute l'image des erreurs quadratiques locales pondérées calculées sur des blocs de taille 8x8.

$$\text{EQMP}_p(a,b) = \frac{1}{N^2} \sum_{u=1}^N \sum_{v=1}^N W_{u,v} (a(u,v) - b(u,v))^2 \quad (2-3-3)$$

a et b sont respectivement l'image originale et l'image compressée, $W_{u,v}$ est le poids de la pondération à la fréquence spatiale u,v et N est la taille du bloc, ici égale à 8x8.

De la même manière on peut calculer l'EAM pondérée ou l'EAMP qui est la moyenne sur toute l'image des valeurs de l'EAM pondérée localement ou EAMP_p . L'EAMP_p est donnée par la formule :

$$\text{EAMP}_p = \frac{1}{N^2} \sum_{u=1}^N \sum_{v=1}^N W_{u,v} |a(u,v) - b(u,v)| \quad (2-3-4)$$

Le plus souvent, le PSNRP donne une meilleure mesure de la dégradation visuelle de l'image que le PSNR, sauf dans les cas où les images dégradées sont très floues. La méthode n'est cependant applicable qu'à des images compressées en JPEG ou à des méthodes qui considèrent une partition en bloc de l'image pour la compresser.

La méthode de Fränti [Fra-98] est une méthode d'évaluation de la qualité visuelle qui intègre trois aspects différents : les propriétés du système visuel humain, les facteurs de qualité et la pondération des différents facteurs. Parmi les facteurs de qualité on peut noter l'erreur de contraste, l'erreur de structure et l'erreur de quantification en niveaux de gris. Les différents critères sont calculés sur des blocs de pixels de taille 3x3. Le critère de contraste pour un pixel donné est défini par la formule suivante :

$$d_1 = \frac{(\sigma_a - \sigma_b)}{\max(1, \sigma_a)} \quad (2-3-5)$$

Dans cette formule, a représente l'image originale, b l'image compressée et σ désigne l'écart type des niveaux de gris dans le voisinage considéré. Le critère de structure d_2 est défini par la formule suivante :

$$d_2 = \frac{|Gx_a - Gx_b| + |Gy_a - Gy_b|}{2\max(1, \sigma_a)} \quad (2-3-6)$$

Gx et Gy représentent les résultats d'une détection de contours horizontaux et verticaux obtenue avec les filtres de Sobel. La normalisation par σ du bloc original exprime le masquage des erreurs par le contraste. Le critère de quantification est défini par la formule :

$$d_3 = (Q_a - Q_b)^2 \quad (2-3-7)$$

où Q_a et Q_b sont les nombres de niveaux de gris distincts des deux images. On calcule ensuite les valeurs moyennes sur toute l'image D_1 , D_2 et D_3 des 3 critères d_1 , d_2 et d_3 définis précédemment. Le facteur de distorsion D qui mesure la qualité de l'image compressée est calculé comme une fonction linéaire de ces trois valeurs D_i selon la formule :

$$D = w_1f(D_1) + w_2f(D_2) + w_3f(D_3) \quad (2-3-8)$$

La fonction f étant définie par :

$$f(D_i) = 1 - \min\left(1, \frac{D_i}{k_i}\right) \quad (2-3-9)$$

Les valeurs des poids w_i et des coefficients k_i ont été déterminés empiriquement par Fränti à partir d'une unique image de référence, à savoir la célèbre image *Lena*, afin de maximiser la corrélation entre le facteur de distorsion D et la qualité subjective sur cette image. Les valeurs des coefficients obtenues sont alors de $k_1=3$, $k_2=k_3=32$, $w_1=0,45$, $w_2=0,30$, $w_3=0,25$. Cette méthode a été testée sur des images compressées avec les codages BTC, QT et JPEG. Les résultats obtenus par cette mesure ont été comparés avec ceux obtenus avec le MSE et le PQS défini par Miyahara et al. [Miy-98]. L'évaluation comparative consiste à calculer le coefficient de corrélation entre la mesure objective D , MSE ou PQS et une mesure de qualité subjective estimée par un groupe d'observateurs humains sur des images à différents taux de compression. Les résultats obtenus avec la méthode de Fränti sont meilleurs que ceux obtenus avec les deux autres méthodes, la corrélation avec la mesure subjective étant plus importante, de plus cette méthode est plus simple à implémenter et son temps de calcul est plus réduit.

La méthode proposée par Tamtaoui et Aboutajdine [Tam-03] est basée sur la densité de pertinence. La pertinence des erreurs peut ici être définie comme la visibilité des distorsions dues à leur concentration dans des zones de l'image, en effet les erreurs sont d'autant plus visibles qu'elles sont concentrées en une même région. La méthode consiste à calculer dans une fenêtre d'analyse une valeur de la dégradation locale à l'aide d'une méthode standard, et à pondérer cette mesure en fonction de sa densité de pertinence. Comme avec la méthode de Fränti, trois erreurs sont calculées : l'erreur de contraste, l'erreur de structure et l'erreur de quantification. Les erreurs sont calculées dans des fenêtres d'analyse dont la taille varie entre 3x3 pixels et la taille de l'image. La taille des fenêtres permet de jouer sur le niveau de détail de l'analyse. Les fenêtres doivent recouvrir toute l'image de manière glissante ou non. Pour obtenir une seule mesure finale de dégradation, on cumule les valeurs pondérées en une seule valeur en utilisant une fonction de cumul. La densité de pertinence locale est calculée sur des blocs de 3x3 pixels à l'intérieur d'une fenêtre d'analyse. Elle se calcule à l'aide de la formule suivante :

$$D^i = \frac{N_p^i}{N} \quad (2-3-10)$$

N_p^i est le nombre de blocs pertinents dans la fenêtre i et N est le nombre total de blocs de cette fenêtre. Un bloc est considéré comme pertinent si la moyenne de ses erreurs de contraste, de structure et de quantification est supérieure à un seuil S prédéterminé. La mesure des dégradations dans la fenêtre d'analyse I est donnée par la formule :

$$DP_i = M_i(1 - \alpha D^i) \quad (2-3-11)$$

où α est un paramètre d'adaptation constant identique pour toutes les fenêtres de l'analyse. M_i est une mesure de dégradation sur la fenêtre d'analyse i . La mesure de dégradation utilisée ici est l'EAM. La comparaison entre l'EAM et la qualité subjective fait apparaître deux cas : dans le premier cas, l'EAM augmente rapidement en passant d'une image peu dégradée à une autre plus dégradée alors que la qualité subjective diminue moins rapidement, dans ce cas α est positif. Dans le second cas, l'EAM augmente légèrement en passant d'une image peu dégradée à une autre plus dégradée alors que la qualité subjective augmente rapidement, dans ce cas α est négatif.

Pour la fonction de cumul, sachant qu'un seul défaut, même minime, peut gâcher toute l'image, la mesure finale choisie est la valeur maximale de la dégradation DP_i .

La mesure définie précédemment, appelée EAM-DEL, a été comparée à l'EAM pondérée, au PSNR, au PSNRP et à la mesure de Fränti. La comparaison a été effectuée en calculant un coefficient de corrélation avec une mesure subjective.

2.5 Les critères perceptuels

Les critères perceptuels sont basés sur une modélisation explicite du principe de fonctionnement du système visuel humain considéré du point de vue de sa perception de la qualité d'une image.

Le critère proposé par Osberger et al. [Osb-98a] est basé sur un modèle qui exploite plusieurs propriétés du système visuel humain. Parmi ces facteurs, on peut notamment citer la sensibilité au contraste et le masquage. La sensibilité au contraste est approximée par la loi de Weber aux niveaux de luminosité photopiques, et par la loi de de Vries aux niveaux de luminosité plus faibles. La sensibilité différente aux stimuli visuels selon la fréquence peut être modélisée par une fonction de sensibilité au contraste (CSF). Le masquage représente notre habileté réduite à détecter un stimulus visuel dans un arrière-plan complexe, ce qui signifie que les erreurs sont moins visibles à proximité des contours, dans les régions texturées et immédiatement après un changement de scène. Selon Watson et al. [Wat-97], le masquage dépend non seulement du contraste du fond, mais aussi de l'incertitude que génère ce fond. Ainsi les zones d'incertitude élevée comme les textures complexes génèrent un masquage plus important que les zones de même contraste et d'incertitude plus faible comme les contours. D'autres facteurs de plus haut niveau interviennent sur la perception par un observateur humain de la qualité visuelle d'une image, ils concernent le niveau d'attention visuelle apportée aux différentes régions d'une image. Ainsi, notre sensibilité visuelle est plus élevée pour les régions situées au centre de l'image. L'attention visuelle dépend également du contenu de l'image, c'est ainsi que nous accordons une grande importance aux visages humains et, à l'intérieur du visage, aux yeux et à la bouche, c'est pourquoi une distorsion dans ces parties de l'image nous paraît plus visible. Le modèle proposé par Osberger et al. [Osb-98a] permet de prendre en compte ces différents facteurs d'attention visuelle afin de mesurer la distorsion d'une image.

On calcule d'abord le contraste local de bande limitée défini par Peli. Pour une bande de fréquence donnée, c'est le rapport entre l'image filtrée par un filtre passe-bande et la moyenne locale de la luminance. Ce calcul est effectué pour l'image initiale comme pour l'image compressée. La fonction de sensibilité au contraste est utilisée pour déterminer le seuil de détection d'erreur qui dépend de la fréquence. Un masquage est effectué en classifiant les points de l'image en textures et contours, afin de modéliser le fait que les distorsions sont

moins apparentes au niveau des textures que des contours. On effectue pour cela une détection de contour à l'aide d'un filtre de Sobel. Le masquage est effectué par élévation du seuil en fonction du contraste. Si le contraste local C_m est supérieur à un seuil de base théorique prédéfini C_{th0} , on multiplie le seuil par $(C_{th0}/C_m)^\epsilon$, la valeur de ϵ étant différente suivant que le point appartient ou non à un contour. On augmente ainsi le seuil de détection dans les régions fortement contrastées. Ce seuil est utilisé pour déterminer la visibilité de l'erreur pour chaque point de l'image et pour chaque bande de fréquence. Pour un point (x,y) , la visibilité de l'erreur v_k pour une bande de fréquence k est donnée par la formule :

$$v_k(x,y) = (c(x,y) - c'(x,y)) / C_{thk}(x,y) \quad (2-5-1)$$

où $c(x,y)$ et $c'(x,y)$ représentent respectivement le contraste de l'image initiale et de l'image compressée, et C_{thk} est le seuil précédemment calculé. On calcule ensuite pour chaque point (x,y) la visibilité de l'erreur $V(x,y)$ qui est la somme des valeurs de $v_k(x,y)$ de toutes les bandes de fréquence. On obtient ainsi une carte de distorsion qui donne pour chaque pixel la visibilité de l'erreur introduite par la compression. Afin d'obtenir une mesure de la qualité globale de l'image, on calcule l'indice de qualité PQR (Picture Quality Ratio) sur une échelle de 1 à 5, en utilisant une sommation de Minkowski selon la formule suivante :

$$PQR = \frac{5}{1 + p^E} \quad (2-5-2)$$

où p est une constante égale à 0,8 et E est la somme de l'erreur sur toute l'image, calculée à l'aide de la formule :

$$E = \left(\frac{1}{N} \sum_{x=1}^M \sum_{y=1}^N V(x,y)^\alpha \right)^{1/\alpha} \quad (2-5-3)$$

La valeur $\alpha=2$ assure un résultat conforme à l'opinion subjective humaine. Les erreurs étant plus visibles dans les régions perceptuellement importantes de l'image, on peut prendre en compte ce fait en utilisant une *carte d'importance* qui définit l'importance visuelle de chaque région de l'image. Cette carte d'importance est calculée en utilisant la méthode décrite dans [Os98b]. L'utilisation de cette carte d'importance se fait en pondérant la visibilité de l'erreur $V(x,y)$ en chaque point de l'image par l'importance visuelle de ce point de l'image. On obtient ainsi la carte de distorsion pondérée de l'image.

Cette méthode a pour intérêt de prendre en compte la perception visuelle pour le calcul de la mesure de la distorsion de l'image. Elle est indépendante de la méthode de compression utilisée. Elle permet en outre de localiser l'erreur dans l'image à l'aide de la carte de distorsion, elle peut donc être utilisée soit comme méthode numérique, soit comme méthode graphique.

Le PQS (Picture Quality Scale) est une mesure objective de la qualité d'image proposée par Miyahara et al. [Miy98]. Cette méthode utilise différentes caractéristiques structurelles de l'image, aussi bien locales que globales pour déterminer la qualité de l'image. Le principe général de cette méthode consiste à calculer, à partir de l'image originale $a(m,n)$ et de l'image dégradée $b(m,n)$, des cartes locales de distorsion qui expriment la valeur locale, en chaque point (m,n) de l'image, de différents facteurs de distorsion f_i . A partir de ces cartes on calcule la valeur globale F_i du facteur sur toute l'image. On utilise ensuite des méthodes de régression afin de combiner ces facteurs en une seule valeur numérique représentant la qualité d'une image : le PQS. Tout d'abord, afin d'obtenir une échelle perceptuelle uniforme, on transforme les images en utilisant une loi puissance qui approxime la loi de Weber-Fechner

pour la sensibilité lumineuse. L'image ajustée $x(m,n)$ est calculée à partir de l'image originale $a(m,n)$ selon la formule :

$$x(m,n)=k a(m,n)^{1/2,2} \quad (2-5-4)$$

où k est une constante d'échelle qui permet l'ajustement de l'échelle de la variable x . L'image de l'erreur ajustée du contraste $e(m,n)$ est définie comme la différence entre les valeurs de $x(m,n)$ calculées pour l'image originale et l'image compressée. La plupart des facteurs de distorsion utilisés par la méthode sont des fonctions de $e(m,n)$. Les erreurs sont pondérées suivant la fréquence spatiale afin de modéliser la sensibilité différente de la vision humaine selon la fréquence spatiale du signal visuel. La méthode de Miyahara et al. utilise 5 facteurs de distorsion pour calculer l'erreur. Pour chacun de ces facteurs, on calcule une carte de distorsion $\{f_i(m,n)\}$ à partir de laquelle on calcule le facteur de distorsion $\{F_i\}$.

Le premier facteur F_1 mesure le bruit dans l'image. Il est calculé à partir de l'erreur locale pondérée en fréquence, celle-ci est le produit de l'erreur $e(m,n)$ par une fonction $w_{TV}(m,n)$ qui représente la pondération isotropique du domaine spatial définie par le CCIR 567-1. Le facteur de distorsion est alors calculé comme la somme des carrés des erreurs de fréquence locales pondérées. Le facteur F_1 est normalisé par l'énergie pondérée de l'image, il s'agit en fait d'un rapport signal/bruit normalisé.

Le facteur F_2 est un facteur de distorsion qui prend en compte la loi de Weber et l'anisotropie de la réponse fréquentielle. Il est calculé à partir de la formule :

$$f_2(m,n)=I_T(m,n)[e_w(m,n)*s_a(m,n)]^2 \quad (2-5-5)$$

où $s_a(m,n)$ est une fonction de pondération *anisotropique* et où l'on ignore les valeurs de l'erreur inférieures à un certain seuil, $e_w(m,n)$ représente cette erreur seuillée. I_T représente une fonction indicatrice pour un seuil perceptuel de visibilité. Comme F_1 , F_2 est normalisé par rapport à l'énergie pondérée de l'image.

Le facteur F_3 mesure les distorsions de fin de bloc, c'est à dire les discontinuités entre les blocs de la compression JPEG. Il est constitué de deux composantes F_{3h} et F_{3v} qui représentent les discontinuités de blocs horizontaux et verticaux. On calcule de la même manière, mais séparément, les composantes horizontale et verticale de ce facteur de distorsion. Pour la composante horizontale, on calcule :

$$f_{3h}(m,n)=I_h(m,n)\Delta_h^2(m,n) \quad (2-5-6)$$

où Δ_h est la différence des niveaux de gris entre deux pixels adjacents d'une même ligne de l'image d'erreur et I_h une fonction qui permet de considérer uniquement les pixels adjacents appartenant à deux blocs différents. Le facteur F_{3h} est la moyenne des valeurs de f_{3h} sur toute l'image. La composante verticale F_{3v} est calculée exactement de la même manière, mais selon la direction verticale. Le facteur F_3 est enfin calculé comme la racine carrée de la somme des carrés des deux composantes.

Le facteur de distorsion F_4 est la moyenne sur l'image des erreurs de corrélation. Les erreurs de corrélation sont calculées dans une fenêtre W de taille 5×5 centrée autour de chaque pixel par la formule :

$$f_4(m,n)=\sum_{(k,l) \in W} |r(m,n,k,l)|^{0,25} \quad (2-5-7)$$

où $r(m,n,k,l)$ est une fonction de corrélation locale.

Le facteur F_5 mesure les erreurs dans le voisinage des transitions de contraste élevé. Il modélise l'effet de masquage visuel qui fait que l'oeil humain est moins sensible aux erreurs dans les régions de contraste élevé. L'erreur locale $f_5(m,n)$ est calculée par la formule :

$$f_5(m,n) = (I_m(m,n)) / e_w(m,n) / (S_h(m,n) + S_v(m,n)) \quad (2-5-8)$$

où $I_m(m,n)$ est une fonction qui sélectionne les pixels proches des transitions d'intensité, $S_h(m,n)$ et $S_v(m,n)$ sont ses facteurs de masquage horizontal et vertical, ils sont définis à partir d'une fonction locale d'activité. Le facteur F_5 est la moyenne des valeurs de f_5 sur les pixels dont la réponse à l'opérateur de détection de contours de Kirsch est supérieure à un certain seuil.

Une analyse en composantes principales est ensuite effectuée sur les facteurs de distorsion précédemment calculés, afin de quantifier la corrélation entre ces différents facteurs. On calcule la matrice de covariance $C_F = E\{(\mathbf{F} - \mu_F)(\mathbf{F} - \mu_F)^T\}$ où \mathbf{F} est le vecteur des facteurs de distorsion et μ_F leur moyenne. L'espérance se calcule par rapport aux valeurs globales $\{F_1 \dots F_5\}$ des facteurs de distorsion sur l'image testée. Les valeurs propres de cette matrice représentent les contributions relatives des différents facteurs. Les vecteurs propres sont non corrélés et sont par conséquent plus robustes et plus efficaces pour une estimation objective de la qualité de l'image. Le PQS est calculé comme une combinaison linéaire des coefficients Z_j de l'analyse en composantes principales selon la formule suivante :

$$PQS = b_0 + \sum_{j=1}^3 b_j Z_j \quad (2-5-9)$$

Les Z_j sont ici les trois premières composantes principales de la décomposition de la matrice de covariance C_F .

Les résultats obtenus sur des images de test ont permis de déterminer ce que représentent ces différents coefficients Z_j . Z_1 représente une quantité d'erreur moyenne sur l'image, Z_2 mesure la position des erreurs par rapport aux bords de l'image et Z_3 caractérise la structure de l'ensemble des erreurs. Les b_j sont les coefficients de régression partielle qui sont calculés en utilisant une analyse de régression multiple du PQS avec les résultats d'une évaluation subjective du MOS d'une base d'images de test. Le PQS donne une bonne estimation de la qualité visuelle des images, cependant il est complexe à calculer, du fait notamment de l'analyse en composantes principales.

La méthode proposée par Lai et Kuo [Lai-97] utilise les ondelettes pour mesurer la qualité des images compressées. Cette méthode est basée sur des mesures de contraste sur les différentes bandes de fréquence d'une transformée en ondelettes de Haar. La transformation est effectuée sur l'image originale et sur l'image compressée, et le contraste C est calculé pour chaque bande de fréquence en utilisant la formule de Michaelson :

$$C = \frac{L_{\max} - L_{\min}}{L_{\max} + L_{\min}} \quad (2-5-10)$$

où L_{\max} et L_{\min} sont respectivement la luminance maximale et minimale. La sensibilité au contraste du système visuel humain étant différente selon la fréquence, on modélise ce fait en calculant une réponse au contraste R qui dépend d'un seuil de contraste C_T . Le seuil C_T dépend de la bande de fréquence considérée, il est calculé en utilisant la formule suivante :

$$C_T = C_{T0} \left(\sum_{k=0}^N \left(\frac{C_k}{C_{T0}} \right)^m \left(\frac{f_k}{f_0} \right)^{m_2} \right) \quad (2-5-11)$$

Dans cette formule, C_{T0} est le contraste de la bande de fréquence la plus basse, $N+1$ est le nombre de bandes de fréquence affectant la perception de la fréquence considérée, C_k et f_k sont le contraste et la fréquence de la bande k et m_1 et m_2 sont des coefficients qui dépendent de la bande de fréquence considérée. La réponse R est donnée par la formule suivante :

$$R=k(C-C_T)^p \quad (2-5-12)$$

La valeur de l'exposant p ayant été fixée à 0,45. La valeur de la réponse R est calculée en chaque pixel (i,j) pour chacune des bandes de fréquence, tant pour l'image originale que pour l'image compressée. On calcule ensuite la distance D entre les deux images en utilisant la formule suivante :

$$D=\frac{1}{HV} \left(\sum_{k=1}^N \sum_{j=1}^V \sum_{i=1}^H (R_{b,k}(i,j) - R_{a,k}(i,j))^2 \right) \quad (2-5-13)$$

Dans cette formule, H et V représentent la largeur et la hauteur de l'image et N le nombre de bandes de fréquence, $R_{b,k}$ et $R_{a,k}$ sont respectivement la valeur de la réponse R pour la bande. La distance D mesure la qualité de l'image compressée par rapport à celle de l'image originale sous la forme d'un nombre sans dimension.

2.6 Les critères graphiques

Contrairement aux critères précédents qui donnent l'évaluation de la distorsion sous la forme d'une simple valeur scalaire, les critères graphiques permettent de représenter la distorsion de l'image par une représentation graphique. On peut ainsi non seulement évaluer quantitativement la distorsion de l'image mais aussi sa répartition spatiale et/ou sa nature. Ces méthodes calculent des caractéristiques qui permettent de visualiser les distorsions de l'image et de déterminer la position des erreurs. Ces méthodes ne sont pas limitées à la détection d'erreurs de compression, elles peuvent être utilisées pour représenter d'autres types de distorsions dues par exemple au bruit.

Certaines méthodes déjà présentées comme méthodes numériques peuvent être également utilisées en tant que méthodes graphiques. C'est le cas des méthodes qui utilisent une mesure locale de l'erreur représentée par des cartes de distorsion, comme la méthode de Osberger et al. [Os-98a] déjà présentée. Dans ce cas, au lieu de calculer une mesure globale à partir des mesures locales, on utilise la carte de distorsion afin de déterminer les zones de l'image où la distorsion est la plus importante.

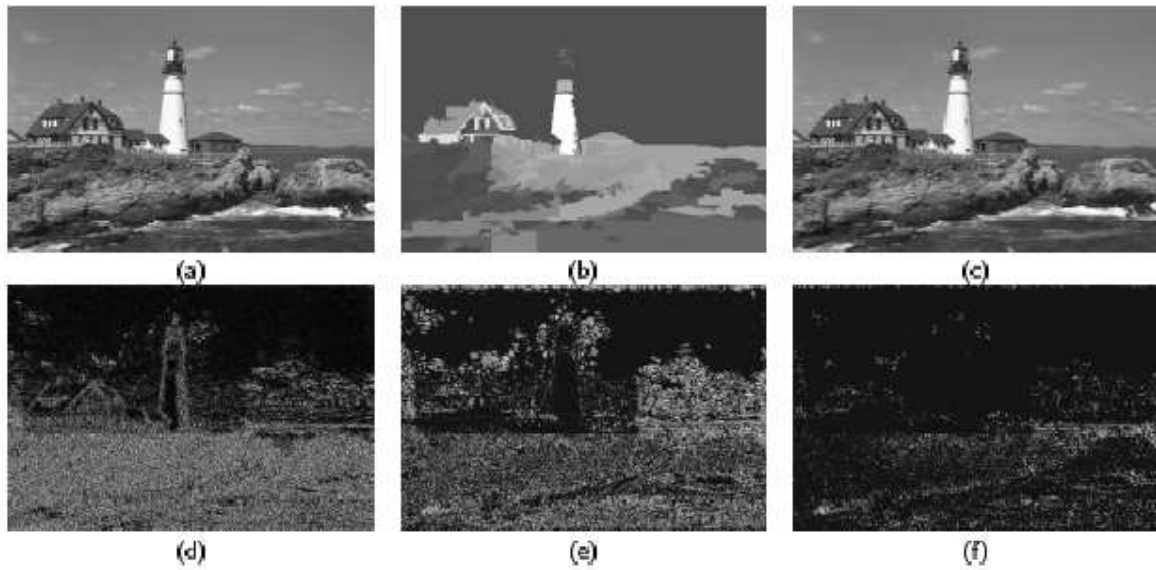


Fig.1 : Méthode de Osberger et al. Image originale(a) , carte d'importance (b), Image compressée (c) , mesure locale du MSE (d), carte de distorsion (e) , carte de disorsion pondérée par la carte d'importance (f). Source : [Os-98a]

Les diagrammes de Eskioglu [Esk-96] ont été mis au point pour évaluer la qualité des images médicales. Ils s'appuient sur une modélisation du système visuel humain pour calculer des mesures locales représentées sous forme graphique. Les diagrammes de Hosaka [Hos-86] sont une représentation des déformations de l'image sous forme d'un graphe en coordonnées polaires. L'image est divisée en blocs, et on calcule pour chaque bloc la différence entre la moyenne et l'écart type des niveaux de gris de l'image originale et de l'image décompressée, ces valeurs sont ensuite représentées dans le diagramme dont la forme permet de déterminer la qualité et le type de dégradation subie par l'image.

Le « Visual Differences Predictor » de Daly [Dal-93] calcule une carte de différence qui prédit la probabilité de détection de dissimilarités entre l'image originale et l'image dégradée. Les deux images sont filtrées par une fonction de réponse non linéaire qui modélise à la fois l'adaptation et la réponse non-linéaire des neurones de la rétine. Dans ce modèle, on suppose que le système visuel s'adapte séparément à chaque pixel. Le modèle d'adaptation utilisé approxime la relation entre la sensation de luminosité et la luminance réelle de l'image. Pour une faible luminance, cette relation suit une loi en racine cubique et pour une luminance élevée elle suit une loi de dépendance logarithmique. On applique ensuite une transformée en fréquence sur l'image filtrée. Le spectre d'amplitude est ensuite pondéré par une fonction de sensibilité au contraste (CSF) : pour chaque fréquence, l'amplitude est multipliée par la valeur de cette fonction pour la fréquence considérée. La fonction de modulation analytique de Barten donne de bons résultats en tant que fonction de sensibilité au contraste. Le signal est ensuite divisé en 31 canaux qui correspondent aux différentes orientations et bandes de fréquence (6 bandes d'orientation pour chacune des 5 bandes de fréquence, et la bande de base indépendante de l'orientation). Les signaux de chacun des canaux sont ensuite transformés à nouveau dans le domaine spatial. Chaque canal est associé à un contraste de masque qui est une fonction de la position dans l'image. La présence d'information de masquage à une position spécifique, la fréquence spatiale et l'orientation augmentent le seuil de détection pour un signal ayant ces caractéristiques. Pour chaque canal, une carte

d'élévation du seuil est calculée comme une fonction du contraste de masque. On effectue ensuite un masquage des deux ensembles de cartes d'élévation du seuil pour les deux images, afin d'obtenir une seule carte d'élévation par canal. On peut maintenant calculer la probabilité de détection. On soustrait le contraste des canaux d'une image de celui de l'autre image, et la différence est pondérée par l'élévation du seuil. Les probabilités de détection sur les différents canaux sont ensuite combinées en faisant l'hypothèse qu'elles sont indépendantes. On obtient alors la carte de prédiction qui donne la probabilité de détection pour chaque point de l'image.

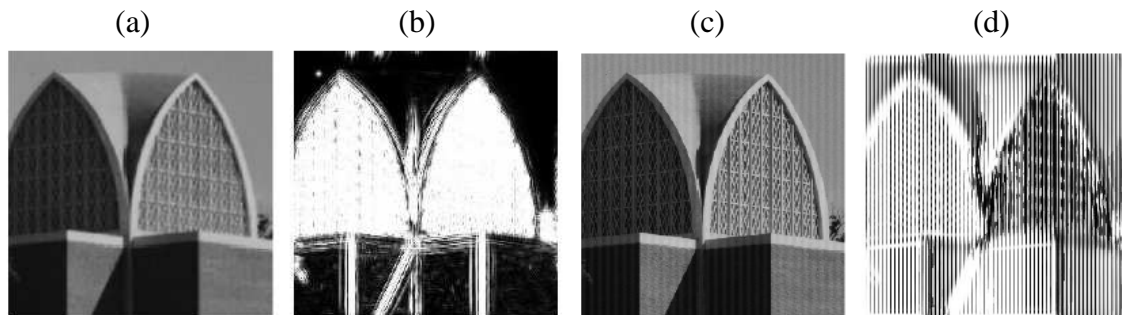


Fig. 3 : Application du modèle de Daly sur deux images bruitées. Source : [Li-98] . De gauche à droite : (a) image bruitée par un flou, (b) carte de différence VDP pour le flou, (c) image bruitée par un bruit sinusoïdal, (d) carte de différence VDP pour le bruit sinusoïdal.

Le modèle de discrimination visuelle (VDM) de Sarnoff, proposé par Lubin [Lub-95] opère dans le domaine spatial. Ce modèle comprend un rééchantillonnage spatial, une décomposition pyramidale, un transducteur qui calcule la différence que l'on peut juste remarquer (Just Noticeable Difference = JND), une normalisation et une modélisation de l'effet de plongée. A partir de deux images et d'un jeu de paramètres qui prennent en compte les conditions de visualisation, le modèle calcule une carte de différence juste remarquable.

La première partie du modèle modélise l'œil et la mosaïque rétinienne. Une fonction d'étalement des points modélise la réponse de l'optique de l'œil. L'image est ensuite découpée en blocs. On effectue une décomposition en ondelettes à l'aide d'une pyramide laplacienne de l'image échantillonnée en blocs et on utilise une paire symétrique en quadrature de noyaux de convolution pour représenter les informations sur 4 bandes d'orientation. Le signal original est alors transformé en unités de contraste local. On utilise une pyramide laplacienne pour décomposer le signal en sept bandes de fréquences. La pyramide utilisée est une pyramide directionnelle, qui est une transformation multi-échelle et multi-orientation comportant des composantes de fréquence et d'orientation. L'étape finale de la décomposition est le calcul d'une réponse d'énergie indépendante de la phase par calcul de la somme quadratique des coefficients de phase paire et impaire.

La phase de normalisation consiste à normaliser l'énergie par la réciproque de la fonction de sensibilité au contraste. Un transducteur est utilisé pour affiner la carte de différence juste remarquable en prenant en compte l'effet de plongée du masque spatial, en effectuant une transformation par une non-linéarité sigmoïde. On effectue ensuite la moyenne locale des sorties du transducteur en les convoluant par un noyau en forme de disque. On obtient ainsi une carte de différence juste remarquable pour chaque canal. Les informations des 28 canaux sont enfin fusionnées en calculant pour chaque point la distance entre les deux vecteurs à 28 dimensions représentant les différences juste remarquables sur les 28 canaux pour les deux images. On obtient ainsi la carte des différences entre les deux images.

Les performances des modèles de Daly et de Sarnoff ont été comparées par Li et al. [Li-98]. Les deux modèles permettent de détecter avec une bonne précision la position des distorsions

dans les images. Le modèle de Sarnoff est plus robuste que celui de Daly et nécessite moins de recalibration. Cependant, son nombre limité de bandes d'orientation fait qu'il peut ne pas détecter certaines erreurs quand les fréquences à détecter se situent entre deux des bandes disponibles. Le modèle de Sarnoff a un temps de calcul plus réduit que celui de Daly mais nécessite davantage de mémoire.

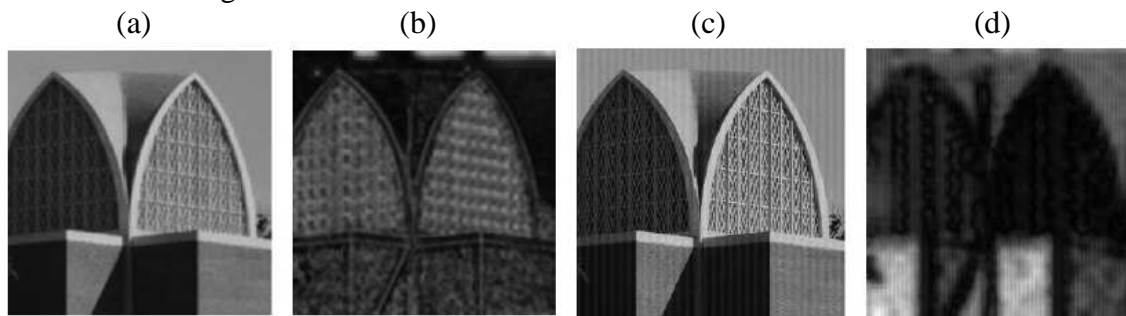


Fig. 4 : Application du modèle de Sarnoff sur deux images bruitées. Source : [Li-98] . De gauche à droite : (a) image bruitée par un flou, (b) carte de différence VDM pour le flou, (c) image bruitée par un bruit sinusoïdal, (d) carte de différence VDM pour le bruit sinusoïdal.

2.7 Conclusion

De nombreuses méthodes ont été développées afin d'évaluer la qualité d'une image compressée. Les méthodes subjectives, basées sur l'évaluation humaine des images, ne sont pas toujours utilisables dans la pratique, cependant elles sont utiles afin d'établir une référence pour l'évaluation des autres méthodes. Les méthodes univariées ont l'avantage de ne pas nécessiter de disposer de l'image originale, cependant elles sont basées sur une connaissance a priori du type de dégradation subie par l'image, elles ne sont donc pas adaptables à tous les cas. Les critères numériques comme les L_p -normes, le PSNR ou les méthodes statistiques mesurent la fidélité de l'image compressée par rapport à l'image originale, en revanche, comme ils ne donnent qu'une mesure globale de la différence entre les deux images, ils ne permettent pas d'estimer comment la dégradation sera perçue par l'oeil humain. Ils présentent par contre l'avantage d'être indépendants du contenu de l'image et du type de compression utilisée. Les critères basés sur un modèle de la vision humaine permettent de prendre en compte le fait que l'oeil humain n'est pas sensible de la même manière à toutes les dégradations, ils donnent donc des résultats plus conformes à la perception visuelle humaine. En revanche, ils sont plus complexes à calculer et sont souvent optimisés pour un type particulier de compression, ils ne sont donc pas efficaces sur toutes les images. Les méthodes graphiques, quant à elles sont surtout utiles pour analyser la position ou le type des dégradations dans l'image.

En définitive, les différentes méthodes d'évaluation de la qualité de la compression ont chacune leurs avantages et leurs inconvénients, et surtout leurs applications ne sont pas les mêmes, certaines méthodes étant mieux adaptées que d'autres à certains types de compression.

3 Mesure de la qualité de compression à l'aide de la loi de Zipf

On peut envisager d'utiliser la loi de Zipf afin de décrire la qualité de la compression d'une image. En effet, la loi de Zipf permet de décrire la structuration du contenu de l'image, on peut donc imaginer de l'utiliser pour mettre en évidence les distorsions de cette structure provoquées par l'application d'une compression avec pertes.

L'allure de la courbe de Zipf varie selon le type de l'image considérée, et il n'y a a priori pas de relation directe entre la qualité perçue d'une image et la forme de sa courbe de Zipf, c'est pourquoi il semble difficile d'utiliser la courbe de Zipf de l'image en tant que critère univarié, c'est à dire de mesurer la qualité de l'image en utilisant uniquement la courbe de Zipf de l'image compressée, sans utiliser d'image de référence.

C'est pourquoi on va ici définir des critères bivariés pour mesurer la qualité d'une image compressée à l'aide de la courbe de Zipf. Ces critères seront utilisés pour comparer la courbe de Zipf de l'image compressée à celle de l'image originale. On aura donc besoin, pour utiliser ces critères, de calculer les courbes associées aux deux images. Nous allons d'abord mettre en évidence les modifications de la courbe de Zipf introduites par la compression, avant de définir des critères permettant d'évaluer ces distorsions sous forme de valeurs numériques. On s'intéressera dans un premier temps essentiellement au cas d'une compression de type JPEG, qui est le standard le plus employé actuellement, avant de s'étendre au cas d'autres méthodes de compression comme SPIHT. Enfin, nous présenterons une méthode permettant de modifier une image afin de la rendre plus facile à compresser grâce aux propriétés de la courbe de Zipf.

3.1 Mise en évidence des distorsions introduites par la compression

Une image compressée diffère essentiellement de l'image originale par ses détails fins, qui sont altérés par la compression. Les zones homogènes de l'image sont peu modifiées, du moins tant que la compression reste faible. C'est pourquoi on utilisera pour l'évaluation de la qualité de la compression un codage qui permet de mettre en évidence les faibles variations de niveaux de gris à l'intérieur des motifs. On utilisera donc le codage des rangs généraux.

Avant de définir un critère de qualité basé sur la courbe de Zipf, il convient de vérifier que la compression de l'image a un effet visible sur la courbe de Zipf. C'est pourquoi des essais préliminaires ont été effectués sur des images avant et après l'application d'une compression. La figure 5 montre un exemple de résultat sur une image compressée en JPEG avec un facteur de qualité de 20%.

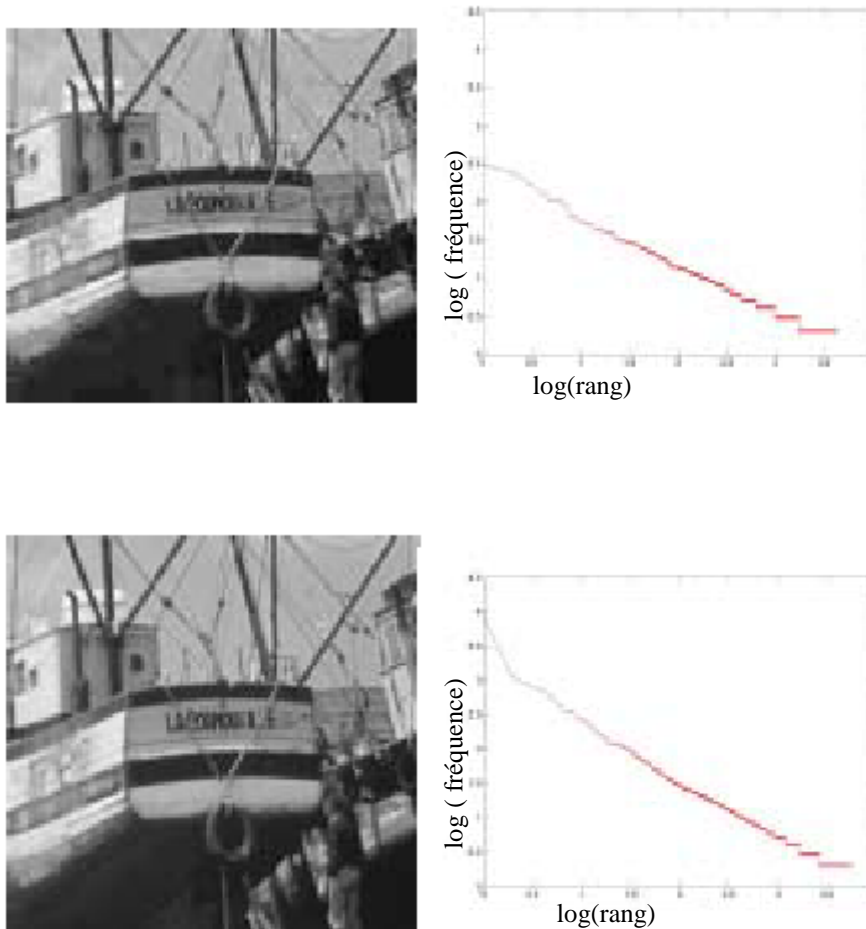


Figure 5 : Courbe de Zipf d'une image originale non compressée (en haut) et de l'image compressée en JPEG (en bas)

On remarque que les courbes de Zipf associées aux deux images sont différentes, surtout dans leur partie gauche. Sur la courbe obtenue avec l'image compressée, l'ordonnée à l'origine est beaucoup plus élevée, ce qui traduit le fait que la compression a rendu l'image plus uniforme. En effet, l'image compressée comprend moins de motifs différents que l'image originale, du fait que la compression a supprimé certains détails fins de l'image. Comme l'image compressée est de même taille que l'image originale, cela veut dire que les motifs les plus fréquents de l'image apparaissent proportionnellement beaucoup plus fréquemment dans l'image compressée. La courbe de Zipf d'une image compressée, codée avec les rangs généraux, est donc différente de l'image originale. Nous allons donc pouvoir maintenant définir des critères numériques afin de quantifier cette différence entre les courbes de Zipf associées aux deux images.

3.2 Utilisation de la distance euclidienne

Un premier critère utilisable pour évaluer la qualité de compression d'une image est la distance euclidienne entre la courbe de Zipf associée à l'image originale et celle associée à l'image compressée. Cette distance s'exprime comme la moyenne des différences entre la fréquence d'apparition des motifs de même rang dans l'image originale I et celle

correspondant dans l'image compressée I'. Comme les deux images n'ont pas le même nombre de motifs différents, les deux courbes n'ont pas la même longueur, cette moyenne est calculée sur le minimum des longueurs des deux courbes. La distance euclidienne est calculée au moyen de la formule suivante :

$$DE(I, I') = \frac{1}{\min(T, T')} \sqrt{\sum_{i=1}^{\min(T, T')} (N_{\sigma(i)} - N'_{\sigma(i)})^2} \quad (3-2-1)$$

Dans cette formule, $N_{\sigma(i)}$ et $N'_{\sigma(i)}$ représentent respectivement les fréquences d'apparition du motif de rang i dans l'image originale et dans l'image compressée et T et T' représentent les longueurs des courbes, c'est à dire le nombre de motifs apparaissant plus d'une fois. La figure ci-dessous montre un exemple de résultat obtenu avec cette méthode sur une image compressée en JPEG avec différents facteurs de qualité :

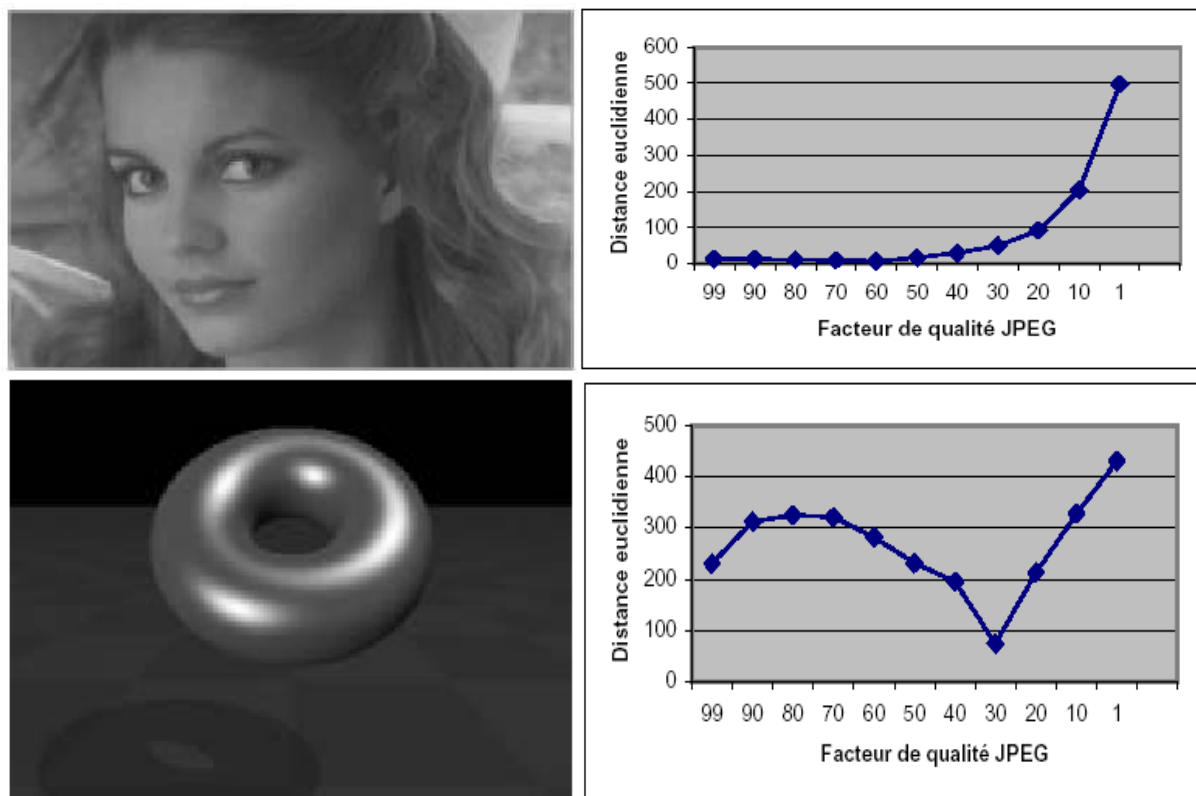


Fig. 6 : Mesure de la distance euclidienne entre les deux courbes de Zipf sur des images JPEG

Sur l'image du haut, on voit que la distance euclidienne est très faible pour un faible taux de compression, pour un facteur de qualité supérieur à 30, et augmente considérablement ensuite. Or, sur cette image, la dégradation devient visible pour un facteur de qualité inférieur à 30. La distance euclidienne entre les deux courbes de Zipf donne ici une bonne évaluation de la qualité de la compression. En revanche, ce n'est pas le cas pour l'image du bas où la variation de la distance euclidienne entre les deux courbes de Zipf n'est pas monotone, alors que la dégradation visuelle augmente avec la compression. La distance euclidienne entre les deux courbes n'est donc pas toujours un critère fiable pour mesurer la distorsion d'une image

compressée par rapport à l'original. C'est pourquoi nous allons utiliser d'autres critères plus élaborés afin d'évaluer la qualité d'une image compressée à l'aide de la courbe de Zipf.

3.3 Le critère ZQ (Zipf Quality)

Nous allons maintenant définir un nouveau critère d'évaluation de la qualité de la compression qui prend davantage de paramètres en compte que la simple distance entre les deux courbes. Ce critère sera appelé **ZQ** (Zipf Quality). Ce critère sera calculé à partir des courbes de Zipf de l'image originale et de l'image compressée et il sera défini de manière à ce qu'il soit nul si les deux images sont identiques et à augmenter si la distorsion entre les deux images augmente. Le ZQ sera défini comme une somme pondérée des valeurs absolues des différences entre les fréquences d'apparition des motifs de l'image originale I et de l'image compressée I'. Il est calculé à partir des courbes normalisées, c'est à dire que toutes les valeurs des fréquences d'apparition des motifs sont divisées par la fréquence d'apparition du motif le plus fréquent. Le ZQ est défini par la formule suivante :

$$ZQ(I, I') = -\frac{2 \sum_{i=1}^M (M-i+1) \cdot |\log_{10}(N_{\sigma(i)}) - \log_{10}(N'_{\sigma(i)})|}{M^2 \log_{10}(N_{\sigma(M)})} \quad (3-3-1)$$

Dans cette formule, $N_{\sigma(i)}$ représente la fréquence d'apparition du motif de rang i dans l'image I et $N'_{\sigma(i)}$ la fréquence d'apparition de ce même rang i dans l'image compressée I'. M représente le nombre de motifs considérés. Comme les différences se concentrent dans la partie gauche de la courbe, on peut se limiter aux premiers motifs de celle-ci. Une valeur de M égale à 40 peut convenir comme mesure dans la plupart des cas et suffit pour avoir une bonne idée de la qualité de l'image. Ce paramètre mesure globalement l'écart entre les courbes de l'image I et de l'image I'. Il est d'autant plus élevé que la distorsion de l'image I' est importante et il est nul si les deux images sont identiques. Le terme (M-i+1) permet d'appliquer un poids fort aux premiers points de la courbe et le terme $2/M^2$ permet de compenser les poids. Le ZQ est calculé en utilisant les valeurs normalisées des fréquences des motifs, c'est à dire que toutes les fréquences sont divisées par la fréquence d'apparition du premier motif. Le ZQ est nul si les deux images sont identiques, et il augmente au fur et à mesure que la distorsion de l'image compressée par rapport à l'image originale augmente.

Ce critère a été testé sur des images compressées en JPEG avec des valeurs différentes du facteur de qualité variant de 99% à 1%. Afin d'évaluer les résultats obtenus avec le nouveau critère de qualité, il est utile de pouvoir les comparer avec un critère de référence, c'est pourquoi les résultats obtenus avec le critère de qualité ZQ ont été comparés avec ceux obtenus avec un critère plus classique, à savoir le PSNR. La formule du PSNR est rappelée ci-dessous :

$$PSNR = 10 \cdot \log_{10} \left(\frac{255^2}{RMS} \right) \quad (3-3-2)$$

$$\text{avec } RMS = \frac{1}{N} \sum_{i=1}^N (a_i - b_i)^2 \quad (3-3-3)$$

Dans cette formule, a et b représentent les niveaux de gris du pixel d'indice i respectivement dans l'image originale I et dans l'image compressée I', et N représente le nombre de pixels dans l'image.

La figure ci-dessous montre un exemple de l'évolution du ZQ en fonction du taux de compression sur une image compressée en JPEG. Le taux de compression est ici défini de la manière suivante : Taux de compression = 100- (Facteur de qualité)
 On a représenté sur le même graphe le ZQ et le PSNR. Contrairement au ZQ, le PSNR diminue avec la dégradation de l'image par rapport à l'original.

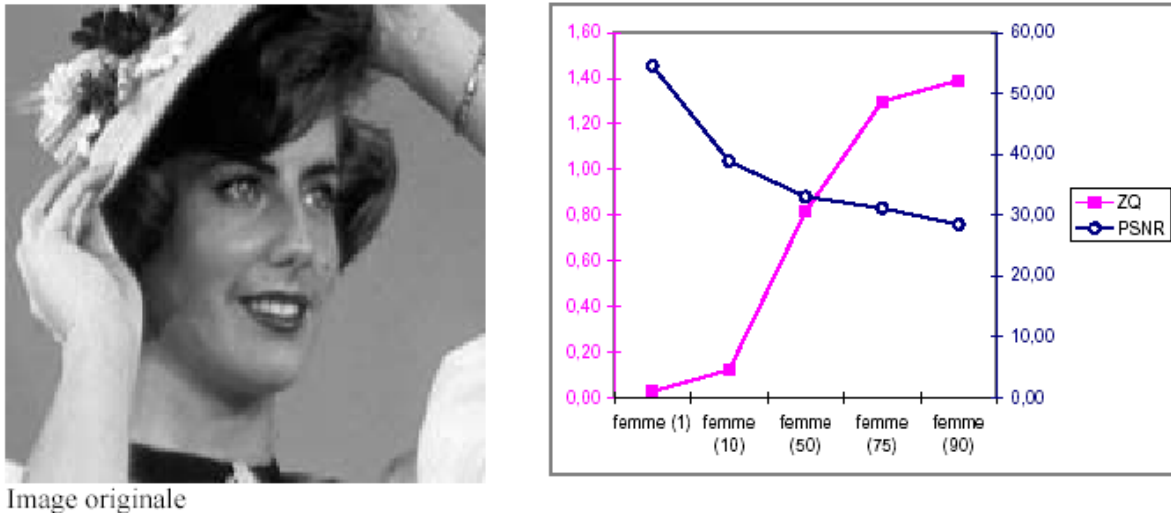


Fig.7 : Evolution du ZQ et du PSNR en fonction du taux de compression



Fig. 8 : Image de la Fig. 6 avec différents taux de compression

Sur ces images, on remarque que le ZQ est très faible pour les deux premières images, et augmente considérablement pour les suivantes. Or, si l'on s'intéresse aux images compressées, on remarque que sur les deux premières images, la dégradation est pratiquement invisible à l'œil nu. La dégradation devient visible à partir de la troisième image pour devenir réellement importante sur les deux dernières. Or, c'est ce que montre la courbe où l'on a un ZQ proche de 0 pour les deux premières images et qui augmente ensuite considérablement. En comparaison, la courbe du PSNR ne présente pas de discontinuité nette au moment où la compression devient visible. On peut donc dire que le ZQ donne une meilleure évaluation de la perte de qualité de l'image introduite par la compression, en ce sens que la mesure donnée par le ZQ correspond mieux à la perception visuelle de la qualité de l'image. Si l'on représente sur un même graphe les courbes de Zipf associées à chaque image, on remarque que ces courbes peuvent être classées en deux catégories comme le montre la figure 9. La première classe regroupe les courbes dont la distorsion est la plus faible, c'est à dire celles associées à des images qui sont les plus proches de l'image d'origine, alors que la deuxième

classe comprend les courbes dont la distorsion est la plus importante. Ces courbes ont une ordonnée à l'origine nettement plus élevée que celle des courbes de la première classe, ce qui traduit le fait que la compression uniformise l'image et augmente par conséquent la fréquence des motifs les plus nombreux dans l'image.

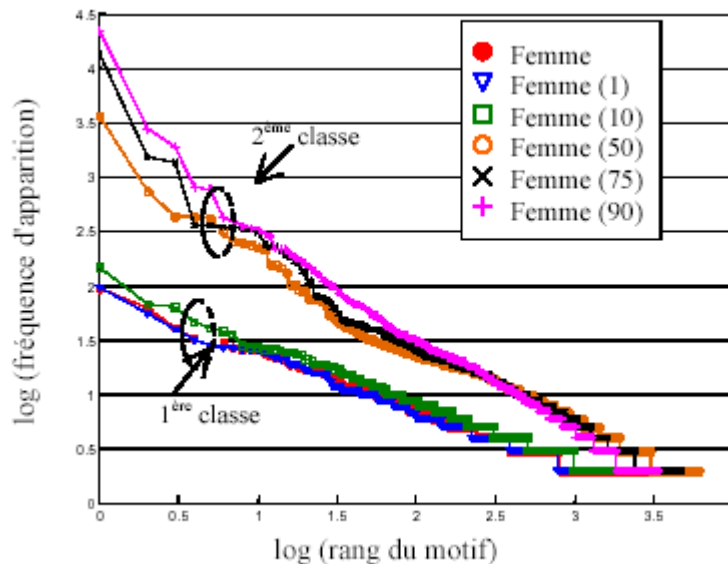


Fig.9 : Courbes de Zipf obtenues avec l'image de la figure 7 pour différents taux de compression.

Cependant, cette formulation du critère de qualité ZQ ne donne pas de résultats satisfaisants sur toutes les images. En effet, pour certaines images comme celle de la figure 6, la variation du ZQ en fonction du taux de compression n'est pas monotone et les images les plus compressées ont un ZQ plus petit que les images moins compressées. On peut remarquer que les valeurs du ZQ obtenues dans ce cas sont beaucoup plus faibles que dans le cas précédent. Or, ce résultat ne correspond pas du tout à la réalité, en effet les images qui présentent un ZQ plus faible que les images les moins compressées présentent en fait une distorsion plus importante que celles-ci, et cette dégradation est perceptible visuellement.

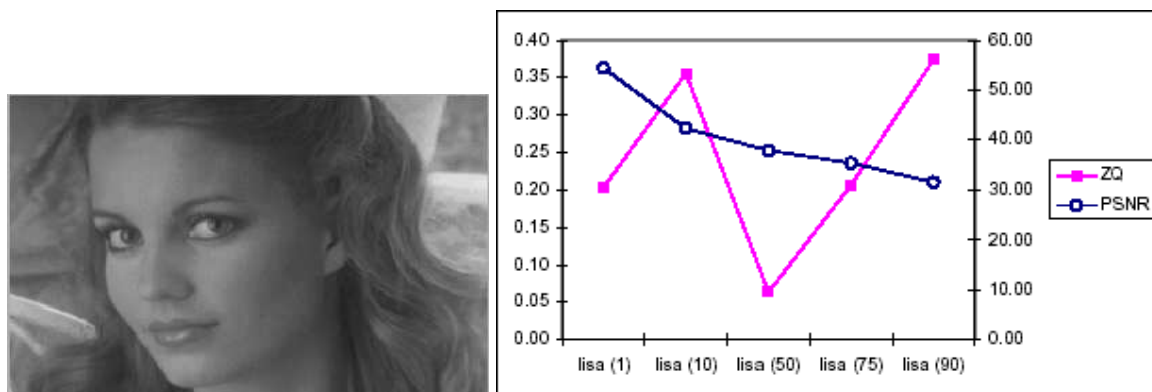


Fig. 10 : Cas d'une image où le ZQ n'est pas monotone

Dans l'ensemble, la formule du ZQ utilisée ici ne donne des résultats vraiment satisfaisants que dans environ 50% des cas. Cela est dû au fait que la formule ne prend pas en compte suffisamment de paramètres pour évaluer correctement la distorsion de l'image compressée par rapport à l'image originale. C'est pourquoi on va définir une nouvelle formule pour le ZQ qui prend en compte d'autres paramètres supplémentaires afin de mieux caractériser la distorsion de l'image compressée par rapport à l'image originale.

3.4 Amélioration du ZQ

Afin de mieux caractériser la qualité de l'image compressée, différents paramètres qui n'étaient pas pris en compte auparavant ont été inclus dans la formule du ZQ. L'un de ces paramètres est le nombre T de motifs qui apparaissent plusieurs fois dans la courbe, et qui représentent donc la longueur de la courbe de Zipf. Jusqu'à présent, cette information n'était pas prise en compte car on ne se préoccupait que des 40 premiers motifs, toute l'information sur la partie droite de la courbe était alors perdue. Or, cette information est importante pour caractériser l'uniformité, et l'on sait que l'un des effets de la compression est d'uniformiser l'image. De même, on inclura le nombre total L de motifs différents. En effet, lorsque les courbes de Zipf sont calculées, seuls les motifs qui apparaissent plus d'une fois dans l'image sont représentés sur la courbe. Or, le nombre total de motifs différents peut être une information utile pour l'évaluation de la qualité de l'image. On prendra aussi en compte la pente de la droite de régression des courbes de Zipf. En effet, il paraît normal de prendre en compte ce paramètre puisqu'il représente l'exposant de la loi puissance. Enfin, on prendra en compte le rapport B des ordonnées à l'origine des deux courbes. Le nouveau critère de qualité ZQ_1 est donné par la formule suivante :

$$ZQ_1(I,I') = \frac{1}{M} \cdot \frac{\sum_{i=1}^M |\log_{10}(N_{\sigma(i)}) - \log_{10}(N'_{\sigma(i)})|}{\log_{10}(N_{\sigma(M)})} \cdot \frac{T/L}{T'/L'} \cdot (P - P') \cdot B \quad (3-4-1)$$

A l'exception du rapport des ordonnées à l'origine, les valeurs de tous les paramètres de la formule sont calculés à partir des valeurs normalisées des courbes de Zipf.

Des essais ont été effectués sur les mêmes images que précédemment avec la nouvelle formule du ZQ_1 . Ici aussi, les résultats obtenus ont été comparés avec le PSNR. La figure ci-dessous montre quelques exemples de résultats obtenus sur des images compressées en JPEG.

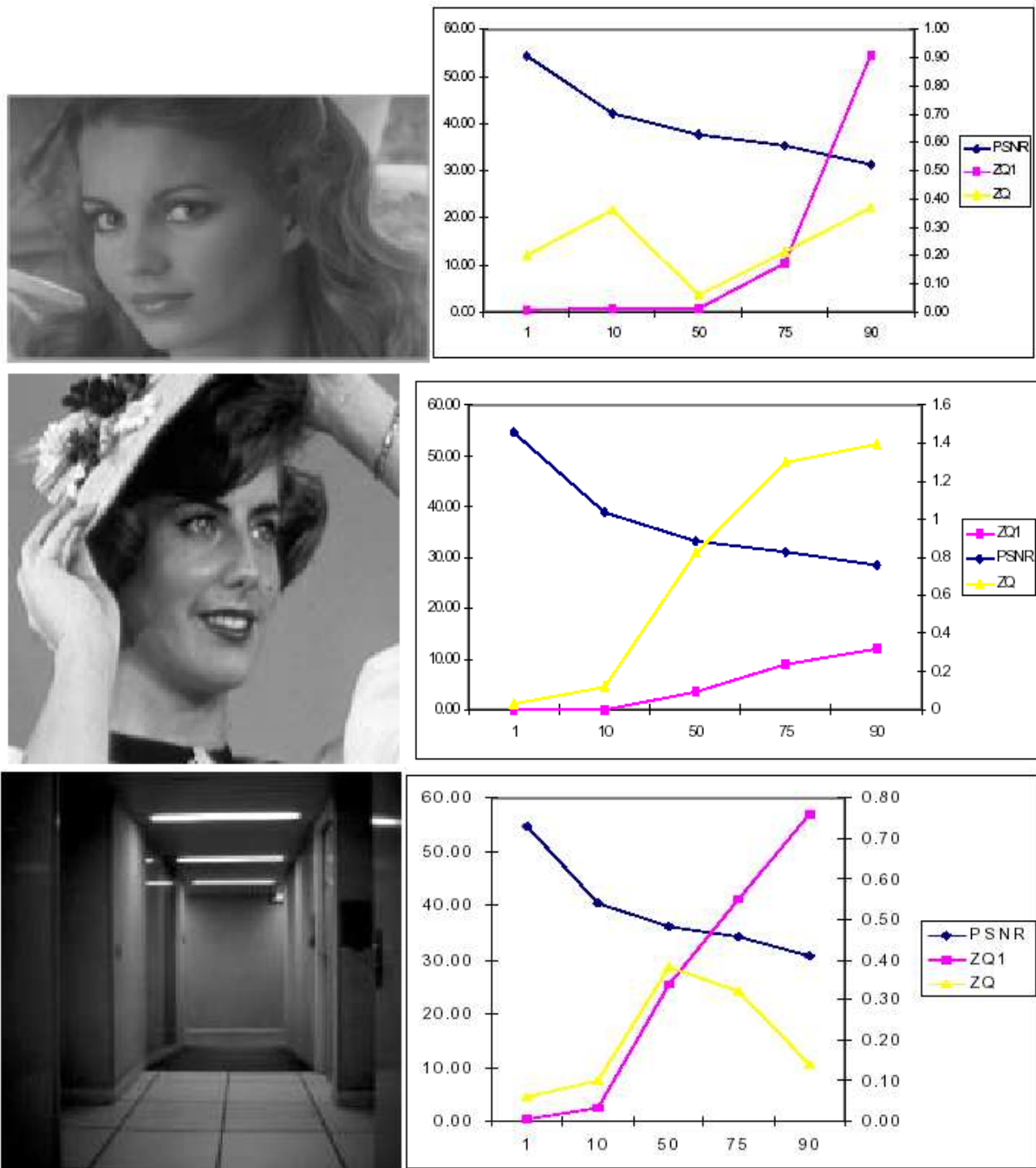


Fig. 11 : Exemples de résultats obtenus avec le nouveau critère ZQ_1

On remarque ici que la variation du PSNR en fonction du taux de compression est quasiment linéaire, alors que la variation du ZQ_1 présente une cassure très nette, le ZQ_1 étant proche de 0 pour un facteur de qualité inférieur à un certain seuil et augmentant fortement ensuite. Or, l'évaluation visuelle des images montre que les images pour lesquelles le ZQ_1 a une valeur proche de 0 ne présentent pas de distorsion visible à l'œil nu, et que la perte de qualité devient visible seulement à partir du moment où le ZQ_1 s'éloigne de zéro. Ce résultat s'observe sur l'ensemble des images testées. De ce fait, on peut dire que le ZQ_1 donne une évaluation de la qualité de la compression qui est conforme à la perception visuelle humaine.

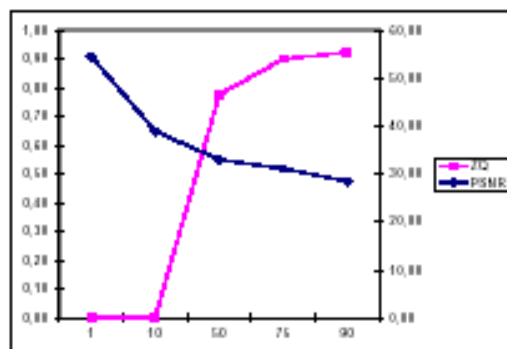
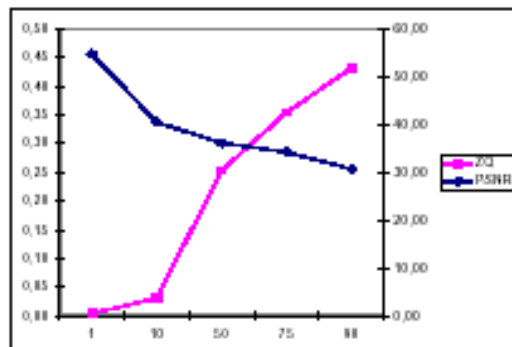
Les paramètres qui ont l'influence la plus grande sur la valeur du ZQ_1 sont principalement la somme des différences des fréquences des motifs ainsi que la valeur normalisée de la fréquence $N_{\sigma(M)}$ du dernier motif pris en compte pour le calcul du ZQ_1 .

La valeur du ZQ_1 obtenue pour une image donnée dépend de l'allure de la courbe de Zipf de l'image d'origine, et par conséquent de l'uniformité de l'image. Dans le cas d'une image très uniforme, comme c'est le cas pour les images de synthèse, la valeur du ZQ_1 reste petite car la somme des fréquences des M motifs les plus fréquents sur l'image d'origine est grande, par conséquent la somme des différences des fréquences des motifs entre l'image originale et l'image compressée sera faible. Il en sera de même pour les images qui comportent des zones localement uniformes. A l'inverse, si l'image originale est peu uniforme et ne comporte que peu de motifs qui apparaissent plusieurs fois dans l'image la valeur du ZQ_1 pour un même taux de compression sera plus grande, car la différence entre les fréquences des motifs sur l'image compressée et sur l'image originale sera plus importante.

Le fait que la plage de variation du ZQ_1 soit différente d'une image à l'autre peut poser problème lorsque l'on veut comparer des résultats d'une image à l'autre. Il est difficile de déterminer un seuil fixe valable pour toutes les images à partir duquel l'image peut être considérée comme dégradée. Afin de remédier à ce problème, on peut utiliser une valeur normalisée du ZQ_1 , appelée ZQ' qui est définie de la manière suivante :

$$ZQ' = \frac{ZQ_1}{1 + ZQ_1} \quad (3-4-2)$$

Contrairement au ZQ_1 qui peut varier entre 0 et $+\infty$, le ZQ' ne varie plus qu'entre 0 et 1, ce qui facilite la comparaison des résultats d'une image à l'autre. La figure 12 montre les courbes de variation du ZQ' pour les images de test de la figure précédente.



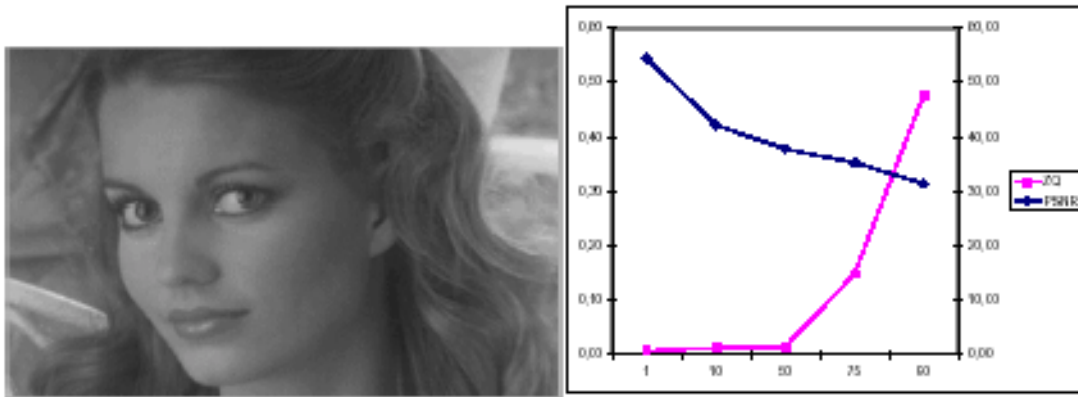


Fig. 12 : Evaluation du ZQ' sur des images compressées en JPEG

3.5 Utilisation du ZQ avec la compression SPIHT

Maintenant que nous avons vu que le ZQ donne de bons résultats avec une compression JPEG, nous allons voir ce qu'il en est avec d'autres types de compression. Nous avons donc appliqué la formule du ZQ définie au paragraphe précédent sur des images compressées avec la méthode SPIHT. Le paramètre de réglage de la méthode SPIHT est le taux de compression (bit-rate) exprimé en bits par pixel. Les essais ont été effectués avec des valeurs du taux de compression comprises entre 0,1 et 1 bit par pixel. La figure 13 montre une image compressée avec SPIHT pour un bit-rate de 0,2 bit par pixel.



Image originale



Image compressée

Fig. 13 : Exemple d'image compressée avec SPIHT

Les effets de la compression SPIHT sont différents de ceux de la compression JPEG. En effet, alors que JPEG introduit un effet de mosaïque, la compression SPIHT introduit un flou sur l'image. Les paramètres de réglage des deux méthodes de compression étant différents, afin de comparer les résultats obtenus, le ZQ a été calculé en fonction du taux de compression réel de l'image, défini comme étant le rapport entre la taille du fichier image original au format .TIF et la taille du fichier compressé. La figure 14 montre le résultat de l'évaluation du ZQ sur l'image de la figure 13 avec les deux méthodes.

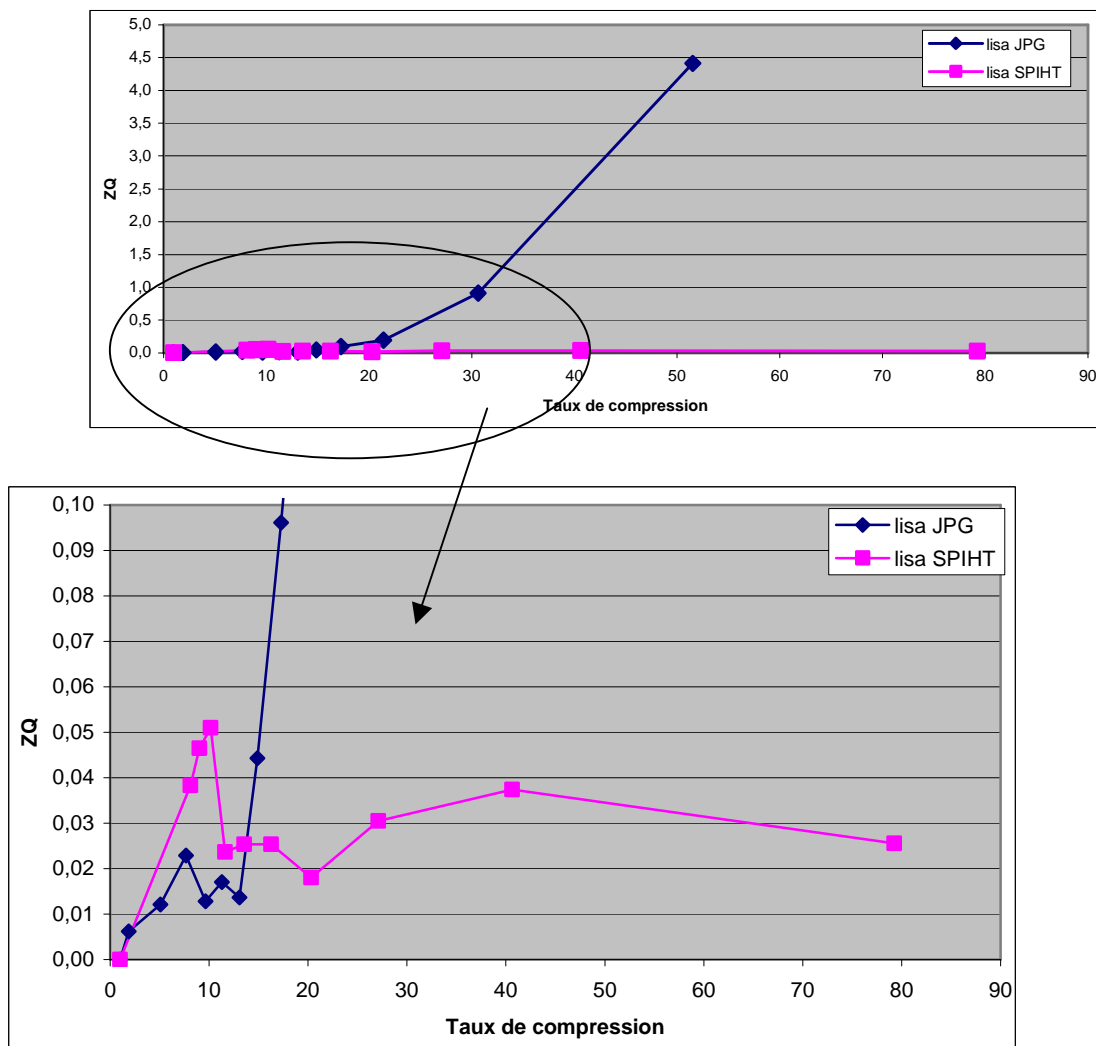


Fig. 14 : Résultat de l'évaluation du ZQ sur une même image compressée au format JPEG et SPIHT selon la compression. En (b) un zoom sur la zone entourée sur le graphe (a).

On remarque ici que la méthode SPIHT permet d'atteindre un meilleur taux de compression que la méthode JPEG et que le ZQ reste très faible pour la méthode SPIHT. C'est pourquoi on a agrandi la courbe sur la figure du bas afin de mieux observer ce qui se passe pour de faibles valeurs du ZQ. On observe que le ZQ de la méthode SPIHT n'évolue pas de manière croissante comme c'est le cas quand on considère la méthode JPEG. De plus, même si la valeur du ZQ reste faible pour un fort taux de compression avec SPIHT, l'image apparaît visuellement nettement dégradée. Le ZQ tel qu'il a été défini précédemment ne donne donc pas une évaluation correcte de la qualité d'une image compressée avec SPIHT. C'est pourquoi il est nécessaire de modifier la formule du ZQ afin de pouvoir l'utiliser de manière fiable avec la méthode SPIHT. Afin de déterminer quels sont les paramètres à modifier, on a effectué une étude des variations des différents paramètres de la formule du ZQ sur les images compressées en SPIHT. On a ainsi pu déterminer que deux de ces paramètres n'évoluent pas de manière croissante, à savoir l'aire comprise entre les deux courbes et la pente de régression des courbes de Zipf. En ce qui concerne la surface entre les deux courbes, le problème vient en fait de la normalisation. On a en effet divisé la fréquence des motifs de chacune des deux courbes par la fréquence d'apparition du motif le plus fréquent de chaque courbe, ce qui

correspond à une translation de la courbe dans un graphe logarithmique. Or, l'amplitude de la translation est différente pour chacune des deux courbes. C'est pourquoi on a utilisé une normalisation différente. On divise la fréquence d'apparition de chacun des motifs dans les deux courbes, aussi bien celle de l'image originale que celle de l'image compressée, par la fréquence du motif le plus fréquent dans l'image originale. On a également supprimé de la formule la différence des pentes et le rapport des ordonnées à l'origine. Le nouveau critère de qualité ZQs s'exprime par la formule suivante :

$$ZQs(I, I') = \frac{1}{M} \cdot \frac{\sum_{i=1}^M |N_{s(i)} - N'_{s(i)}|}{N_{s(M)}} \cdot \frac{T/L}{T'/L'} \quad (3-5-1)$$

$$\text{avec } N'_{s(i)} = \frac{\log_{10}(N'_{\sigma}(i))}{\log_{10}(N'_{\sigma}(1))} \quad \text{et} \quad N_{s(i)} = \frac{\log_{10}(N_{\sigma}(i))}{\log_{10}(N_{\sigma}(1))} \quad (3-5-2)$$

La figure 15 montre l'évaluation du ZQs sur l'image de la figure 12 :

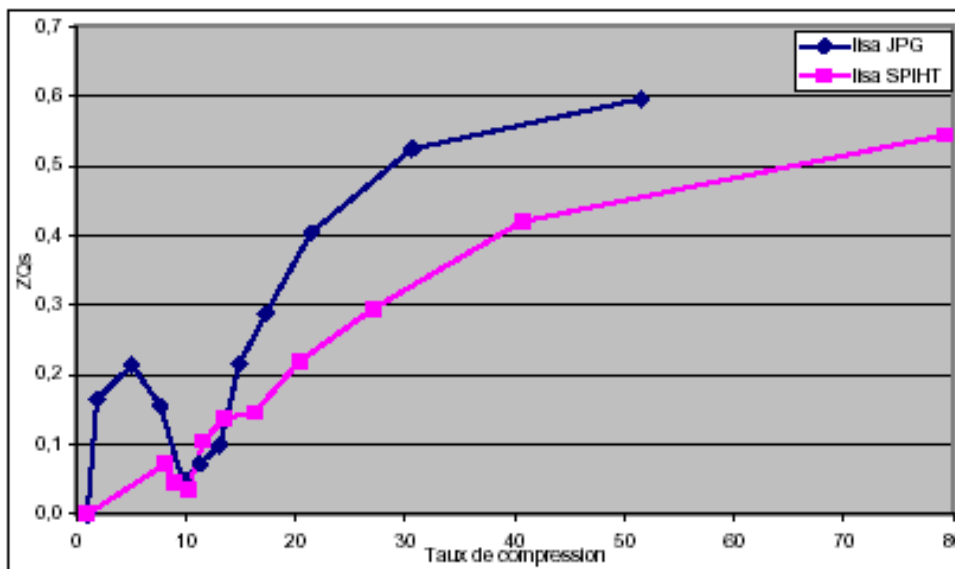


Fig. 15 : Evaluation du ZQs sur une image compressée à l'aide des deux méthodes

Sur cette figure, on remarque que le ZQs donne de meilleurs résultats que le ZQ pour l'évaluation d'une image compressée en SPIHT, en effet la valeur du ZQs est croissante avec l'augmentation du taux de compression de l'image, ce qui traduit bien l'impression visuelle. En revanche, le ZQs donne des résultats moins bons pour les images compressées en JPEG. En effet, la variation du ZQs en fonction du taux de compression n'est pas monotone dans le cas d'une compression JPEG. On peut donc dire que l'efficacité du critère d'évaluation de la qualité d'une image dépend de la méthode de compression utilisée, et que chaque critère est mieux adapté à une méthode particulière de compression. Ainsi, le critère ZQ, s'il est particulièrement efficace sur des images JPEG, n'est en revanche pas du tout adapté aux images SPIHT. En revanche, c'est l'inverse pour le ZQs. Cependant, le problème est moins grave car la zone dans laquelle le ZQs n'est pas croissant se situe dans le domaine des faibles taux de compression où la distorsion est peu visible.

3.6 Evaluation de la compression JPEG2000

Nous allons maintenant évaluer la qualité des images compressées avec la norme JPEG 2000. On rappelle que la norme JPEG2000 est une méthode de compression à base d'ondelettes, elle permet, pour une même taille de fichier, d'obtenir des images compressées de meilleure qualité que le JPEG. La norme JPEG2000 prévoit de nombreux paramètres de réglage, mais tous n'interviennent pas dans la qualité du résultat final. Les principaux paramètres qui interviennent dans la qualité de l'image sont le taux de compression (bit-rate) exprimé en bits par pixel, le nombre de niveaux de résolution lors de la restitution de l'image (C-layers) et le nombre de niveaux de décomposition en ondelettes (C-levels). En fait, la qualité de l'image compressée dépend essentiellement du bit-rate, les autres paramètres n'ayant qu'une influence très réduite en comparaison. La figure ci-dessous montre un exemple d'image compressée en JPEG2000 pour différentes valeurs du bit-rate.



Fig. 16: Image compressée en JPEG2000 avec différentes valeurs du taux de compression (bit-rate)

La dégradation des images compressées en JPEG2000 a été évaluée à l'aide du PSNR et du critère ZQ1 défini précédemment sur des images compressées avec différentes valeurs du bit-rate. La figure ci-dessous montre un résultat de ces essais sur une image de test.

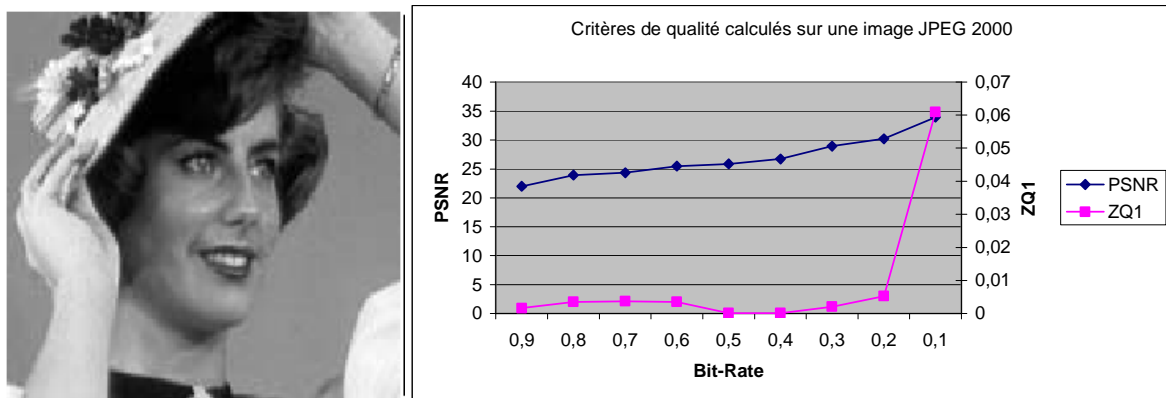


Fig. 17: Evaluation du ZQ1 et du PSNR sur une image JPEG 2000

Sur cette image, la dégradation devient visible pour un bit-rate inférieur à 0,4. On remarque que l'évolution du ZQ en fonction du bit-rate n'est pas monotone. La formule du ZQ1 n'est donc pas adaptée à la norme JPEG2000. En fait, ce défaut vient du fait que certains paramètres de la formule du ZQ1 n'évoluent pas de manière croissante avec le taux de compression : il s'agit de l'aire comprise entre les deux courbes et de la différence des pentes des droites de régression.

On va donc définir un autre critère qui ne prenne pas en compte les paramètres qui posent problème. Ce nouveau critère, que l'on appellera ZQ2, est défini par la formule :

$$ZQ2 = B \frac{\sum_{i=0}^M |\log_{10}(N_{\sigma(i)}) - \log_{10}(N_{\sigma(i)})|}{\log_{10}(N_{\sigma(M)})} \quad (3-6-1)$$

Dans cette formule, $N_{\sigma(i)}$ représente la fréquence du motif de rang i , M le nombre de motifs pris en compte, ici égal à 40 et B le rapport des ordonnées à l'origine des courbes de Zipf de l'image originale et de l'image modifiée. La figure ci-dessous montre le résultat de l'évaluation du nouveau critère sur l'image de la figure 17

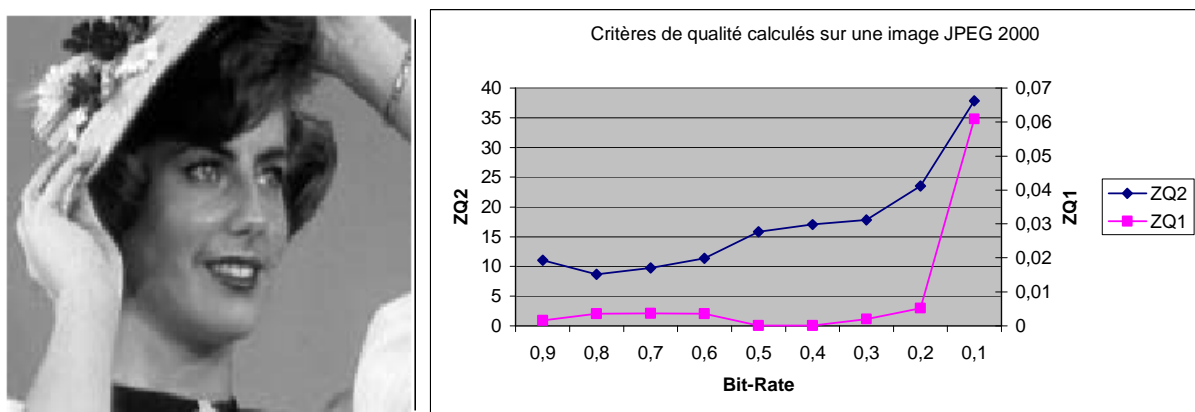


Fig. 18 : Comparaison du ZQ et du ZQ2 sur une image JPEG200

Le résultat obtenu avec ce nouveau critère est davantage conforme à la perception humaine. En effet, on observe une forte élévation du ZQ2 pour un bit-rate inférieur à 0,3, ce qui correspond au moment où la compression devient visible. Cependant, pour les valeurs élevées du bit-rate, la variation du ZQ n'est pas monotone.

En réalité, les problèmes rencontrés avec les critères basés sur la loi de Zipf viennent du fait que la compression JPEG2000 rend les textures de l'image plus uniformes. Or, les faibles variations de niveaux de gris introduites par cette uniformisation, si elles ne sont pas perceptibles à l'œil nu, sont en revanche détectées par le codage des rangs généraux.

Pour remédier à ce problème, on va utiliser le codage des rangs généraux avec seuil, ce qui permet de limiter l'influence des faibles variations de niveaux de gris. La figure ci-dessous montre un exemple de l'application du codage des rangs généraux avec un seuil de 10 sur l'image de la figure 18.

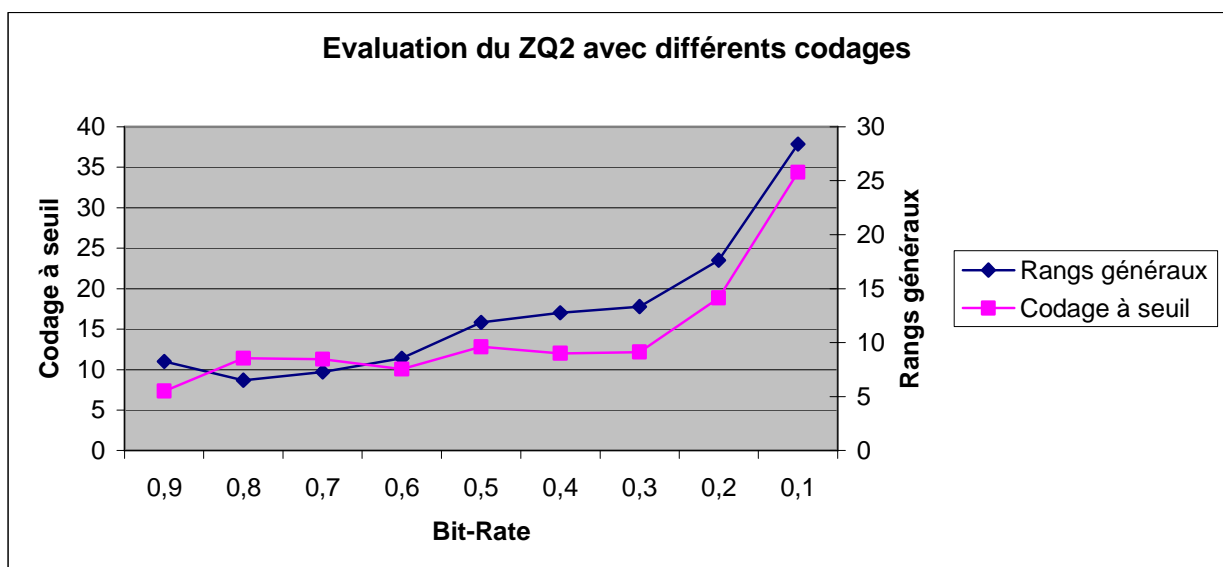


Fig. 19 : Evaluation du ZQ2 avec le codage des rangs généraux avec et sans seuil

Les résultats obtenus avec le codage à seuil sont assez peu différents de ceux obtenus avec les rangs généraux sans seuil, les deux courbes ayant une allure assez proche. L'augmentation de la variation du ZQ2 quand l'image est très dégradée est plus importante avec le codage à seuil, donc on peut dire que les résultats sont davantage en conformité avec l'impression humaine. Le codage à seuil présente en outre l'avantage d'être plus rapide à calculer. Cependant, les résultats obtenus avec le ZQ2 sur les images compressées avec JPEG2000 sont relativement moins bons que ceux obtenus avec le ZQ1 sur les images JPEG.

3.7 Conclusion

La loi de Zipf peut être utilisée afin d'évaluer la qualité d'une image compressée. Il est en effet possible, sur une courbe de Zipf codée avec les rangs généraux, de mettre en évidence les modifications de la structure de l'image introduites par la compression. On peut alors définir un critère bivarié qui permet d'évaluer la distorsion entre l'image originale et l'image compressée à partir des caractéristiques des courbes de Zipf. La mesure de la distance euclidienne peut être utilisée mais ne donne pas toujours des résultats satisfaisants. On a donc défini un critère de qualité ZQ qui prend en compte plusieurs caractéristiques des courbes de Zipf. Ce critère permet d'obtenir une évaluation de la qualité de la compression des images JPEG qui soit conforme à la perception visuelle humaine et qui donne de meilleurs résultats que le PSNR. Une normalisation de ce critère permet de comparer plus facilement les résultats obtenus d'une image à l'autre. En revanche, le même critère appliqué à des images compressées avec SPIHT donne des résultats nettement moins satisfaisants, tout comme le critère ZQs optimisé pour les images SPIHT donne de moins bons résultats si on l'applique à une image JPEG. Dans le cas des images JPEG2000, c'est encore un autre critère qui donne les meilleurs résultats, en utilisant un codage à seuil. Les critères d'évaluation de la qualité d'une image basés sur la qualité de la compression ne sont donc pas universels, et un critère efficace sur un type particulier de compression ne le sera plus sur un autre. Il faut donc définir le critère de qualité, en choisissant les caractéristiques de la courbe de Zipf que l'on va évaluer, en fonction du type de compression que l'on applique à l'image. Pour cela, on est généralement amené à procéder de manière empirique. Cela étant, il est possible d'obtenir de bons résultats avec ce critère, et notamment de déterminer le seuil à partir duquel la compression devient visible pour l'œil humain, alors même que ce critère n'est pas du tout basé sur un modèle de la vision humaine.

4 Optimisation de la compression

4.1 Principe de l'optimisation multicritères

La compression d'images a pour objectif de réduire le plus possible la taille du fichier image tout en préservant une qualité d'image suffisante pour que la dégradation de l'image ne soit pas trop visible. L'optimisation de la compression consiste donc à déterminer de manière automatique les meilleures valeurs des paramètres de compression, c'est à dire celles qui permettent d'obtenir le fichier le plus petit possible tout en préservant une bonne qualité de l'image. Il s'agit donc d'un problème d'optimisation multicritères : on cherche en effet à minimiser en même temps deux critères qui sont la taille du fichier et la dégradation de l'image. Ces deux critères étant contradictoires, l'objectif est alors de trouver une solution qui assure le meilleur compromis entre les deux.

La figure ci-dessous montre le principe de l'optimisation multicritères. Les deux fonctions objectifs à minimiser F_1 et F_2 sont représentés par l'abscisse et l'ordonnée du graphe, et la courbe représente l'ensemble des solutions possibles du problème en fonction d'un ou plusieurs paramètres. La recherche d'une solution optimale revient alors à rechercher le point de la courbe pour lequel la distance à l'origine d soit minimale.

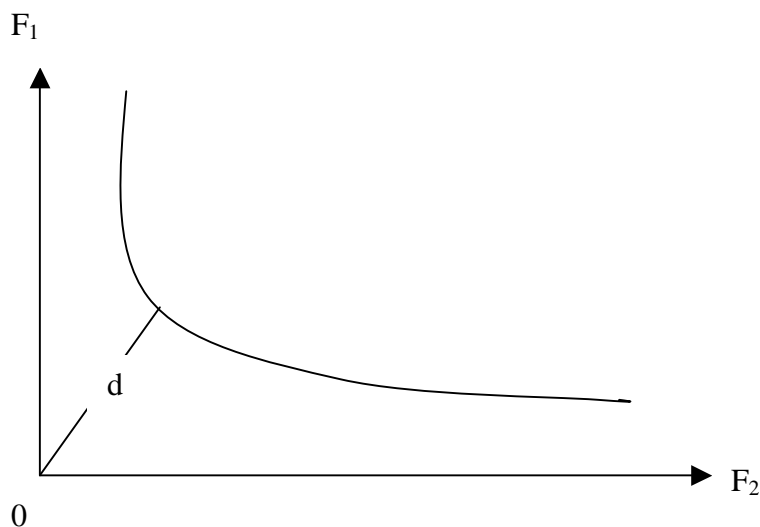


Fig. 20 : Représentation des solutions d'un problème d'optimisation multicritères

Dans notre cas, les deux fonctions à minimiser sont la taille de l'image compressée et la dégradation de l'image, mesurée par un critère numérique de qualité. Les paramètres à optimiser sont les paramètres de l'algorithme de compression utilisé. Dans le cas de la compression JPEG2000, il s'agit essentiellement du bit-rate, les autres paramètres ayant une influence négligeable sur le résultat.

Il est généralement difficile de calculer de manière exacte la solution optimale, c'est pourquoi on utilise des méthodes heuristiques comme les algorithmes génétiques, la méthode taboue ou

le recuit simulé. Ces méthodes ne garantissent pas de trouver à coup sûr la solution optimale, mais elles permettent d'obtenir une solution de bonne qualité dans un temps raisonnable.

Dans notre cas, nous avons choisi d'utiliser la méthode taboue. En effet, les résultats de la compression d'images dépendent essentiellement d'un seul paramètre, à savoir le bit-rate. L'utilisation d'un algorithme génétique serait alors assez mal adapté à ce problème. La méthode taboue a été expérimentée sur des images JPEG2000 avec deux critères de qualité : le ZQ2 et le critère statistique de Wang [Wan-02b] défini au paragraphe 2.4.

4.2 La méthode taboue

La méthode taboue est une méthode d'optimisation qui permet d'éviter les minima locaux de la fonction objectif. Le principe de la recherche taboue consiste à garder la trace du cheminement passé en mémoire et de s'y référer pour améliorer la recherche. La méthode consiste à se déplacer d'une solution vers une autre par observation du voisinage de la solution de départ et à définir des transformations taboues que l'on garde en mémoire. Une transformation taboue est une transformation que l'on s'interdit d'appliquer à la solution courante.

A partir de n'importe quelle configuration, la méthode taboue génère une séquence de configurations qui doit atteindre la configuration optimale. A chaque itération, le mécanisme qui permet de passer d'une configuration (x_n) à la configuration suivante (x_{n+1}) consiste à construire la liste des voisins de x_n , c'est à dire un ensemble de configurations de paramètres qui sont proches de x_n , puis à évaluer la fonction objectif f pour chacun des voisins de la liste. La configuration choisie x_{n+1} est celle qui a la valeur minimale de f parmi toutes les configurations de la liste. Cette configuration est acceptée même si $f(x_{n+1}) > f(x_n)$, cela permet à la méthode taboue d'éviter les minima locaux de f .

Pour éviter d'évaluer plusieurs fois les mêmes solutions, on construit une liste taboue qui contient les changements de configuration déjà effectués et que l'on s'interdit de faire. Cette liste, qui donne son nom à la méthode, contient les changements inverses ($x_{n+1} \rightarrow x_n$) des 7 derniers changements ($x_n \rightarrow x_{n+1}$). La recherche d'une nouvelle configuration est restreinte aux voisins de x_n qui peuvent être atteints sans utiliser les changements de la liste taboue.

Le processus peut être arrêté au bout d'un nombre fixé d'itérations sans amélioration de la solution.

Dans notre cas, la fonction objectif f sera définie par :

$$f = \sqrt{(taille)^2 + (ZQ2)^2} \quad (4-2-1)$$

où *taille* représente la taille du fichier image JPEG2000 exprimée en kilo-octets et ZQ2 est calculé en utilisant la formule (3-7-1) sur des courbes obtenues à l'aide du codage des rangs généraux avec un seuil de 10. On arrêtera le processus au bout de 100 itérations.

4.3 Résultats de l'optimisation

La méthode taboue à été essayée sur différentes images de test. Deux critères d'évaluation de la qualité de la compression ont été essayés, le ZQ2 et le critère statistique de Wang. On rappelle la formule du critère de Wang :

$$Q = \frac{\sigma_{ab}}{\sigma_a \sigma_b} \times \frac{2\bar{a}\bar{b}}{(\bar{a})^2 + (\bar{b})^2} \times \frac{2\sigma_a \sigma_b}{\sigma_a^2 + \sigma_b^2} \quad (4-3-1)$$

Où \bar{a} et \bar{b} sont les moyennes des niveaux de gris des deux images, σ_a et σ_b les écarts type et σ_{ab} la covariance des deux images.

La figure suivante montre le résultat optimal trouvé par la méthode taboue avec les deux critères sur une image de 256x256 pixels.



Fig. 21 : Exemple de résultat obtenu avec la méthode taboue sur une image de test

Sur cet exemple, le résultat optimal a été obtenu pour un bit-rate de 0,85 avec le critère de Wang et un bit-rate de 0,70 avec le ZQ2. La distorsion des images compressées est peu visible à l'œil nu dans les deux cas. La taille du fichier étant plus faible dans le cas du ZQ2 pour un résultat visuel comparable, on peut dire que le ZQ2 est un meilleur critère que le critère de Wang pour l'optimisation de la compression. Cependant, la solution trouvée n'est pas optimale, car sur cette image on peut compresser jusqu'à un bit-rate de 0,3 avant d'obtenir une dégradation visible.

En revanche, sur des images plus grandes comme celle de la figure ci-dessous, les résultats obtenus se rapprochent de la solution optimale. Sur cette image de taille 512x512, le meilleur résultat trouvé par la méthode taboue a été obtenu pour un bit-rate de 0,54 avec le critère de Wang et un bit-rate de 0,37 avec le ZQ2. On se rapproche davantage du résultat optimal qui est de 0,3 pour cette image.



Image originale
Format BMP, 768Ko

Résultat optimal avec le
critère de Wang , 19Ko

Résultat optimal avec le
ZQ2, 12Ko

Fig. 22: Résultats obtenus avec la méthode taboue sur une image de taille 512x512

Sur l'ensemble des images testées, le ZQ2 utilisé comme critère d'évaluation pour la méthode taboue a permis d'obtenir des images plus compressées, donc des fichiers images plus petits, pour une qualité visuelle comparable. Le ZQ2 est donc un meilleur critère d'évaluation de la compression que le critère de Wang. La méthode taboue donne toujours une image moins compressée que celle pour laquelle la compression devient visible, donc l'image obtenue est de bonne qualité. Les résultats dépendent de la taille de l'image, la solution donnée par la méthode taboue se rapproche de la solution optimale au fur et mesure que la taille de l'image augmente, ce qui veut dire que l'on perd moins de place sur les gros fichiers images que sur les petits. La taille minimale de l'image pour que le résultat obtenu par le tabou se rapproche de l'optimal est de l'ordre de 200000 pixels.

Conclusion

La compression d'images avec pertes introduit une dégradation de l'image que l'on est amené à vouloir quantifier. Pour cela, de nombreux critères très divers ont été proposés. Il est possible d'utiliser la loi de Zipf pour définir des critères bivariés de mesure de la dégradation de l'image, c'est à dire qui permettent de mesurer par une valeur numérique la perte de qualité de l'image compressée par rapport à l'image originale. Ces critères basés sur la loi de Zipf permettent de donner une information plus conforme à la perception humaine que des critères plus classiques comme le PSNR. En particulier, ils permettent de détecter facilement le taux de compression maximal à partir duquel la compression devient facilement perceptible à l'œil, et cela bien qu'ils soient basés sur un modèle totalement différent de celui de la vision humaine. Cependant, il n'existe pas de critère unique qui donne des résultats satisfaisants quel que soit le type de compression d'image utilisé, mais plusieurs critères adaptés chacun à un standard particulier de compression. Ces critères peuvent être utilisés pour l'optimisation automatique de la compression d'image, en tant que fonction objectif d'une méthode taboue. Nous allons maintenant, dans le chapitre suivant, traiter une autre application des modèles de lois puissance, à savoir la détection d'objets et de zones d'intéerêt dans les images.

Chapitre 4 : Détection d'objets et de zones d'intérêt

Introduction

Nous allons maintenant utiliser la loi de Zipf pour une autre application, à savoir la détection d'objets ou de zones d'intérêt dans une image. Nous avons vu au chapitre 2 que la répartition des fréquences d'apparition des motifs de l'image, modélisée par les lois de Zipf ou de Zipf inverse est différente selon le contenu structurel de celle-ci. Ceci devrait permettre de détecter une zone particulière de l'image qui se distingue par son caractère naturel ou artificiel ou par une quantité de détails différente de celle de l'arrière-plan. Nous allons donc définir et utiliser des méthodes de détection d'objets dans les images qui utilisent les modèles de lois puissances. Après avoir défini la notion de zone d'intérêt et présenté différentes méthodes de détection de zones d'intérêt existant dans la littérature, nous allons d'abord traiter ce problème au travers d'un cas particulier, celui de la détection d'un objet artificiel dans un milieu naturel. Pour cela, nous allons proposer une méthode basée sur la loi de Zipf qui doit permettre de détecter la présence et estimer la position dans l'image d'un objet artificiel. Ensuite, nous généraliserons cette approche à la détection de la zone d'intérêt d'une image quelconque, et nous définirons différentes méthodes de détection de zones d'intérêt basées sur lois puissance, que nous comparerons entre elles et avec des méthodes plus classiques.

1 Etat de l'art

Nous allons ici présenter différentes méthodes qui peuvent être utilisées pour détecter des régions d'intérêt dans une image. La détection automatisée de régions d'intérêt dans les images est un problème difficile qui a fait l'objet de nombreuses études, et conduisant à diverses méthodes. Avant de les présenter, nous allons d'abord définir la notion de région d'intérêt.

1.1 Notion de région d'intérêt

D'une manière générale, une région d'intérêt est une région de l'image qui intéresse particulièrement le regard de l'observateur. La notion de région d'intérêt se définit donc, au moins en partie, par rapport à un référent humain et comporte donc une part subjective.

Pour qu'il y ait région d'intérêt, il faut que l'on soit capable de distinguer une *région particulière de l'image* qui attire plus notre attention, nous paraisse plus intéressante que le reste de l'image. Pour cela il faut que cette région soit visuellement distincte (du moins aux yeux d'un observateur humain) du reste de l'image. Cependant, la réciproque n'est pas vraie, toute région visuellement distincte du reste de l'image n'est pas nécessairement région d'intérêt.

On désigne souvent par « fond » ou « arrière-plan » le reste de l'image n'appartenant pas à la zone d'intérêt. Cependant, dans la réalité cette zone de moindre intérêt n'est pas forcément située entièrement en arrière de ce qui est qualifié de zone d'intérêt, mais il est exact que dans

la plupart des images, les parties de l'image représentant ce qui est le plus éloigné du point d'observation appartiennent à cette catégorie.

Une zone d'intérêt ne peut pas englober toute l'image. Dans le cas d'une image où il n'est pas possible de déterminer une zone particulière sur laquelle se porte l'intérêt de l'observateur, il n'y a alors pas de zone d'intérêt, ce qui ne veut cependant pas dire, par ailleurs, que l'image elle-même n'est pas intéressante. A l'inverse, une zone d'intérêt, pour être significative, doit représenter une part suffisamment importante de l'image, elle ne peut pas, par exemple, être réduite à un simple point.

La notion de zone d'intérêt étant une notion subjective, pour une même image plusieurs personnes pourront définir, en fonction de leur sensibilité particulière ou de leurs centres d'intérêt, des régions d'intérêt différentes, comme c'est le cas pour les images de la figure 1. Ainsi, pour une image représentant un animal, certaines personnes choisiront comme région d'intérêt l'animal tout entier, et d'autres seulement la tête de celui-ci. Il en est de même pour l'image représentant une voiture devant une maison, certaines personnes choisiront la voiture, d'autres la maison, et d'autres encore les deux.



Fig.1 : Images où la zone d'intérêt choisie est différente selon les observateurs

Souvent, la région d'intérêt est constituée d'un objet facilement identifiable et reconnaissable par un être humain. Le sujet humain se sert alors de ses connaissances pour identifier l'objet. Le problème est qu'une machine ne dispose pas des connaissances et des référents culturels lui permettant d'identifier de cette manière la zone d'intérêt. Certes, il est possible d'utiliser des outils qui comportent une phase d'apprentissage, tels que les réseaux de neurones, pour simuler le comportement de l'être humain face à une image. Mais ces outils sont limités, ils ne peuvent être utilisés que sur un nombre relativement réduit de modèles et ne sont utilisables que si l'on a une connaissance a priori du type d'image ou d'objet à détecter. Il ne permettent pas de détecter n'importe quel objet ou zone d'intérêt dans n'importe quelle image.

C'est pourquoi nous allons plutôt nous intéresser à des méthodes perceptuelles ne reposant pas sur la reconnaissance d'un objet, elles se contentent de rechercher une zone de l'image

qui soit *saillante*, c'est à dire qui se distingue du reste de celle-ci par certaines caractéristiques visuelles de bas niveau. Différentes caractéristiques peuvent être utiles pour tenter d'identifier une région d'intérêt, comme la couleur, le contraste, les contours d'un objet, les formes, les textures, la netteté ou même la position d'un objet dans l'image.

Les méthodes utilisées pour la détection de région d'intérêt sont diverses. Certaines sont directement inspirées du fonctionnement du système visuel humain de bas niveau, alors que d'autres n'ont pas de fondement biologique et sont simplement basées sur une analyse de la structure de l'image. Nous indiquerons d'abord quelques études essayant de simuler le système visuel humain. Après avoir rappelé les différents éléments du système visuel humain et leur rôle ainsi que les facteurs auxquels nous sommes particulièrement sensibles nous présenterons différentes méthodes de détection de régions d'intérêt basées sur une modélisation de ces facteurs d'attention visuelle. Nous présenterons ensuite diverses méthodes basées sur l'analyse de différentes caractéristiques structurelles de l'image au moyen d'outils divers tels que les transformations de symétrie, l'entropie, la transformée en ondelettes, les orientations locales dominantes ou encore la dimension fractale

1.2 Le système visuel humain

Un certain nombre de méthodes de détection de zones d'intérêt sont fondées sur une modélisation du fonctionnement du système visuel humain. Le processus de fonctionnement de la vision humaine, ainsi que les structures cérébrales qui sont affectées à cet usage, se répartissent entre vision de haut et de bas niveau. La vision de bas niveau s'attache essentiellement à la *perception* de l'environnement visuel alors que la vision de haut niveau s'intéresse plutôt à l'*interprétation* de ce qui a été perçu. Sachant que notre problématique s'intéresse principalement à la *détection* de zones d'intérêt et non à leur identification, la vision de haut niveau intervient certainement peu ici. Avant d'aborder ces méthodes, il convient de présenter succinctement le fonctionnement du système visuel humain.

La vision de bas niveau met en œuvre différents éléments qui sont l'œil, le nerf optique, le corps genouillé latéral et le cortex visuel primaire qui est situé à l'arrière du cerveau. Pour une description plus détaillée du système visuel humain, on se référera à des ouvrages comme ceux de Marr [Mar-82], Imbert [Imb-83], ou Hubel [Hub-88].

L'image se forme sur la rétine qui est une membrane tapissant le fond de l'œil et qui est composée de plusieurs couches de cellules photosensibles : les *bâtonnets* et les *cônes*. Les bâtonnets sont responsables de la vision *scotopique*, c'est à dire de nuit, ils ne sont pas sensibles aux couleurs et permettent surtout de percevoir le mouvement. Les cônes sont de trois types, ils sont sensibles respectivement au rouge, au vert et au bleu. Ils permettent ainsi la vision des couleurs. Les cônes nécessitent pour fonctionner une intensité lumineuse élevée, c'est pourquoi ils ne sont efficaces que pour la vision en plein jour (vision *photopique*). Les cônes se concentrent dans la région centrale de la rétine, la *Fovea centralis* qui constitue la zone d'acuité visuelle maximale. Les bâtonnets, eux se répartissent plutôt dans la périphérie de la rétine et participent à la vision périphérique.

Les informations issues de la rétine sont ensuite acheminées par le *nerf optique* jusqu'au *corps genouillé latéral*. Les deux nerfs optiques se croisent au niveau du *chiasma optique*, l'influx visuel issu de l'oeil gauche étant analysé par l'hémisphère droit du cerveau et réciproquement. Le corps genouillé latéral est un ganglion nerveux dont le rôle est avant tout d'amplifier le signal issu du nerf optique, il joue essentiellement un rôle de transmission du

signal plutôt que de traitement de l'information. Après avoir traversé le corps genouillé latéral, le signal visuel atteint *le cortex visuel primaire* qui est le siège du traitement de bas niveau de l'information visuelle.

Le cortex visuel primaire est composé de plusieurs couches de neurones qui sont spécialisés dans différents traitements. Certains de ces neurones servent dans la détection de contours, ils sont regroupés par « colonnes d'orientation » qui sont spécialisées dans la détection de contour suivant une direction privilégiée. D'autres de ces neurones sont spécialisés dans la détection du contraste. Ces neurones fonctionnent suivant le principe du contraste *center-surround*, c'est à dire qu'ils sont sensibles aux différences des caractéristiques visuelles comme la luminosité ou la couleur entre le centre du champ visuel et la périphérie de celui-ci. Le cortex visuel primaire est également chargé de la vision stéréoscopique en reconstituant la troisième dimension à partir des images issues des deux yeux.

Il a été montré, notamment par Hubel [Hub-88] que les neurones du cortex visuel primaire peuvent être modélisés par des filtres de Gabor. La figure 2 montre le schéma du système visuel humain.

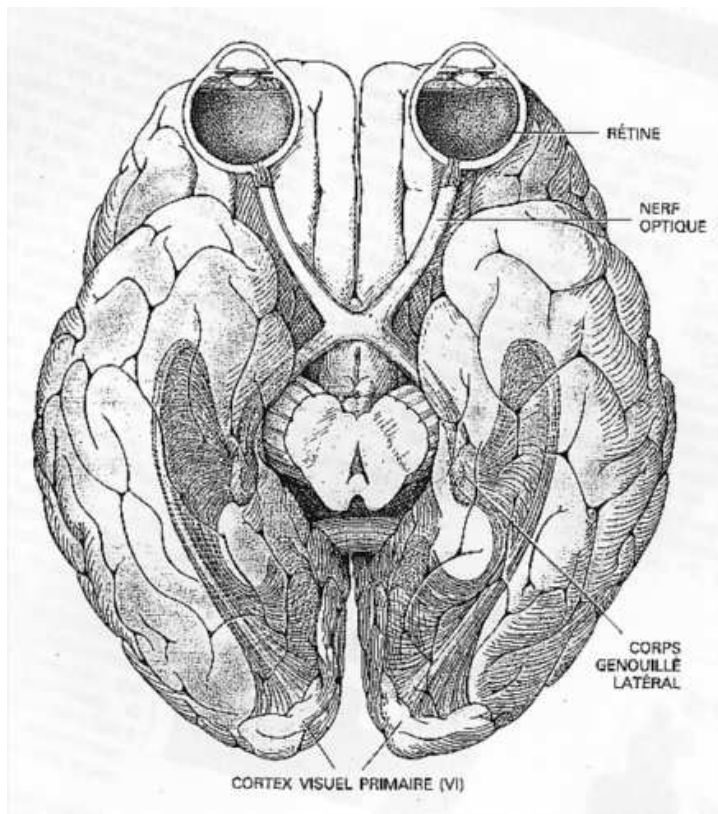


Fig . 2 : Le système visuel humain (source : [ZEK-92])

De nombreuses caractéristiques visuelles de la scène influencent l'attention du spectateur. Différentes études, comme celles de Finlay [Fin-80], ou de Yarbus [Yar-67] ont permis de déterminer ces facteurs, notamment à partir des observations des mouvements des yeux d'une personne qui regarde une scène. Parmi ces facteurs on distingue des facteurs de bas niveau et de haut niveau. Les facteurs de haut niveau les plus importants sont les suivants :

- Le contraste. L'œil humain est fortement attiré par un contraste élevé.

- La taille. Les régions de grande taille sont plus attractives, lorsque celles-ci ne dépassent pas une certaine limite.
- La forme : les régions longues et fines seraient d'après certains auteurs comme Osberger et Maeder [Os-98b] plus attractives que les régions plus arrondies. Cependant, d'autres études comme celles de Luo et Singhal [Luo-01] affirment le contraire. Le contexte joue probablement un rôle important en ce qui concerne l'attractivité des formes.
- La couleur : certaines couleurs, telles que le rouge, attirent fortement l'attention visuelle. L'attention visuelle est également influencée par les objets dont la couleur est différente de celle du fond.
- Le mouvement : La vision périphérique humaine étant optimisée pour détecter les mouvements, le mouvement est l'un des facteurs qui attirent le plus l'attention visuelle. Cependant, le mouvement ne constitue pas un critère pertinent pour caractériser une région d'intérêt dans une image fixe.

Les facteurs de haut niveau comprennent notamment :

- La position dans l'image : le regard du spectateur se dirige le plus souvent vers le centre de l'image.
- Le premier plan qui est plus attractif que l'arrière-plan.
- La présence de personnes dans la scène. L'attention visuelle humaine est attirée par les personnes, et en particulier par les visages de celles-ci.
- Le contexte. Le mouvement du regard du spectateur peut être considérablement différent selon ses objectifs lorsqu'il observe l'image.

Il n'est pas toujours facile de déterminer parmi tous ces facteurs lesquels sont plus importants que les autres, cela dépend des images. C'est pourquoi il est utile de prendre plusieurs facteurs en considération pour modéliser l'attention visuelle. Les facteurs de haut niveau sont très utiles pour déterminer une région d'intérêt, mais ils sont difficilement utilisables si le contenu de la scène ou le contexte de visualisation n'est pas connu a priori. C'est pourquoi, dans le cadre de cette étude, on se limitera essentiellement aux méthodes d'analyse faisant appel aux facteurs de bas niveau.

1.3 Mesures de saillance basées sur les caractéristiques visuelles de bas niveau

Le principe général de ces méthodes est d'utiliser les informations de bas niveau définies précédemment afin de déterminer les régions perceptuellement saillantes dans l'image en calculant des mesures de *saillance absolue* et/ou de *saillance relative*. La saillance absolue ou saillance propre est ce qui détermine combien une région est remarquable à conserver. Elle se définit à partir de paramètres propres à la région comme la couleur et la taille de celle-ci, sans tenir compte des propriétés des régions voisines. La saillance relative détermine comment une région se distingue des régions voisines, elle est quant à elle définie par des paramètres comme le contraste ou la différence de taille entre une région et les régions voisines. Le système visuel humain utilise beaucoup la saillance relative, en particulier la différence entre les comportements du centre et de l'entourage du champ visuel, ou contraste *center-surround*. Les cellules nerveuses du système visuel sont plus sensibles dans une petite région centrale de l'espace visuel, alors que les stimuli présentés sur une région antagoniste plus large qui entoure la région centrale inhibent la réponse neuronale. Une telle architecture qui est sensible

aux discontinuités locales spatiales est particulièrement bien adaptée pour détecter les régions saillantes dans l'image.

D'une manière générale, les méthodes basées sur la modélisation de la vision humaine de bas niveau cherchent à dresser des *cartes de saillance* pour chacune des régions de l'image

La méthode proposée par Osberger et Maeder [Os-98] permet de déterminer les zones d'intérêt dans une image par la construction de cartes de saillance. Cette méthode consiste à segmenter l'image en régions, à calculer différents facteurs d'attention pour chacune des régions et à combiner ces différents facteurs afin de construire une carte d'importance qui permet de déterminer quelles régions sont les plus importantes du point de vue de l'attention visuelle. La méthode de segmentation est une méthode récursive de séparation et de fusion qui utilise la variance locale des niveaux de gris. Le seuil de variance est de 250 et les régions de moins de 16 pixels sont fusionnées avec les régions voisines les plus similaires. Ensuite, on calcule 5 facteurs d'attention visuelle pour chaque région, à savoir le contraste, la taille de la région, la forme de celle-ci, sa position par rapport au centre de l'image et son appartenance au fond. Le contraste est défini comme la différence entre la moyenne des niveaux de gris de la région et la moyenne des niveaux de gris des régions voisines. La taille de la région, est pondérée afin d'éviter d'accorder une importance excessive aux plus grandes régions qui constituent souvent le fond de l'image. Pour la forme de la région, un indice de forme est calculé de manière à accorder une importance plus grande aux régions longues et fines. L'appartenance de la région au fond est calculée en déterminant la proportion de la bordure de l'image qui est contenue dans chaque région.

D'autres paramètres pourraient également être utilisés pour caractériser l'attention visuelle. Ces 5 facteurs sont normalisés dans l'intervalle [0,1]. On peut supposer que les 5 facteurs ont une importance égale, ou bien afin d'assigner une importance plus grande aux régions qui ont des valeurs plus importantes pour ces facteurs, l'on calcule la somme quadratique des différents facteurs afin d'obtenir l'importance visuelle de chaque région. On normalise de nouveau l'importance visuelle de chaque région en attribuant la valeur 1 à la région d'importance visuelle la plus élevée, et on construit la carte d'attention visuelle de l'image. Cette méthode permet d'identifier les régions visuellement les plus importantes de l'image sur des images simples ou complexes.

La méthode proposée par Privitera et Stark [Pri-98] permet de reconstituer des séquences de régions d'intérêt qui modélisent les fixations successives de l'œil humain sur les différents points d'intérêt d'une scène au cours de l'observation de celle-ci. Différentes mesures de saillance sont utilisées pour identifier les différentes régions d'intérêt de l'image, notamment une transformation de symétrie, le contraste de Michaelson, le contraste *center-surround*, les orientations des contours, l'entropie et la densité de contours. Des séquences de régions d'intérêts sont déterminées à partir des maxima locaux de ces différentes caractéristiques. A partir de ces séquences, des matrices de Markov sont construites pour modéliser les transitions des mouvements oculaires entre les différentes régions d'intérêt d'une image.

Une méthode basée sur une segmentation de l'image en régions de couleur et une mesure de saillance basée sur la couleur a été proposée par Syeda-Mahmood [Sye-97]. Contrairement à la méthode précédente qui utilise l'image en niveaux de gris, cette méthode nécessite une image en couleur. L'espace couleur utilisé ici est l'espace HSV. Cet espace couleur a été choisi car la teinte, la saturation et la luminance sont les attributs significatifs au niveau perceptuel. On utilise une table prédéfinie de 220 catégories de couleurs selon les valeurs de H, S et V. La segmentation de l'image en régions de couleurs consiste à affecter chaque pixel

de l'image à une catégorie de couleur et à regrouper ensemble les pixels selon les composantes connexes afin de former des régions qui sont affectées à une catégorie de couleur. L'image est divisée en blocs de taille 8x8, pour chaque bloc on liste les catégories de couleurs qui s'y trouvent et on regroupe les blocs voisins contenant la même couleur pour former des sous-ensembles connexes, un même bloc peut appartenir à plusieurs sous-ensembles. On utilise ensuite un algorithme d'étiquetage afin de regrouper les différents sous-ensembles qui se recouvrent et de définir les régions de couleurs selon la catégorie de couleur dominante.

Afin de caractériser les différentes régions détectées, on définit des mesures de saillance propre et de saillance relative qui sont combinées linéairement en utilisant une pondération heuristique afin de déterminer la saillance globale. En ce qui concerne la couleur, on utilise uniquement ici la saturation et la luminance. La taille est définie comme le rapport entre l'aire de la région et l'aire de l'image.

La saillance couleur d'une région est définie comme la somme de 5 paramètres. Les deux premiers paramètres sont la saturation et la luminance. Le troisième paramètre est la taille de la région pondérée par une fonction de pondération qui varie suivant la taille de la région afin de réduire l'importance des régions trop grandes ou trop petites. Les régions sont réparties en 5 catégories suivant leur aire, pour chacune de ces catégories on utilise une formule différente pour calculer une fonction de pondération. Le quatrième paramètre est le contraste couleur qui est défini comme la distance entre les couleurs de deux régions dans l'espace RGB normalisé, à laquelle on ajoute une constante de pondération si les régions ont des teintes différentes dans l'espace HSV. Le dernier paramètre est calculé à partir du contraste de taille, c'est à dire du rapport entre l'aire d'une région et celle de la région voisine. Ce contraste est mesuré par rapport à la « meilleure » région voisine, qui est définie comme la région de rang le plus élevé dans le classement selon la taille, puis selon le contraste. Afin de favoriser les régions de taille sensiblement égale à celle de leur voisine, le contraste de taille est pondéré par une fonction exponentielle $f(t(R))=e^{-12t(r)}$, où $t(R)$ représente le contraste de taille de la région.

La saillance d'une région de texture peut être déduite du contraste relatif entre les régions claires et sombres de cette texture. On peut alors définir une mesure de saillance de texture qui extrait les zones relativement claires et sombres d'une texture et analyse leurs attributs saillants tels que le nombre, l'aire, la forme et la distribution des trous dans une région claire. Cependant, toutes les textures naturelles ne peuvent être modélisées de cette manière de façon satisfaisante.

Un estimateur de l'attention visuelle basé sur les dissimilarités de voisinage dans une image a été proposé par Stentiford [Ste-01]. Cet estimateur est destiné à déterminer les points de l'image qui attirent l'attention visuelle. L'application envisagée est de concevoir un système de compression à taux de compression variable. Le principe de cette méthode est de mesurer les différences des composantes couleurs entre des points choisis aléatoirement dans l'image. On cherche à mettre en correspondance différents voisinages de l'image. Les voisinages correspondent si les différences entre les composantes couleurs R,G,B des pixels de ce voisinages sont inférieures à un certain seuil. Les zones qui attirent l'attention visuelle, autrement dit les zones d'intérêt, sont les groupes de pixels qui ont le moins de correspondants dans l'image. Cela signifie que les pixels qui possèdent des couleurs nouvelles qui n'apparaissent pas ailleurs dans l'image seront associés à une estimation d'attention visuelle élevée. Cela veut dire aussi que les contours qui séparent des régions de couleurs différentes recevront des scores d'attention visuelle élevés. L'estimateur de Stentiford fournit,

à partir d'une image, une matrice de scores d'attention visuelle. Un seuillage est effectué afin de conserver comme zones d'intérêt les pixels ayant les scores les plus élevés. De cette manière, les configurations de pixels qui représentent des structures communes dans l'image sont éliminées. Pour réduire le temps de calcul, on peut utiliser une approche multirésolution afin d'éliminer les grandes zones uniformes. Les cartes d'attention visuelle générées par cette méthode sont ensuite utilisées par un algorithme de compression à résolution variable dérivé du JPEG afin de compresser différemment le contenu de l'image selon son degré d'attention visuelle.

1.4 Modèles d'attention visuelle multi-échelle pour l'analyse de scènes

Le modèle d'attention visuelle proposé par Itti et al. [Itt-98] utilise une approche ascendante basée sur une analyse multi-échelle de caractéristiques de bas niveau afin de construire une carte de saillance de l'image. Le modèle simule le fonctionnement des centres nerveux qui sont responsables de la vision de bas niveau chez l'être humain.

A partir de l'image initiale, on construit 9 images réduites en les rééchantillonnant à l'aide de la méthode des pyramides gaussiennes dyadiques. Chaque caractéristique est calculée en utilisant le modèle *center-surround* comme dans le système nerveux humain. Ce modèle est ici représenté par une différence d'échelle. On définit un opérateur de différence d'échelle noté Θ qui est une interpolation à l'échelle la plus fine et soustraction point à point.

Cet opérateur de différence d'échelle est utilisé pour réaliser des cartes d'attention visuelle à partir de caractéristiques d'intensité, de couleur et d'orientation. La caractéristique d'intensité consiste en une mesure de contraste entre deux échelles différentes. Les caractéristiques de couleurs simulent la perception des couleurs par le cortex visuel humain, dans le cortex primaire, des neurones sont activés par une couleur (par exemple le rouge) dans le centre de leur champ de vision et inhibés par une autre couleur (comme le vert), et c'est l'inverse sur le pourtour du champ de vision. Ce phénomène est ici modélisé par des cartes de double opposition de couleurs calculées à partir de 4 canaux couleurs R,G,B,Y définis à partir des trois composantes couleurs (r,g,b) de l'espace RGB, $R=r-(g+b)/2$ pour le bleu, $G=g-(r+b)/2$ pour le vert, $B=b-(r+g)/2$ pour le rouge et $Y=(r+g)/2-|r-g|/2-b$ pour le jaune. On crée 4 pyramides gaussiennes à partir de ces 4 canaux couleurs. Les cartes d'opposition couleur bleu/jaune et vert/rouge sont calculées à différentes échelles. On calcule également des cartes d'orientation en utilisant des pyramides de Gabor à différentes orientations. Les filtres de Gabor simulent les neurones sensibles à l'orientation du cortex visuel. On calcule au total 42 cartes de caractéristiques : 6 pour l'intensité, 12 pour la couleur et 24 pour l'orientation.

Des cartes de remarquabilité sont construites à partir de ces cartes de caractéristiques, les caractéristiques sont normalisées à l'aide d'un opérateur de normalisation $\mathcal{N}(\cdot)$ qui permet de favoriser les cartes dont la valeur de la caractéristique présente un petit nombre de pics. Cet opérateur simule le fonctionnement des mécanismes d'inhibition du cortex latéral. Les cartes de caractéristiques sont combinées en trois cartes de remarquabilité, \bar{I} pour l'intensité, \bar{C} pour la couleur et \bar{O} pour l'orientation qui sont obtenues en utilisant un opérateur d'addition inter-échelle. Pour finir, on calcule la carte de saillance par sommation des trois cartes de remarquabilité.

Le maximum de la carte de saillance définit le point le plus saillant de l'image, c'est à dire celui qui attire le plus l'attention. La zone d'intérêt est définie comme un cercle dont le rayon est égal à 1/6 de la plus petite dimension de l'image (largeur ou hauteur) centré au point d'attention maximale. On peut définir plusieurs zones d'intérêt successives en prenant les valeurs maximales de la carte de saillance dans l'ordre décroissant. Le modèle conçu par Itti

et al. a été comparé au modèle SFC (Spatial Frequency Content) défini par Reinagel et Zador [Rei-97] qui se base sur une transformée de Fourier sur des blocs 16x16 de l'image. Par rapport au modèle SFC, le modèle de Itti et al. est plus performant surtout sur les images présentant de fortes variations de couleur et d'intensité, car il implémente une compétition spatiale entre les différentes régions saillantes de l'image. Ce modèle se caractérise également par une grande robustesse au bruit.

Différents modèles de voisinages pour le calcul d'une mesure de saillance relative ont été comparés par Luo et Singhal [Luo-01]. Le modèle I définit le voisinage comme la région adjacente qui a la plus forte valeur de la caractéristique étudiée. Le modèle II considère toutes les régions voisines définies par un graphe d'adjacence. Le modèle III s'intéresse à un voisinage étendu constitué de toutes les régions de l'image qui se trouvent dans une fenêtre centrée approximativement sur la région à évaluer. On détermine d'abord le rectangle englobant minimal de la région. Ensuite, on agrandit ce rectangle dans toutes les directions, par exemple d'un facteur 2. Toutes les régions qui sont incluses au moins partiellement dans le rectangle agrandi sont sélectionnées comme voisines de la région à évaluer. C'est la région en totalité qui est incluse dans le voisinage, et pas seulement la partie de celle-ci incluse dans le rectangle. Les dimensions de la fenêtre de voisinage sont approximativement un multiple du rectangle englobant minimal, le voisinage final n'étant pas nécessairement une fenêtre rectangulaire. Ce modèle de voisinage étendu ressemble à celui qui est utilisé par le système visuel humain, en ce sens qu'il évalue une mesure relative entre le centre et la périphérie et que l'extension du champ de vision n'est pas isotrope mais dépend du contenu de l'image. On utilise une pondération des différentes régions qui associe un poids maximal aux régions voisines de taille comparable à celle de la région considérée.

Pour évaluer la saillance d'une région de l'image, différentes mesures de saillance relative, basées sur la couleur et la forme sont utilisées. Les mesures basées sur la couleur sont le contraste de teinte, la saturation, le contraste de saturation, la luminance et le contraste de luminance. Le rapport longueur/ largeur est défini comme le rapport entre le plus grand axe et le plus petit axe de la région.. On définit également trois mesures de convexité, la convexité du périmètre, qui est le rapport entre le périmètre de l'enveloppe convexe et le périmètre de la région, la convexité de la surface, qui est le rapport entre l'aire de la région et celle de l'enveloppe convexe, et enfin la complexité, définie comme le rapport entre la convexité du périmètre et la convexité de la surface. On calcule également la rectangularité, qui est le rapport entre l'aire du rectangle englobant minimal et l'aire de la région, et la circularité, qui est le rapport entre le carré du périmètre de la région et l'aire de celle-ci. Pour chacune de ces caractéristiques, on calcule une mesure de saillance relative. D'après Luo et Singhal, le modèle III donne de meilleures performances pour la détection des zones d'intérêt avec l'ensemble des caractéristiques que les modèles de voisinage I et II dont les performances sont nettement moins bonnes.

1.5 Utilisation de transformations de symétrie

Les objets constituant la zone d'intérêt d'une image présentent souvent des symétries dans leurs formes, c'est pourquoi des méthodes basées sur des mesures de symétrie ont été proposées afin de détecter ces zones d'intérêt.

La méthode proposée par Di Gesu et al. [DiG-97] pour détecter des régions d'intérêt dans l'image se base sur deux transformations de l'image, la transformée des moments discrets (DMT) et la transformée de symétrie discrète (DST). La transformation DMT est calculée en évaluant les moments centrés autour de chaque pixel, et la transformation DST mesure la symétrie annulaire locale. Les résultats des deux transformations sont complémentaires, la DMT est sensible à la distribution des intensités locales alors que la DST est plus sensible aux relations spatiales.

La transformée des moments est effectuée sur un noyau de taille $2k+1$, on a $p, q = 0, 1, 2, \dots$. La transformée des moments de l'image $a_{i,j}$ est donnée par la formule suivante :

$$DMT_{i,j}(p,q) = \sum_{r=-k}^{r=+k} a_{i-r,j-s} (i-r)^p (j-s)^q \quad (1-5-1)$$

Les résultats de la DMT dépendent des valeurs de p et q , avec deux nombres pairs on lisse l'image, avec d'autres combinaisons on met en évidence les contours.

La transformée en symétrie discrète est le produit de deux opérateurs locaux $F_{i,j}$ et $E_{i,j}$.

Le premier opérateur est une fonction des moments axiaux calculés dans une fenêtre C_k de taille linéaire $2k+1$ centrée en (i,j) , qui s'exprime par

$$T_{i,j}^h = \sum_{r=-k}^{+k} \sum_{s=-k}^{+k} \left| r \sin\left(\frac{h\pi}{n}\right) - s \cos\left(\frac{h\pi}{n}\right) \right| f_{i-r,j-s} \quad (1-5-2)$$

Avec $h=0, 1, 2, \dots, n-1$, où n est le nombre d'axes de symétrie utilisé. La fonction F dépend du type de symétrie que l'on veut détecter. Dans le cas d'une symétrie annulaire, la formule est la suivante :

$$F_{i,j} = 1 - \sqrt{\frac{\sum_h (T_{i,j}^h)^2}{n} - \left(\frac{\sum_h (T_{i,j}^h)}{n} \right)^2} \quad (1-5-3)$$

Le second opérateur $E_{i,j}$ sert à pondérer F suivant que la texture de l'image est plus ou moins lisse. Il est définie par la formule :

$$E_{i,j} = \sum_{(l,m) \in C_k, (r,s) \in C_{k+1}} |f_{l,m} - f_{r,s}| \quad (1-5-4)$$

Dans cette formule les pixels de coordonnées (l,m) et (r,s) doivent être voisins. Si $E_{i,j} = 0$ l'image est localement plate. Les deux transformations peuvent être appliquées en série ou en parallèle. Si on les applique en série, on calcule d'abord la transformation DMT, puis on applique la transformation DST sur le résultat de celle-ci.

Si on les applique en parallèle, on calcule deux images transformées M et S , obtenues respectivement par les transformations DMT et DST à partir de l'image d'origine D . Les valeurs de la moyenne et de la variance (μ_M, σ_M) et (μ_S, σ_S) sont utilisées pour sélectionner la région d'intérêt selon la règle suivante :

$$T(DXT(D), \mu_x, \sigma_x, \alpha) = \begin{cases} DXT(D) & \text{si } DXT(D) > \mu_x + \alpha \times \sigma_x \\ 0 & \text{sinon} \end{cases} \quad (1-5-5)$$

Dans cette formule DXT représente la transformation DMT ou DST et α est une constante positive. On fusionne ensuite les résultats des deux transformations à l'aide de l'opérateur \otimes défini comme suit : $y \otimes x = x \Leftrightarrow y > 0$. On sélectionne alors les points d'intérêt de la région x qui appartiennent aussi à la région y . La combinaison en série ou en parallèle des deux transformations permet d'obtenir une meilleure définition des régions d'intérêt : la transformation DMT permet de déterminer les grandes régions d'intérêt et la transformation

DST permet de repérer les détails fins à l'intérieur de ces régions. Cette méthode est efficace, les essais effectués par les auteurs ont permis de détecter la zone d'intérêt préalablement identifiée par un observateur humain dans 80% des images testées.

Une autre transformation de symétrie a été utilisée par Privitera et Stark [Pri-98]. Pour chaque pixel de coordonnées x,y de l'image, on définit une mesure de symétrie locale à l'aide de la formule suivante :

$$S(x,y) = \sum_{(i_1,j_1),(i_2,j_2) \in \Gamma(x,y)} s((i_1,j_1),(i_2,j_2)) \quad (1-5-6)$$

$\Gamma(x,y)$ représente le voisinage de rayon 7 du point x,y défini suivant l'axe horizontal et vertical. :

$$\Gamma(x,y) = (x-r,y) \dots (x,y) \dots (x+r,y) \dots (x,y-r) \dots (x,y+r) \quad (1-5-7)$$

L'équation suivante permet de définir $s((i_1,j_1),(i_2,j_2))$:

$$s((i_1,j_1),(i_2,j_2)) = G_{\sigma}(|d((i_1,j_1),(i_2,j_2))|) \cos(\theta_1 - \theta_2) \quad (1-5-8)$$

G est une gaussienne de variance fixée, $\sigma=3$ pixels et d est une fonction de variance. Le second facteur est une mesure simplifiée de symétrie, θ_1 et θ_2 représentent les angles de gradient de niveau de gris au voisinage des deux pixels (i_1,j_1) et (i_2,j_2) . La gaussienne représente une fonction de pondération de la distance qui introduit une localisation dans l'évaluation de symétrie. La définition de la symétrie est ici basée sur les correspondances des orientations des gradients autour d'un point central.

D'autres transformations de symétrie sont encore possibles, par exemple la transformation radiale proposée par Loy et Zelinsky [Loy-03]. Cette transformation calcule la contribution à la symétrie radiale du gradient à une certaine distance n de chaque point. Cette transformation de symétrie peut être calculée à plusieurs échelles. Pour chaque échelle, on construit une image de projection des orientations et une image de projection des amplitudes. Ces images sont générées en examinant le gradient en chaque point p de l'image et en déterminant un point affecté positivement et un point affecté négativement, situés chacun à une distance n du point p , l'un dans la direction du gradient et l'autre dans la direction opposée.

1.6 Détection de zones d'intérêt à l'aide d'ondelettes

Dans les images photographiques à faible profondeur de champ, les objets d'intérêt constituent le premier plan de l'image et sont plus nets que le reste de celle-ci. Par conséquent ils comportent plus de détails. Des méthodes utilisant les ondelettes ont été proposées pour détecter ces objets situés au premier plan. La présence de détails se traduit par le fait que les coefficients des hautes fréquences de la transformée en ondelettes ont une amplitude et une variance plus importante dans le cas des objets de premier plan. Nous décrivons ici deux méthodes reposant sur ce principe.

La méthode développée par Wang et al. [Wan-01] utilise les ondelettes de Haar et une approche multirésolution pour la classification des objets et repose sur une comparaison entre des zones de l'image. Initialement, l'image est découpée en blocs de taille 32x32 et on applique la transformée en ondelettes de Haar sur chacun des blocs. Les ondelettes de Haar

ont été choisies en raison de leur temps de calcul plus réduit que celui des autres modèles d'ondelettes. Les blocs sont classifiés comme appartenant au fond ou à la région d'intérêt en fonction de l'intensité moyenne et de la variance de chacun des blocs. L'algorithme de classification comporte trois phases : la classification initiale à l'échelle la plus petite, un processus récursif pour ajuster la classification initiale en utilisant une approche multi-échelle et un post-traitement pour lisser les contours et supprimer les petites régions isolées. La classification initiale consiste à calculer la variance $v_{i,j}$ des coefficients d'ondelettes des 3 bandes de plus hautes fréquences pour chacun des blocs i,j à l'échelle 0. On classe alors les blocs en deux classes (objet d'intérêt et arrière-plan) suivant la valeur de v en utilisant l'algorithme des k-means. Pour affiner cette classification initiale, on utilise un algorithme multi-échelle. On découpe chacun des blocs d'échelle r en 4 à l'aide d'un quad-tree et on recalcule la variance $v_{i,j}$ et le niveau de gris moyen $x_{i,j}$. On reclassifie ensuite les blocs de la zone d'intérêt qui ont au moins un voisin appartenant au fond, en fonction de la différence entre l'intensité moyenne d'un bloc et celle des blocs parents. Si le bloc (i,j) a été affecté au fond, alors tous les blocs voisins qui ont été classés dans la région d'intérêt sont à nouveau affectés au fond. On recommence cette classification jusqu'à ce que l'on atteigne la taille minimale des blocs fixée par l'utilisateur et qui peut aller jusqu'à un seul pixel. Avec cette méthode, on peut arriver à une segmentation très précise des objets d'intérêt de l'image. Cette méthode est totalement indépendante des caractéristiques de forme ou de couleur de l'objet.

La méthode développée par Zhen Ye et Cheng-Chang Lu [Ye-02] utilise elle aussi les ondelettes de Haar, mais ici la phase de classification est assurée par les arbres de Markov. La transformée en ondelettes de Haar est ici modélisée sous la forme d'un quad-tree. Chaque coefficient d'ondelette est associé à un label de classe. On définit deux classes : l'objet (O) et le fond (B). Si l'on considère que les coefficients d'ondelettes et les labels de classes sont deux processus aléatoires désignés respectivement par W et C , le processus W est le processus d'observation et le processus C est le processus caché correspondant. Les processus W et C sont indexés par les nœuds du quad-tree. On a donc une série descendante de processus aléatoires, et si l'on suppose que chaque processus dépend uniquement du processus précédent, on obtient donc une chaîne de Markov. La densité de probabilité de chaque coefficient est modélisée par une mixture gaussienne. Dans ce modèle, on suppose que chaque coefficient d'ondelette est conditionnellement indépendant étant donné son label de classe. La corrélation entre les coefficients d'ondelettes donne une information importante sur le contexte. Dans ce modèle, on considère uniquement la relation entre un nœud et ses ascendants et descendants immédiats. Soit $p(i)$ le nœud parent du nœud i , pour chaque paire de labels cachés $\{C_{p(i)}, C_i\}$ on définit par $\mathcal{E}_{i,p(i)}^{n,r}$ la probabilité de transition qui représente la probabilité conditionnelle que le nœud i appartienne à la classe n sachant que $p(i)$ appartient à la classe r . L'information contextuelle de l'image peut alors être décrite par une matrice de transition.

On notera ϕ_c le vecteur de paramètres du modèle de Markov caché pour la classe c . Ce vecteur peut être considéré comme un vecteur de caractéristiques pour cette classe. Le maximum de probabilité a posteriori $p(C|W, \phi_c)$ peut être obtenu par l'algorithme EM multi-échelle et il est utilisé comme critère de segmentation. Le label de classe d'un coefficient d'ondelette C_w est donné par $C_w = \arg\max_{c \in \{O,B\}} p(c|W, \phi_c)$, où B représente le nombre de bandes de fréquence de la transformation en ondelettes.

Chacun des blocs dyadique D de la transformée en ondelettes dans le domaine spatial est classifié comme étant un objet ou le fond en fonction de sa probabilité conditionnelle $p(C|D, \phi_c)$. Chaque bloc est associé à trois sous-arbres dans les trois bandes de hautes fréquences. Si l'on suppose que les trois bandes sont indépendantes, la probabilité

conditionnelle $f(C|D, \phi_c)$ est le produit des fonctions de distribution $f(C|T, \phi_c)$ des trois bandes de haute fréquence. Un algorithme simple de montée et de descente de gradient peut être utilisé pour calculer $f(C|T, \phi_c)$ à l'aide de la probabilité $f(C|W, \phi_c)$ obtenue au moyen de l'algorithme EM. Cependant, le résultat obtenu aux échelles les plus fines n'est pas très bon en raison du manque d'informations statistiques. C'est pourquoi il est nécessaire d'opérer une fusion multi-échelles afin d'améliorer la segmentation. Soit V le vecteur de contexte du bloc dyadique D . Le bloc D étant indépendant de V , et $f(D)$ et $f(D/V)$ étant indépendants de C , on peut calculer $f(C/D, V)$:

$$f(C|D, V) = \frac{\alpha f(C|D) f(C|V)}{f(C)} \quad (1-6-1)$$

où α est une constante de proportionnalité. On peut obtenir $f(C|D)$ par segmentation brute. Pour calculer $f(C)$, on utilise l'algorithme EM. La phase E de l'algorithme consiste à calculer :

$$p(C|D, V) = \frac{f(C|D) f(C|V)}{f(C)} \quad \text{et} \quad f(C|D, V) = \frac{p(C|D, V)}{\sum_{c \in \{O, B\}} p(c|D, V)} \quad (1-6-2)$$

La phase M de l'algorithme permet de calculer $f(C)$:

$$f(C) = \frac{1}{2^k} \sum f(C|D, V) \quad (1-6-3)$$

La probabilité contextuelle $f(C|V)$ peut être calculée en utilisant la méthode HCLT (Hybrid Contextual Labeling Tree) proposée par Fan et Xia [Fan-00]. Avec cette méthode, seule l'échelle précédente et l'échelle courante sont considérées comme étant le contexte. Plusieurs contextes peuvent être implémentés séquentiellement pour obtenir de meilleurs résultats. Ici, trois contextes sont considérés pour le bloc D , le bloc parent V_p , ses 8 voisins $V_{p:n}$ et les huit voisins V_n de D . La segmentation finale, en tenant compte du contexte, s'effectue en appliquant la formule suivante :

$$C_D = \operatorname{argmax}_{c \in \{O, B\}} f(C_D|D, V, \phi_c) \quad (1-6-4)$$

Aux échelles les plus fines, les blocs ont une très petite taille. Il n'y a alors pas de différence très visible entre les objets et le fond. Dans ce cas, la variance des coefficients d'ondelettes n'est pas un très bon indicateur. C'est pourquoi on utilise les informations de contexte pour obtenir la segmentation au niveau le plus fin. Après la fusion du contexte multi-échelle, on utilise un post-traitement pour éliminer les régions isolées et lisser les contours de l'objet.

Les méthodes à base d'ondelettes permettent une segmentation très précise des objets d'intérêt et du fond de l'image. Cependant, elles sont limitées au cas particulier des images à faible profondeur de champ, comme par exemple les images prises au téléobjectif. La méthode risque de ne pas convenir à d'autres types d'images.

1.7 Approche multi-échelles basée sur l'entropie

La notion d'entropie, au sens de la théorie de l'information définie par Shannon, a été utilisée pour détecter des régions saillantes de l'image, notamment à travers des approches multi-échelles.

La méthode développée par Kadir et Brady [Kad-00], [Kad-03] utilise une mesure de saillance qui est définie comme une entropie pondérée de l'histogramme des niveaux de gris sur un voisinage. Afin de pouvoir détecter des régions d'intérêt de taille différente, on utilise une approche multi-échelles où l'on détermine la meilleure échelle pour chaque région d'intérêt. On définit plusieurs échelles qui sont les tailles du voisinage sur lesquelles on calcule une mesure de saillance d'échelle y_D définie à partir de l'entropie. Pour une position x de l'image et pour une échelle s , la saillance d'échelle Y_D est définie comme le produit de l'entropie H_D et de la saillance inter-échelle W_D

$$Y_D(s_p, x) \stackrel{\Delta}{=} H_D(s_p, x) W_D(s_p, x) \quad (1-7-1)$$

L'entropie H_D étant définie par la formule :

$$H_D(s, x) \stackrel{\Delta}{=} - \sum_{d \in D} p_{d,s,x} \log_2 p_{d,s,x} \quad (1-7-2)$$

Dans cette formule, $p_{d,s,x}$ désigne la probabilité en fonction de l'échelle s , de la position x et de la valeur d du descripteur et D est l'ensemble des valeurs possibles du descripteur. La saillance inter-échelle, pour une position x de l'image et une échelle s , est définie par la formule suivante :

$$W_D(s, x) \stackrel{\Delta}{=} \frac{s^2}{2s-1} \sum_{d \in D} |p_{d,s,x} - p_{d,s-1,x}| \quad (1-7-3)$$

On sélectionne les échelles pour lesquelles l'entropie atteint un pic, ces échelles forment un vecteur s_p défini par la formule suivante :

$$s_p = \{s : H_D(s-1, x) < H_D(s, x) \text{ et } H_D(s+1, x) < H_D(s, x)\} \quad (1-7-4)$$

Les points d'intérêt sont les points pour lesquels la saillance dépasse un certain seuil. Afin de rendre la méthode plus robuste au bruit, on regroupe les points d'intérêt en régions en sélectionnant les points d'intérêt qui ont des voisins ayant une valeur de la saillance et une échelle similaires. Pour chacun des points dont la saillance est supérieure à un seuil, on teste les K points les plus proches, K étant une valeur prédéfinie. Si la variance de la saillance est inférieure à un seuil prédéfini et si la distance dans l'espace (x, y, s) à une autre région saillante est inférieure à l'échelle moyenne de la région, alors la région est considérée comme saillante. D'autres mesures que l'entropie pourraient être utilisées pour caractériser la saillance. Cette méthode permet de détecter des objets dans une image fixe, mais elle peut aussi être utilisée pour le suivi d'objets en mouvement, en effet les régions saillantes correspondent aux mêmes objets d'une image à l'autre, et il serait possible de prendre le mouvement en compte dans la mesure de la saillance.

1.8 Approche multirésolution pour l'apprentissage du fond

La méthode proposée par Lefèvre [Lef-02] utilise un apprentissage de l'arrière-plan à l'aide d'une approche multirésolution afin de séparer le fond des objets contenus dans l'image. Il est à noter que cet apprentissage est réalisé sur l'image analysée elle-même, et non pas à partir d'images de référence. Cette méthode utilise une approche multirésolution pour extraire le

fond de l'image, l'hypothèse de départ étant que lorsque la résolution est faible, le fond est majoritairement représenté. Le modèle du fond est obtenu en construisant une représentation pyramidale de l'image par rééchantillonnages successifs en partant de l'image initiale que l'on considère comme ayant une résolution $r=0$ jusqu'à une résolution r_{\max} qui constituera le modèle initial de l'arrière-plan. Chaque pixel de l'image de résolution $r+1$ a pour valeur la moyenne des valeurs des pixels de la région correspondante de l'image à la résolution précédente r . On calcule ensuite une valeur moyenne de la teinte $\varphi(I)$, pour modéliser l'arrière-plan ou une région donnée. La teinte a été choisie comme paramètre en raison de sa robustesse aux conditions d'illumination et aux redimensionnements successifs. En partant de la résolution $r_{\max-1}$, on va comparer chaque région avec le modèle de l'arrière-plan en utilisant la mesure de différence δ . Si la valeur de δ est inférieure à un seuil $S1$, on calcule une mesure de l'homogénéité de la région qui est l'amplitude de la variation de la teinte sur cette région. Si la mesure d'homogénéité est supérieure à un seuil $S2$, la région est étiquetée comme faisant partie du fond. Sinon, on examine les régions correspondant à la résolution $r-1$. Afin d'améliorer la qualité du modèle du fond, la valeur moyenne de la teinte peut être recalculée au cours du processus de segmentation. Dans ce cas, le modèle obtenu à la résolution la plus petite ne représente que l'état initial de l'arrière-plan. A mesure que la résolution devient plus fine, les résultats sont plus précis, et il est possible d'obtenir un modèle de l'arrière-plan plus fiable en ne considérant que les parties de l'image déjà étiquetées comme le fond. On réitère le processus jusqu'à la résolution initiale $r=0$. Les régions qui ne sont pas étiquetées comme faisant partie de l'arrière-plan correspondent a priori à des objets d'intérêt.

Cette méthode offre un temps de calcul très réduit, ce qui a permis de l'utiliser pour l'analyse de vidéos en temps réel. Par contre, elle nécessite un fond de teinte assez uniforme, comme par exemple dans le cas des joueurs évoluant dans un stade de football. Elle risque de ne pas donner de résultats satisfaisants sur d'autres types d'images.

1.9 Détection d'objets par dimension fractale

L'utilisation de la notion de dimension fractale définie par Mandelbrot [Man-83] a été proposée pour détecter des objets et des régions d'intérêt.

La méthode de détection d'objets artificiels proposée par M.J. Carlotto et M.C. Stein [Car-90] a une application qui est un peu fantaisiste au premier abord puisqu'il s'agit de détecter des objets artificiels sur des images de la planète Mars. En effet, selon ces auteurs, certains objets visibles sur les images de Mars prises en 1976 par la sonde Viking pourraient être des vestiges d'une civilisation extraterrestre disparue, en particulier un monticule ressemblant à un visage et un autre à une forteresse. Bien sûr, ce point de vue est extrêmement discutable, et il est bien plus probable que ces reliefs soient le résultat d'un processus géologique naturel. Il n'en reste pas moins qu'en tout état de cause, ils constituent incontestablement des zones d'intérêt qui sont dignes d'être détectées. D'ailleurs, la méthode s'applique aussi sur des images prises sur Terre, par exemple pour détecter des engins militaires dans des vues aériennes.

Cette méthode se base sur une mesure locale d'une dimension fractale de la surface d'intensité. La surface d'intensité est la surface d'une forme 3D dont les dimensions horizontales sont les deux dimensions de l'image et la hauteur représente le niveau de gris.

Une surface fractale, dans un espace en trois dimensions a une dimension fractale comprise entre 2 et 3. Dans le cas des images de milieux naturels, des mesures ont permis de déterminer que les milieux naturels ont généralement une dimension fractale comprise entre 2 et 2,5. Il est ainsi possible de détecter des objets artificiels en calculant la dimension fractale de différentes régions de l'image. Les régions ayant une dimension fractale inférieure à 2 ou

supérieure à 2,5 sont alors considérées comme des objets artificiels. Cette méthode a été d'abord proposée par Mark et Aronson [Mar-84]

La méthode proposée par Stein et Carlotto calcule la dimension fractale et l'erreur du modèle fractal sur des zones de 21x21 pixels. Pour chaque zone de l'image, on applique une série de dilatations et d'érosions en niveaux de gris, qui sont utilisées pour calculer le volume d'une couverture de la surface de l'intensité de l'image en fonction d'un paramètre d'échelle.

La dimension fractale est une fonction de la pente de régression linéaire du logarithme de l'aire de la surface et du logarithme du paramètre d'échelle r . L'erreur du modèle fractal est la moyenne des carrés des résidus de l'approximation linéaire à l'échelle r . Les objets dits artificiels, c'est à dire les zones d'intérêt correspondent aux zones de l'image ayant une dimension fractale inférieure à 2,0 ou supérieure à 2,5 et où l'erreur du modèle fractal est supérieure à 90%.

1.10 Méthode des orientations locales dominantes

Cette méthode proposée par Guérin-Dugué et al [Gué-01] n'a pas pour objectif principal de détecter la position d'une zone dans une image mais de catégoriser des images en scènes urbaines, scènes d'intérieur et scènes naturelles. Pour les scènes naturelles, on distingue les paysages « fermés » comme les montagnes et les forêts, caractérisés par le manque de profondeur de champ et les paysages « ouverts » caractérisés par une ligne d'horizon bien nette, comme les plages, les champs ou les déserts. Il ne s'agit pas d'une méthode de détection à proprement parler, mais il est cependant intéressant d'en parler ici, car elle permet de définir une structuration de l'image qui pourrait être utilisée pour la détection de zones d'intérêt. La catégorisation des images se fait en analysant leur forme spectrale dans l'espace de Fourier. Pour caractériser cette forme spectrale, il y a plusieurs approches possibles, parmi lesquelles les filtres passe-bande orientés et la méthode des orientations locales dominantes.

On peut analyser le spectre d'énergie des images en plaçant des filtres de type passe-bande orientée, comme des filtres de Gabor 2D, dans l'espace de Fourier pour capter la spécificité des formes spectrales des catégories à discerner. Ces filtres sont placés dans les basses fréquences spatiales, les hautes fréquences ne codant que les détails de l'image. La statistique globale sur chaque filtre orienté est uniquement relevée par la moyenne des activités dans l'image filtrée, l'organisation spatiale de l'image étant perdue. Ainsi, l'image est représentée comme un vecteur de N caractéristiques d'énergie.

Le principe de la méthode des orientations locales dominantes consiste à construire pour chaque image une carte d'orientations (orientation principale de chaque pixel au centre d'une région d'intérêt) pour ensuite former l'histogramme de la distribution des orientations spatiales dans l'image. La technique utilisée pour estimer les orientations locales dominantes est celle des filtres ajustables définis par Freeman et Adelson [Fre-91]. Le principe de ces filtres est de combiner les réponses de filtres dérivés de fonctions gaussiennes à l'ordre 2 pour estimer l'orientation locale dominante en chaque pixel et l'énergie locale à cette orientation. En faisant évoluer la largeur du filtre gaussien, on modifie en conséquence l'échelle de l'analyse de l'image.

L'orientation locale dominante correspond en fait à la direction du gradient d'intensité de niveaux de gris en un point. La distribution des orientations est analysée à travers les différentes échelles de l'image, des basses fréquences aux hautes fréquences, par l'histogramme des orientations locales dominantes, pondéré par l'énergie à cette orientation. Selon l'image, les résultats obtenus peuvent être très différents.

Sur les images représentant des villes ou des scènes d'intérieur, c'est à dire des environnements construits par l'homme, on observe surtout des orientations horizontales et verticales. Sur les scènes naturelles « fermées » ne présentant pas d'orientations privilégiées, on a des histogrammes qui présentent des distributions uniformes dans toutes les directions. Sur les scènes naturelles « ouvertes » caractérisées par la présence d'une ligne d'horizon, les effectifs de points sont plus importants sur les verticales, ce qui correspond à la ligne d'horizon. Cette méthode a permis de classer des images en 4 classes : scènes d'extérieur, d'intérieur, ouvertes et fermées. Pour détecter des zones d'intérêt à l'intérieur d'une image, il faut passer de l'approche statistique globale à une approche locale. Pour cela, une approche possible est d'utiliser un modèle statistique de la répartition spatiale de l'information chromatique.

1.11 Conclusion

Il existe de nombreuses méthodes permettant de détecter des régions d'intérêt dans les images. Le principe général de ces méthodes est d'utiliser une ou plusieurs caractéristiques extraites de l'image afin de caractériser une *région saillante* qui se distingue de l'arrière-plan de l'image. Parmi ces méthodes, un certain nombre d'entre elles sont basées sur une modélisation de la vision humaine, et plus particulièrement du fonctionnement du cortex visuel primaire. Ce type de méthode utilise plusieurs caractéristiques de bas niveau comme le contraste, les différences de teintes ou l'orientation des contours. Ces méthodes sont efficaces sur différents types d'images, à condition qu'elles présentent des régions de saillance qui contrastent nettement avec le fond de l'image. D'autres méthodes utilisent des caractéristiques structurelles différentes de celles utilisées par l'œil humain comme des dimensions fractales, des mesures d'entropies ou le résultat d'une transformée en ondelettes. Ces méthodes sont souvent très efficaces pour certains types d'images particuliers, comme celles présentant une faible profondeur de champ ou un fond de teinte uniforme, mais elles le sont nettement moins dans d'autres cas. Les méthodes les plus efficaces sont sans doute celles qui utilisent plusieurs caractéristiques et qui les combinent, comme le fait naturellement l'œil humain, ce qui lui permet de détecter les régions d'intérêt dans les images les plus variées.

2 *Détection d'objets artificiels dans les milieux naturels*

Nous allons maintenant tenter d'appliquer les modèles de répartition statistique en loi puissance des motifs de l'image à un problème de détection d'objets dans une image. Nous nous intéresserons d'abord à un cas particulier, celui de la détection d'objets artificiels dans un environnement naturel. L'application envisagée à l'origine pour la méthode que l'on va étudier ici est la détection de mines antipersonnel ou antichar et de munitions abandonnées sur un champ de bataille.

2.1 **Position du problème**

Il existe de nombreux types de mines terrestres. Une première classification peut être faite selon le type de cible : antipersonnel, antivéhicule (destinée à détruire des engins non blindés) ou antichar, étant donné que certaines mines peuvent convenir à plusieurs types de cibles. Ces mines peuvent être enterrées dans le sol à plus ou moins grande profondeur, posées sur le sol, camouflées ou non par la végétation environnante, placées en hauteur par exemple dans un arbres Les modes d'action et de déclenchement de ces mines sont également très divers. En plus des mines, on peut trouver sur un champ de bataille toutes sortes de munitions non explosées. Toutes ces mines et autres engins explosifs se présentent sous des apparences très diverses, comme on peut le voir sur la figure ci-dessous.



Fig. 3 : Quelques exemples de mines antipersonnel

La détection des mines est plus ou moins difficile en fonction du type de mine et de la nature du terrain. Dans le cadre d'une détection visuelle, on se limitera essentiellement à des mines posées au sol, et détectées depuis une position terrestre. Certes, on ne peut pas détecter ainsi les mines enterrées, mais il existe de nombreux types de mines, parmi les plus dangereuses et

les plus utilisées aujourd'hui, qui sont visibles au sol, notamment les mines déclenchées à distance ainsi que les mines dispersées par voie aérienne ou par des tirs d'artillerie.

Les objets que l'on cherche à détecter ont des caractéristiques physiques et une apparence très différentes les uns des autres. De plus, dans une situation réelle ces objets peuvent ne pas être visibles en totalité, ils peuvent avoir été camouflés ou être cachés par des éléments de leur environnement. Ils peuvent également avoir été endommagés lors de leur chute, ce qui peut rendre leur identification plus difficile. Pour toutes ces raisons il est préférable non pas de chercher à identifier un type d'objet particulier, mais de rechercher des caractéristiques visuelles communes à l'ensemble des objets à traiter.

Tous ces objets ont pour point commun d'être des objets artificiels, c'est à dire des objets manufacturés construits par l'homme dont on recherche la présence dans un environnement naturel, composé d'éléments végétaux et minéraux tels que de l'herbe, du sable, de la terre, etc... C'est pourquoi on va s'attacher à distinguer ce qui différencie, du point de vue des caractéristiques visuelles, les objets naturels des objets artificiels. Parmi ces caractéristiques, on peut notamment citer la forme, la couleur et les matériaux.

En ce qui concerne la forme, les objets artificiels présentent en général des formes géométriques simples et régulières, comme des rectangles, des cercles ou des cylindres. Ils se caractérisent notamment par la présence fréquente de lignes droites ou de courbes régulières. Les contours de ces objets sont en général très nets et se détachent nettement du fond. A l'inverse, les objets naturels n'ont pas en général de formes régulières identifiables ni de contours aussi marqués. Il faut cependant garder à l'esprit que les objets peuvent être camouflés, et l'on sait que le camouflage a pour but de casser les formes.

Les objets artificiels présentent également souvent des zones ayant des couleurs uniformes sur une partie importante de leur surface, contrairement aux objets naturels dont les couleurs sont beaucoup plus variées. De plus, les objets artificiels sont en général fabriqués dans des matériaux ayant une texture de surface homogène, comme le métal, le plastique ou le bois travaillé et peint. L'objet artificiel peut être constitué de plusieurs matériaux, mais chacun de ces matériaux a une texture relativement homogène. Une telle homogénéité de surface ne se rencontre habituellement pas dans le cas des objets naturels. D'une manière générale, un objet artificiel se distingue donc par la présence de surfaces homogènes délimitées par des contours francs. Ces propriétés se manifestent même dans le cas où l'objet a été camouflé. A partir de là, on peut concevoir essentiellement deux manières de détecter un objet : en recherchant le contour de l'objet ou en s'intéressant à l'uniformité de sa surface.

Une approche classique pour détecter les objets artificiels serait d'extraire tous les objets de l'image. Une approche classique de détection de contours ne serait pas efficace pour détecter les objets artificiels, en effet si l'on extrait tous les contours présents dans l'image, l'interprétation de la scène devient difficile, en effet il n'est pas facile d'identifier quels sont les contours qui correspondent aux objets artificiels que l'on veut détecter, si l'on ne dispose pas de connaissances préalables sur la structure de la scène. Si par contre on sélectionne seulement les contours les plus importants, on risque au contraire de perdre trop d'information pour pouvoir interpréter la scène. Les méthodes de détection de régions sont également difficiles à appliquer pour les mêmes raisons, les caractéristiques qui définissent une région étant souvent trop restrictives ou trop vagues, ce qui fait qu'on détecte trop de régions ou trop peu, ce qui rend difficile l'interprétation de la scène avec une bonne fiabilité.

2.2 Application des modèles de lois puissance

L'utilisation des modèles de lois puissance pourrait être une solution pour résoudre ce problème. En effet, l'application de ces modèles permet d'évaluer la complexité structurelle d'une image. Or, l'image d'un objet artificiel a une structure différente de celle d'un objet naturel, sa texture est plus uniforme et ses contours sont plus marqués. Les différences entre les deux types d'objets doivent être visibles sur la courbe de Zipf. Etant donné que l'on s'intéresse ici essentiellement à la distribution des contours et de la texturation fine de l'image, on utilisera le codage des rangs généraux pour la construction des courbes de Zipf.

Avant d'utiliser la courbe de Zipf pour détecter les mines, il convient de vérifier que l'on peut effectivement distinguer à l'aide des caractéristiques de la courbe de Zipf, les images représentant des milieux artificiels de celles représentant des milieux naturels.

La figure ci-dessous montre un exemple de comparaison des courbes de Zipf obtenues à partir de deux images, l'une représentant un milieu naturel, l'autre un milieu artificiel. Les deux images ont une même taille de 256 x 256 pixels et les courbes sont obtenues avec le codage des rangs généraux.

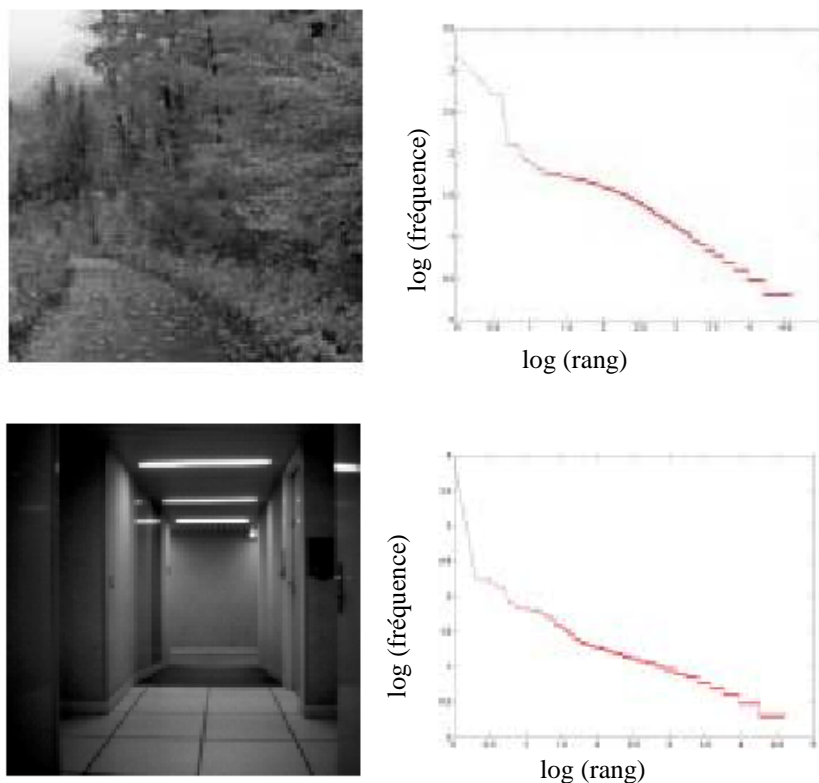


Fig. 4 : Courbes de Zipf associées à l'image d'un milieu naturel (en haut) et d'un milieu artificiel (en bas)

Par rapport à une image naturelle, on remarque que la pente à l'origine de la courbe de l'image représentant un milieu artificiel est plus élevée, mais en revanche la courbe est beaucoup plus plate dans sa partie droite et donc la pente moyenne est plus faible. L'aire sous la courbe est également plus faible pour l'image du milieu artificiel, et le nombre de motifs différents pris en compte pour la construction de la courbe est également plus petit. Dans les deux images présentées sur la figure, on constate une sur-représentation du motif le plus fréquent, à savoir le motif uniforme. Cela est dû dans les deux cas à la présence de zones

totale­ment uniformes sur une petite partie de l'image, et qui ne sont pas représentatives de la structure du reste de l'image. D'une manière générale, les courbes de Zipf sont différentes selon que l'image représente un milieu naturel ou artificiel, aussi bien en ce qui concerne l'aire sous la courbe que la pente de celle-ci.

Si la courbe de Zipf donne une information sur la structure de l'image, en revanche par elle-même elle ne permet pas de déterminer la position d'un objet à l'intérieur de l'image. C'est pourquoi il est nécessaire de segmenter l'image en régions, avant de calculer les courbes de Zipf associées à chacune de ces régions, afin de les comparer entre elles pour pouvoir détecter celles qui sont associées à un objet artificiel.

Etant donné que la taille de l'image a une influence sur les caractéristiques de la courbe de Zipf, il est préférable de travailler sur des régions de même taille pour que la comparaison des courbes soit pertinente. C'est pourquoi l'image sera découpée en im­a­gettes rectangulaires de taille égale, et l'on calculera la courbe de Zipf associée à chacune des im­a­gettes. Les courbes de Zipf seront ensuite comparées entre elles en utilisant différents critères de détection.

Un problème qui se pose pour la comparaison des courbes est que la courbe de Zipf, du fait qu'elle est représentée en échelle logarithmique, possède beaucoup plus de points en partie droite qu'en partie gauche. Or, l'information la plus pertinente pour différencier deux courbes de Zipf se situe au contraire au début de la courbe. C'est pourquoi on ne peut pas utiliser les courbes de Zipf directement. Afin de réduire au minimum le nombre de points pris en compte dans le calcul il est donc nécessaire de procéder à un rééchantillonnage. Le principe du rééchantillonnage est de ne conserver que certains points régulièrement espacés sur la courbe. Le pas d'échantillonnage est fixé au départ. L'algorithme d'échantillonnage fonctionne de la manière suivante : soit $x(k)$ et $y(k)$, l'abscisse et l'ordonnée du $k^{\text{ième}}$ point de la courbe pour k compris entre 1 et n , n étant le nombre de points de la courbe. On calcule l'ordonnée de chaque point d'abscisse i selon la formule :

$$y(i) = \left[\frac{y(j+1) - y(j)}{\log_{10}(j+1) - \log_{10}(j)} (i - \log_{10}(j)) \right] + y(j) \quad (2-2-1)$$

où j est l'entier inférieur à l'inverse du logarithme décimal de i . On effectue ce calcul sur toute la courbe en incrémentant i à chaque itération du pas d'échantillonnage. Le pas d'échantillonnage a ici été fixé à 0,05. La figure ci-dessous montre le résultat de l'échantillonnage d'une courbe de Zipf.

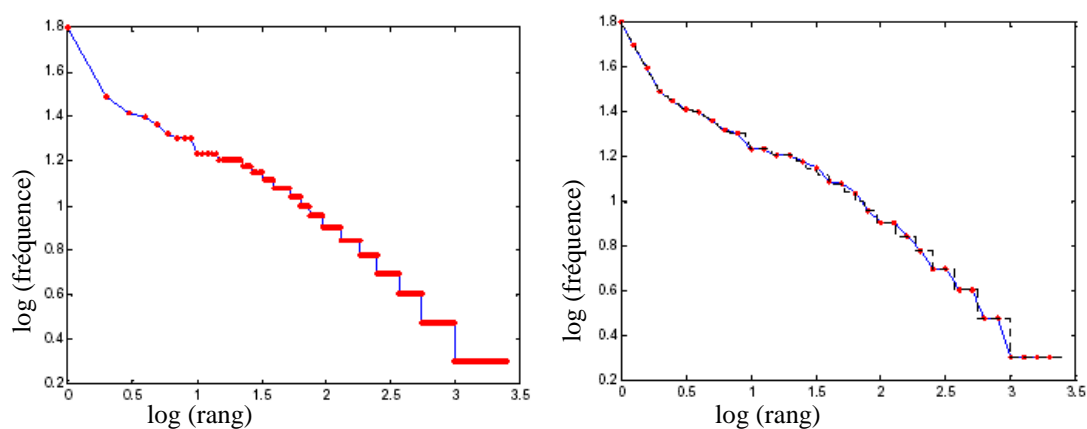


Fig. 5 : Courbe de Zipf originale (à gauche) et rééchantillonnée (à droite)

Une fois les courbes de Zipf calculées, il reste maintenant à les comparer entre elles afin de pouvoir détecter l'objet. Pour cela, il faut définir des critères de détection qui seront calculés à partir des courbes de Zipf afin de déterminer le caractère naturel ou artificiel des objets présents dans l'image. Les critères utilisés seront des critères numériques pour lesquels on calculera une valeur par imagerie. Les résultats de l'évaluation des critères pourront être représentés sous forme de surfaces 3D où les coordonnées horizontales représentent la position spatiale de l'imagerie dans l'image, et l'altitude représente la valeur du critère pour cette imagerie. Les critères seront définis de manière à ce que l'on obtienne un maximum sur les imageries qui représentent des objets artificiels, ce qui se traduira par la présence d'un pic sur la surface 3D.

Un premier critère que l'on peut définir est l'aire délimitée par la courbe de Zipf. Cette aire est d'autant plus faible que l'image est uniforme et donc artificielle. L'aire sous la courbe est calculée pour chaque imagerie, en utilisant la méthode des trapèzes. Soit n le nombre de motifs de la courbe, f_i la fréquence et r_i le rang du motif i , l'aire de la courbe est donnée par la formule suivante :

$$A = \sum_{i=1}^{n-1} \frac{(f_i + f_{i+1})(r_{i+1} - r_i)}{2} \quad (2-2-2)$$

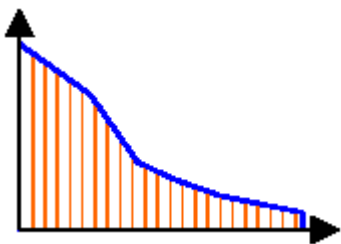


Fig. 6 : Représentation de l'aire sous la courbe de Zipf.

Comme l'aire de la courbe tend à diminuer quand l'image devient artificielle, on affichera en fait $-A$ au lieu de A , afin d'avoir un maximum à l'endroit où l'image est la plus uniforme sur la représentation graphique.

Un autre critère de détection est la pente de la courbe. La pente de la courbe est la pente moyenne calculée en utilisant une interpolation linéaire par la méthode des moindres carrés. La pente moyenne de la courbe est le coefficient directeur de la droite des moindres carrés. Elle est donnée par la formule suivante.

$$P = \frac{n \sum_{i=1}^n y_i x_i - \sum_{i=1}^n y_i \sum_{i=1}^n x_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2} \quad (2-2-3)$$

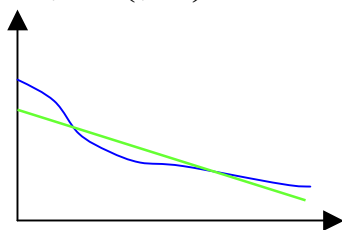


Fig. 7 : Représentation de la pente de la courbe de Zipf.

Les premiers essais effectués avec le critère de l'aire sous la courbe ont montré que la position de l'objet à détecter ne correspond pas toujours au pic d'amplitude maximum dans toute l'image, mais plutôt à celui qui est le plus haut par rapport aux imagerie voisines. D'où l'idée de définir comme critère de détection la différence entre l'aire de la courbe et celle des imagerie voisines. La différence avec la moyenne des aires de la courbe est calculée pour chaque imagerie en prenant la valeur absolue de la différence entre l'aire de l'imagerie et la moyenne des aires des imagerie voisines. Ce critère est calculé de façon à ce que l'on ait un pic pour l'imagerie dont la différence entre son aire et celle des imagerie voisines est la plus importante. L'aire de la courbe est celle calculée par la formule précédente. La formule suivante donne la valeur du critère pour l'imagerie repérée par (i,j).

$$D = \left| A_{ij} - \left(\frac{1}{9} \sum_{k=i-1}^{i+1} \sum_{l=j-1}^{j+1} A_{kl} \right) \right| \quad (2-2-4)$$

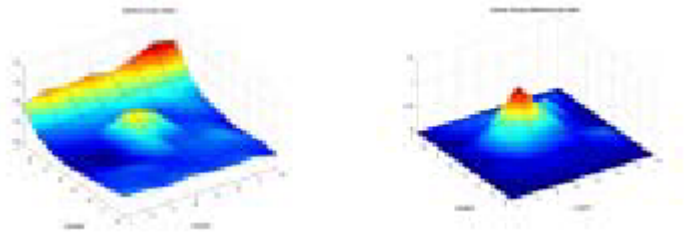


Fig. 8 : Représentation de la différence avec la moyenne des aires.

Il est également possible de calculer ce type de critère pour les autres caractéristiques comme la pente de la courbe.

Un autre critère de détection que l'on peut utiliser est un coefficient d'uniformité de la courbe. Le coefficient d'uniformité est la somme des pentes au voisinage de chaque point de la courbe. Il s'agit donc d'une mesure locale de la pente, contrairement à la mesure globale calculée par la formule (2-2-3). Le coefficient d'uniformité est calculé avec la formule :

$$C = - \sum_{i=2}^{n-1} \frac{f_{i+1} - f_{i-1}}{r_{i+1} - r_{i-1}} \quad (2-2-4)$$

Comme la valeur de la somme diminue quand l'uniformité de l'imagerie augmente, on calcule le résultat avec un signe moins afin que l'on ait un maximum pour l'imagerie la plus uniforme.

Enfin, le dernier critère utilisé est le nombre n de motifs différents par imagerie qui apparaissent plus d'une fois dans l'image.

2.3 Résultats et interprétation

La méthode définie précédemment va maintenant être testée sur des images contenant des objets artificiels dans un environnement naturel. On dispose pour ce faire d'images de test

représentant différents types de mines antipersonnel ou antichar dans des milieux naturels variés. Il s'agit d'images haute résolution en niveaux de gris non compressées, dont la taille varie entre 1500x1000 à 2000x1500 pixels. Des essais complémentaires ont également été effectués sur des images prises à l'aide d'un appareil photo numérique, qui représentent des objets divers posés sur le sol. Ces images ont une taille de 2048x1536 pixels et sont compressées au format JPEG avec un taux de compression faible, l'appareil utilisé ne permettant pas d'enregistrer des images non compressées. La figure 9 représente les résultats de l'évaluation des différents critères de détection sur une image de test représentant une mine antipersonnel posée dans l'herbe.

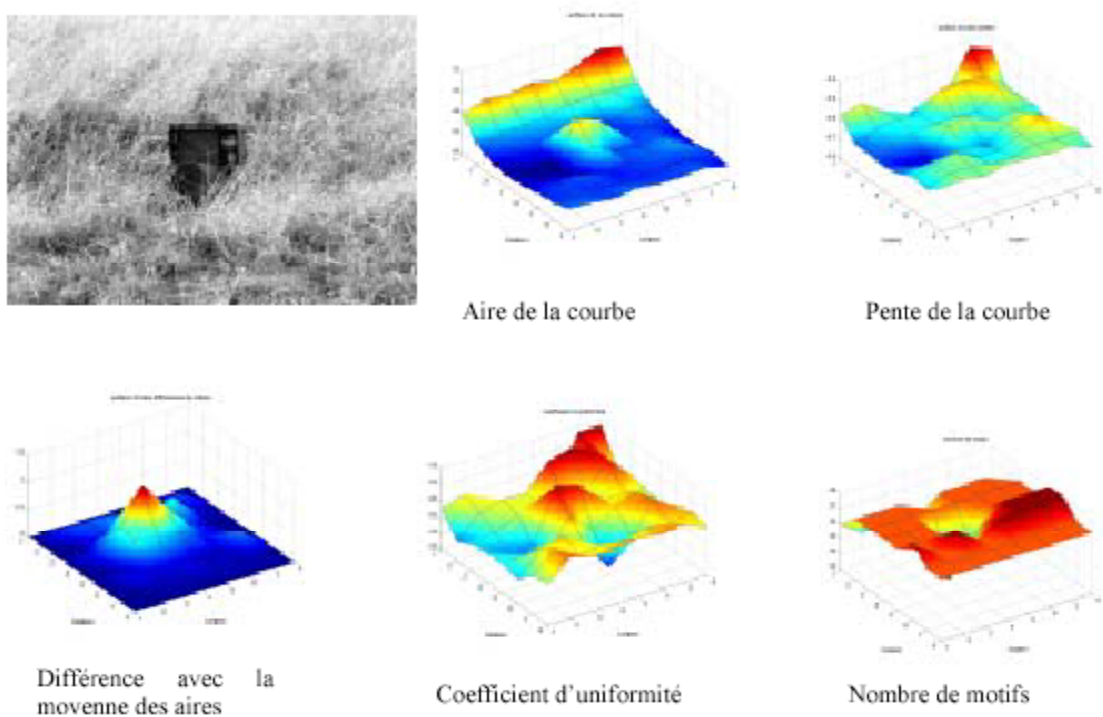


Fig . 9 : Exemple de résultat sur une image représentant une mine

On observe sur le graphe du critère de l'aire sous la courbe une surface uniforme en haut de l'image. Cette surface correspond au fond de l'image qui apparaît flou en raison de la profondeur de champ limitée de l'appareil photographique utilisé, en effet les photos sont prises avec un téléobjectif à longue focale. Cette surface apparaît comme uniforme au niveau de la courbe de Zipf. On observe également un pic isolé à l'emplacement de la mine, mais ce pic est moins élevé que celui qui correspond au fond de l'image. On voit bien ici que la position du pic le plus élevé (qui correspond en fait au minimum de l'aire) ne permet pas à elle seule de détecter la mine.

Le critère de la pente de la courbe est plus difficile à interpréter. L'observation de ce critère ne permet pas facilement de repérer la position de la mine. La représentation graphique ne présente pas véritablement de signe distinctif à l'endroit où se trouve la mine. Par contre on peut apercevoir un pic au fond de l'image. Le critère du coefficient d'uniformité donne des résultats analogues au précédent, en effet il est lui aussi basé sur une mesure de la pente. Cela est dû au fait que sur des imagerie de taille relativement petite, la courbe de Zipf présente des discontinuités qui font que le calcul de la pente moyenne de la courbe de Zipf n'est pas une caractéristique très pertinente.

Le critère de la différence avec la moyenne des aires donne de bien meilleurs résultats. Le calcul des différences permet d'éliminer l'uniformité parasite due à la profondeur de champ et

de renforcer le pic correspondant à l'objet à détecter. Avec cette méthode, le pic le plus élevé correspond bien à la position de la mine. Ce critère est donc le plus efficace pour détecter ce type d'objet, et c'est le seul qui permette une détection automatique sans avoir à interpréter visuellement le graphique.

Enfin, en ce qui concerne le nombre de motifs, on remarque un trou à l'emplacement de la mine, ce qui confirme le fait que l'image d'un objet artificiel contient moins de motifs qu'une image naturelle.

Le critère de la différence avec la moyenne des aires est donc celui qui présente les meilleures performances en matière de détection, en ce sens que c'est le seul pour lequel on a la différence la plus significative entre les imagerie contenant un objet artificiel et les autres. Par rapport à une image ne contenant pas d'objet, l'amplitude du maximum est beaucoup plus importante dans le cas d'une image comportant un objet artificiel. Le rapport entre la valeur du maximum de la différence avec la moyenne des aires D_{ij} et la moyenne des valeurs de ce critère sur l'ensemble des imagerie permet de signaler la présence ou non d'un objet artificiel dans la scène : si la valeur de ce rapport est supérieure à 9, l'image contient un objet artificiel.

L'exemple de la figure 10 montre encore davantage l'intérêt de la méthode utilisée, et son efficacité à distinguer le naturel de l'artificiel. Cette image représente une mine antichar dans le désert, sur un fond de sable et de cailloux. Notre méthode permet de détecter la mine malgré la présence de cailloux qui pourraient être confondus avec la mine si l'on utilise une méthode plus classique à base de détection de contours. En effet, même si les cailloux ressemblent à première vue à la mine, la répartition des motifs à leur surface est différente, ce qui permet de reconnaître, parmi tous les objets présents dans l'image, l'objet artificiel à détecter.

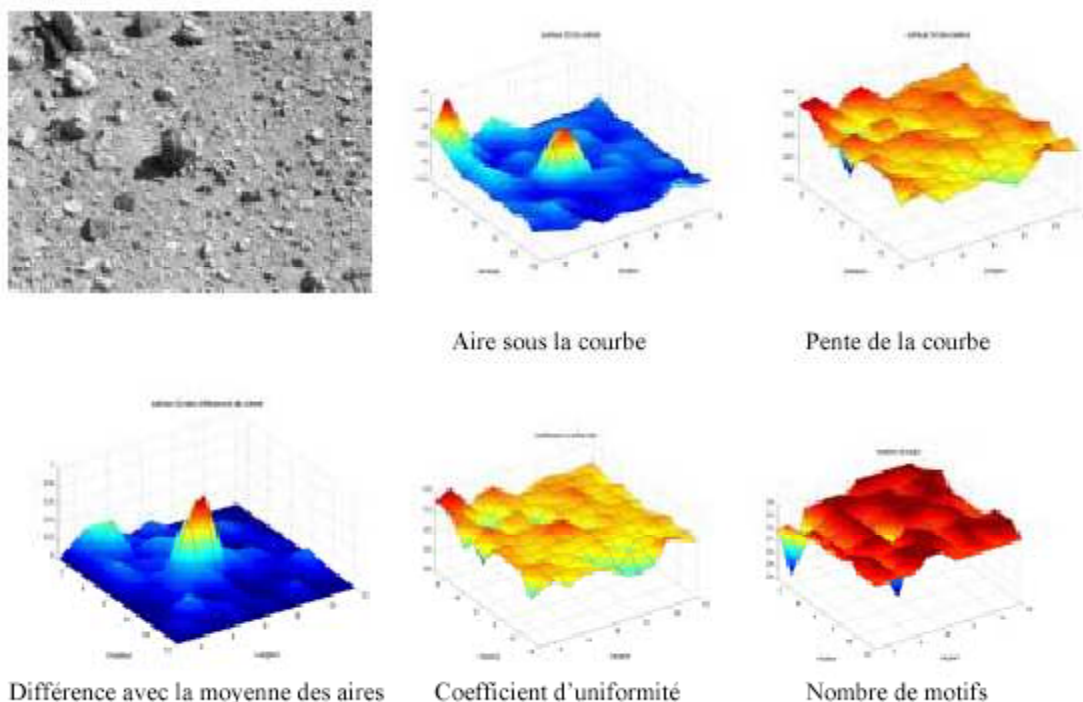


Fig. 10 : Détection d'un objet artificiel au milieu d'autres objets

Sur d'autres images, on trouve au lieu du maximum attendu un minimum de l'aire sous la courbe et de la pente de la courbe, comme c'est le cas de l'objet de la figure ci-dessous. Sur cette image, l'objet à détecter, cadré de près, présente une surface qui comporte de faibles

variations continues de niveaux de gris, sans aucun contour à l'intérieur de l'objet. A l'inverse, les textures naturelles qui entourent l'objet sont riches en contours. Sur cette image, les motifs les plus fréquents sont en fait ceux qui correspondent aux contours, et l'objet artificiel se distingue du reste de l'image par son absence de contours. Le critère de la différence avec la moyenne permet de détecter l'objet, en effet on s'intéresse ici à l'existence d'une différence entre l'objet et son environnement plutôt qu'aux caractéristiques propres de l'objet.

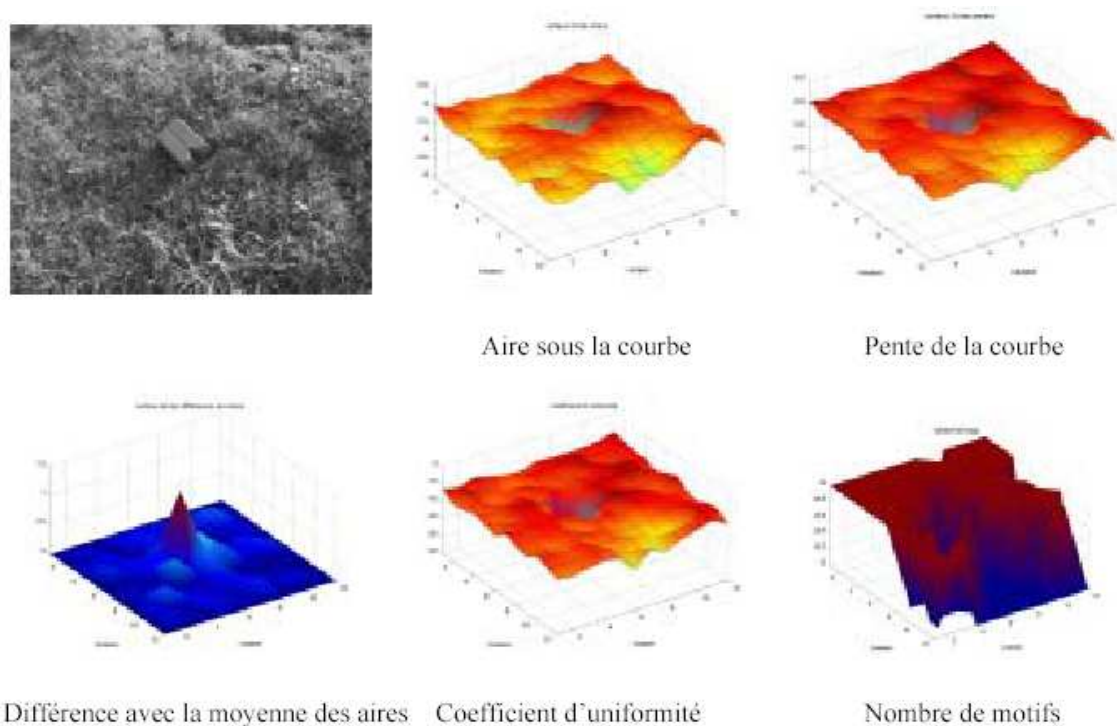
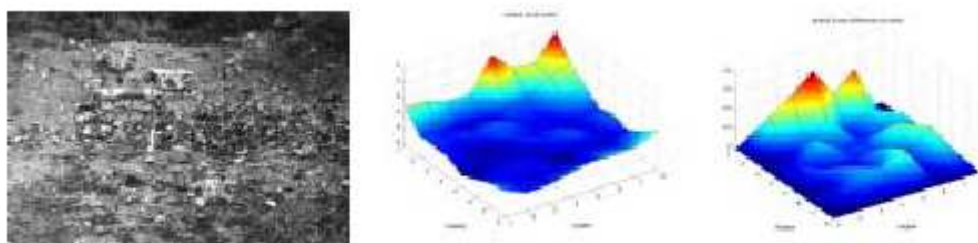


Fig. 11 : exemple de résultat sur un objet plus uniforme que le fond

Cependant, cette méthode de détection n'est pas assez efficace sur toutes les images. En effet, sur certaines images, on ne parvient pas à détecter l'objet à l'aide des courbes de Zipf alors qu'il est parfaitement visible sur l'image. La figure 12 montre quelques exemples pour lesquels l'objet n'a pas pu être détecté correctement. Pour ces images, on a représenté uniquement les critères de l'aire sous la courbe et de la différence des aires.



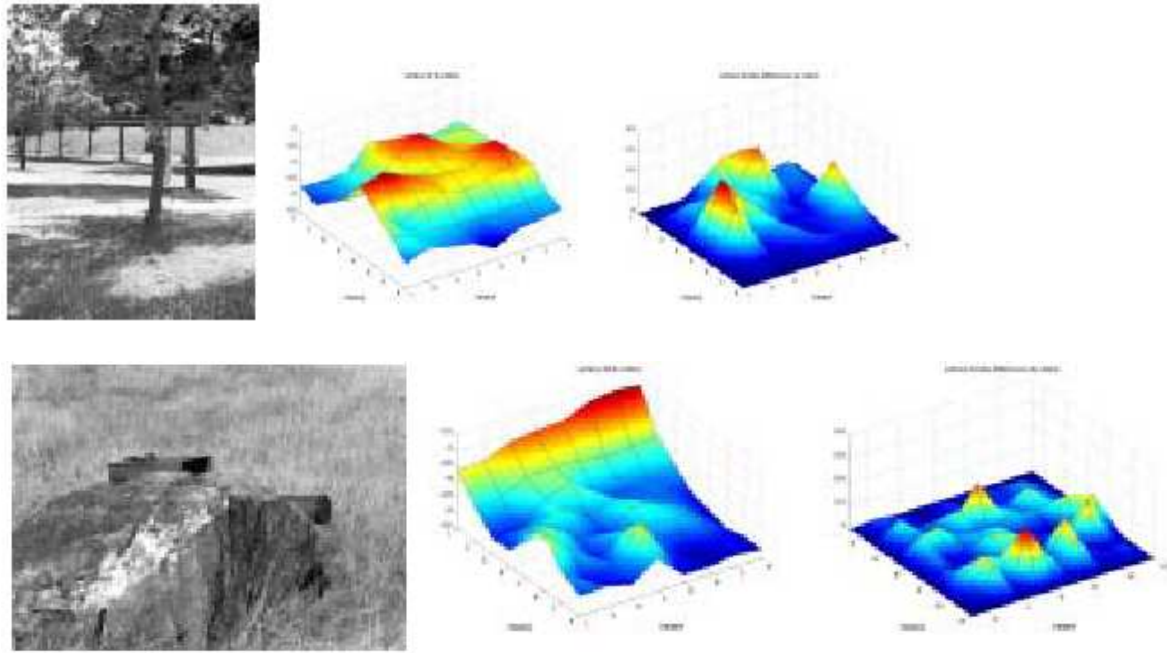


Fig. 12 : Exemples où l'objet n'est pas correctement détecté.

Sur la première image, on a un objet camouflé sur un fond constitué de cailloux, l'objet qui apparaît constitué de plusieurs petites régions de couleurs différentes ne se distingue pas suffisamment du fond pour pouvoir être détecté. Sur la seconde image, l'objet est dans un arbre, et on détecte en fait les zones claires situées derrière l'arbre qui apparaissent uniformes sur l'image. Sur l'image du bas, où l'objet est derrière un rocher, on observe un pic parasite qui correspond à la bande claire sur le rocher. Dans les trois cas, la détection échoue à cause de la présence à proximité de l'objet à détecter de surfaces uniformes d'origine naturelle dont la texture ressemble à celle d'un objet artificiel et l'on ne parvient pas à distinguer l'objet du fond.

2.4 Influence de la résolution de l'image

Afin d'étudier l'influence de la segmentation de l'image sur la détection des objets, des essais ont été effectués sur les images avec différents découpage, en 8x8, 12x12 et 16x16 imagerie. Les figures ci-dessous montrent des essais effectués sur l'image de la figure 9 avec un découpage en 12x12 et 16x16 imagerie.

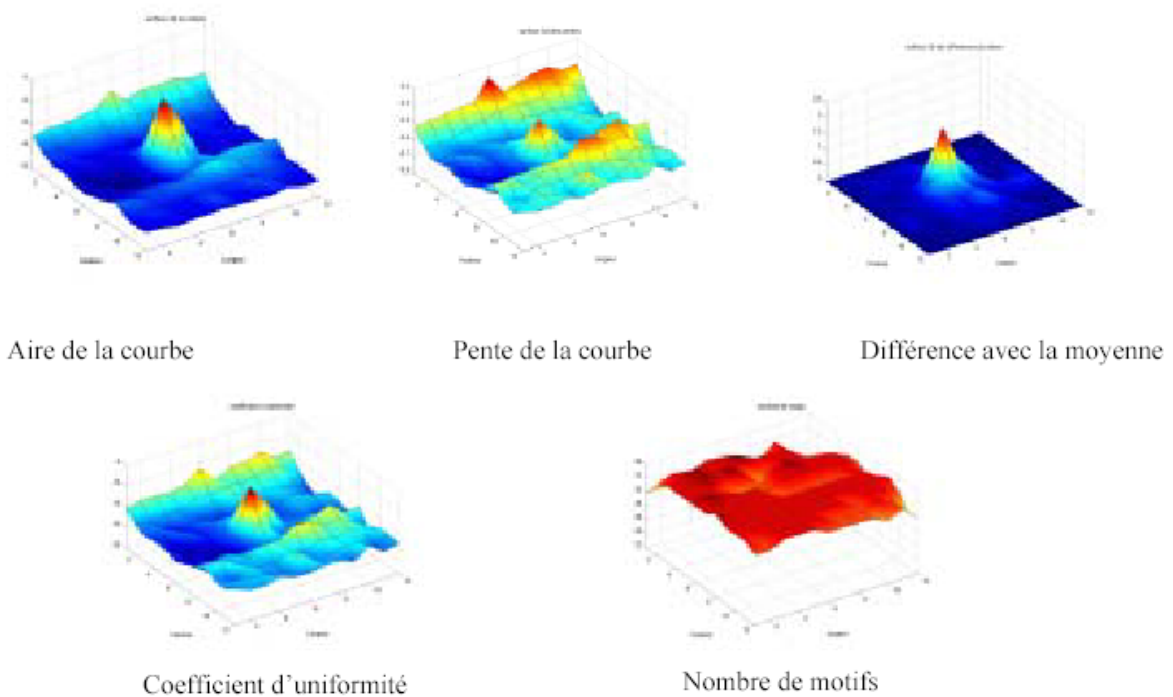


Fig. 13 : Résultats de la détection avec un découpage en 12x12 imagettes

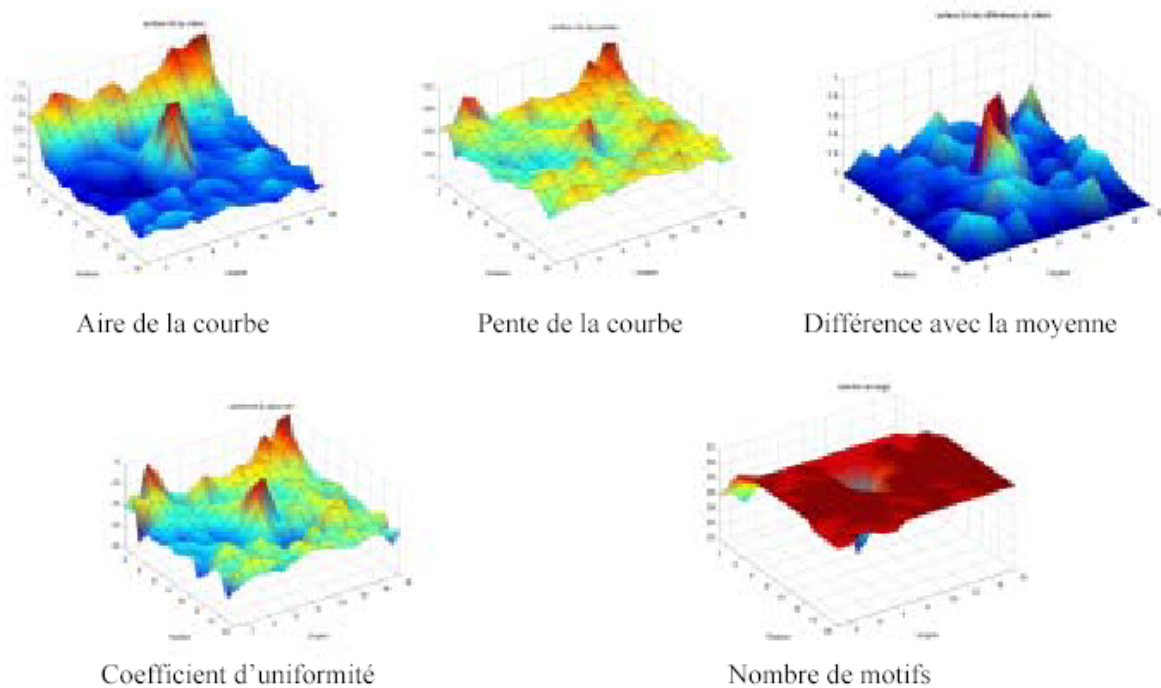


Fig. 14 : Résultat de la détection avec un découpage en 16x16 imagettes

Sur l'image découpée en 12x12 imagettes, on remarque que le pic correspondant à la mine est renforcé sur le critère de l'aire sous la courbe et qu'en revanche l'influence de la profondeur de champ est nettement diminuée avec le nouveau découpage. Le pic est également renforcé pour le critère de la différence avec la moyenne et il apparaît nettement pour la pente de la courbe et le coefficient d'uniformité alors qu'il n'apparaît pas sur le découpage 8x8. On voit donc ici qu'un découpage plus fin de l'image permet d'améliorer la détection. Ceci peut s'expliquer par le fait qu'un découpage en imagettes plus petites permet de mieux séparer les différents objets présents dans l'image ; avec un découpage 12x12 on a

en effet des imagerie qui sont entièrement occupées par la mine, ce que l'on n'avait pas avec le découpage précédent. En revanche, on ne distingue plus l'objet avec le nombre de motifs. Ceci est dû au fait que les imagerie sont plus petites et contiennent par conséquent moins de motifs. Il n'est alors plus possible de distinguer l'artificiel du naturel par ce critère. Ceci fixe une limite au nombre d'imagerie du découpage : un trop grand nombre de petites imagerie ferait que les courbes de Zipf de chaque imagerie contiendraient trop peu de motifs pour être utilisables. C'est le cas sur cette image, où un découpage 16x16 donne de moins bons résultats que le découpage 12x12. On observe en effet l'apparition de pics parasites qui sont dus à la réduction de la taille des imagerie et donc du nombre de motifs, on a alors des surfaces qui apparaissent comme uniformes alors qu'elles ne correspondent pas à un objet.

Le choix de nombre d'imagerie est donc un compromis entre deux exigences contradictoires, des imagerie trop peu nombreuses ne permettent pas une détection précise de la position d'un objet, mais un nombre trop important d'imagerie a pour conséquence qu'elles contiennent trop peu de motifs pour que la répartition des fréquences des motifs de l'imagerie soit significative quant au contenu de l'image. Dans notre cas, pour des images dont la taille est comprise entre 1024x768 et 2500x1500 pixels, un découpage en 12x12 imagerie donne en général les meilleurs résultats.

La détection d'un objet à l'aide de cette méthode nécessite que l'objet occupe une surface suffisamment importante dans l'image. En effet un objet trop petit n'aurait qu'une incidence très faible sur la courbe de Zipf de l'imagerie et ne serait pas détecté. Afin de déterminer la surface minimale que l'objet doit occuper pour espérer être détecté, des essais ont été effectués sur des images d'objets prises à des distances différentes, ce qui se traduit par une surface apparente de l'objet plus ou moins grande. La figure ci-dessous montre un exemple de résultat d'un de ces essais.

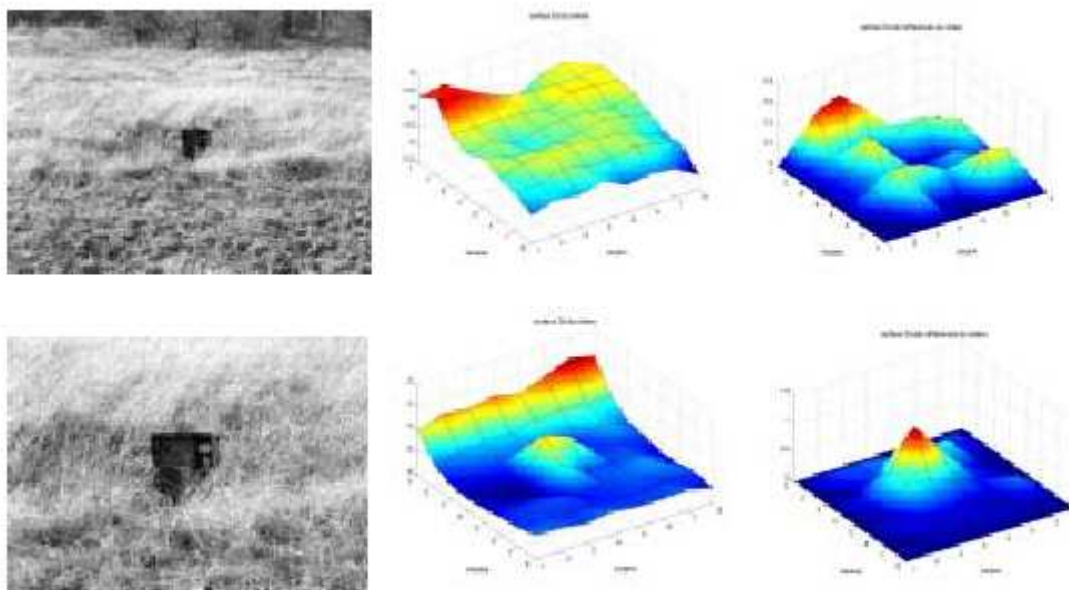


Fig. 15 : Deux images d'un même objet prises à deux distances différentes

Sur la première image, l'objet occupe 0,88% de la superficie totale de l'image et il n'est pas détecté. Sur la seconde, il occupe 4,86% de la surface de l'image et il est détecté. D'une manière générale, les essais effectués ont montré que l'objet doit occuper au minimum 2 % de la superficie totale de l'image pour pouvoir être détecté.

Le calcul des courbes de Zipf est une opération assez longue, surtout sur les images de grande taille. C'est pourquoi il peut être intéressant de travailler sur des images plus petites afin de gagner en temps de calcul. Il faut cependant vérifier que la méthode de détection reste utilisable sur des images de résolution plus faible.

Afin d'évaluer l'influence de la résolution de l'image sur la qualité de la détection, un nouveau paramètre d'évaluation a été défini. Ce paramètre est calculé pour le critère de la différence avec la moyenne des aires défini au chapitre précédent. Il s'agit du rapport entre le maximum du critère sur les imagerie qui contiennent un objet artificiel et le maximum sur les imagerie qui n'en contiennent pas. L'objet est considéré comme détecté si ce rapport est supérieur à 1. Plus la valeur de ce paramètre est élevée et plus la détection est bonne. Les essais ont été effectués avec différents rapports de réduction. Les images ont été réduites successivement à 75%, 66%, 50%, 33% et 25% de leur taille initiale. Les découpages en 8x8 et 12x12 imagerie ont été utilisés. Pour chacun des essais on a calculé le paramètre de détection. Cependant, un problème se pose si les imagerie sont trop petites : il se peut que des imagerie ne contiennent qu'un seul motif, et le programme de détection ne fonctionne pas si la courbe de Zipf est réduite à un unique point. Cela explique l'absence de résultats pour les rapports de réduction les plus importants pour le découpage 12x12. La figure ci-dessous montre les résultats obtenus avec différents objets. Pour chaque objet, les courbes donnent la valeur du paramètre de qualité de détection en fonction du rapport de réduction de l'image.

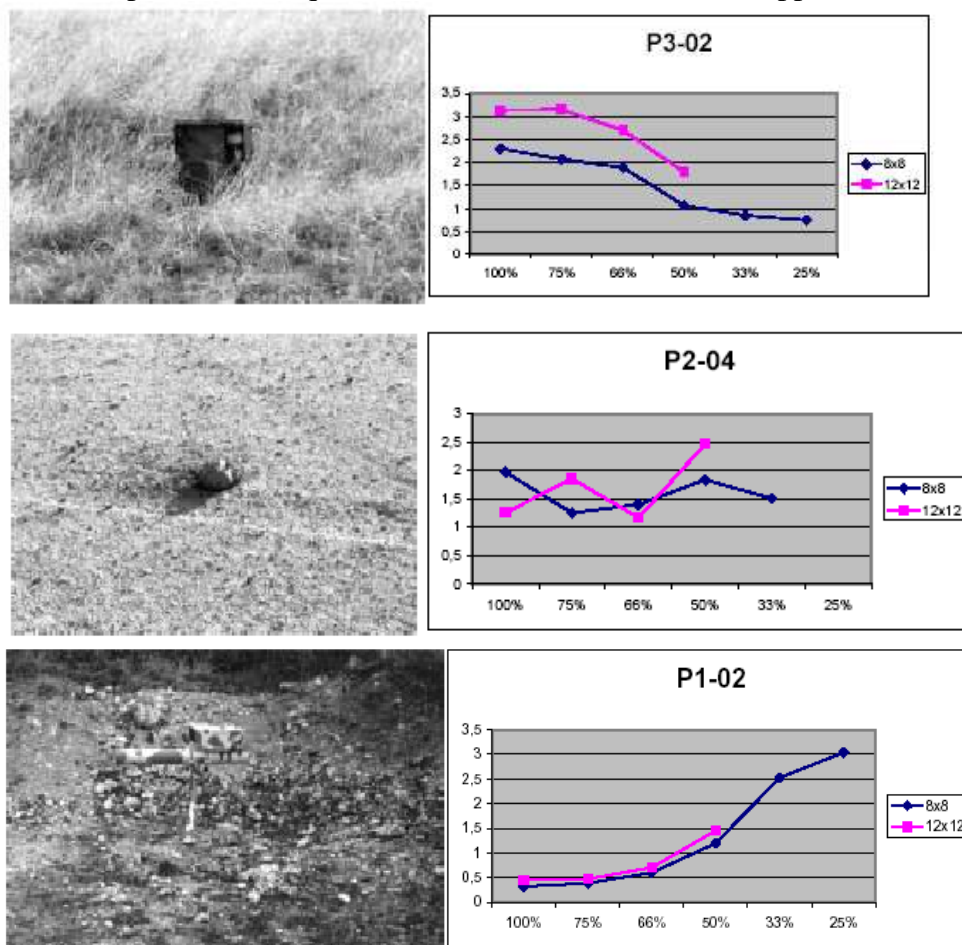


Fig . 16 : Influence de la résolution de l'image sur la détection de quelques objets

Au vu de ces résultats, on remarque que l'influence de la résolution est variable selon les images, et qu'il est difficile de déterminer à priori l'effet que pourra avoir un redimensionnement de l'image sur le résultat de la détection. Dans certains cas, une réduction de la taille de l'image diminue la qualité de la détection, mais dans d'autres elle est au contraire améliorée, et sur quelques images il est même possible de détecter l'objet sur une image réduite alors que cela ne l'est pas sur l'image d'origine. Dans d'autre cas enfin, la variation de la qualité de l'image en fonction de la résolution n'est pas monotone.

En règle générale, une résolution plus faible a globalement pour effet de rendre l'image plus uniforme, mais toutes les textures ne réagissent pas de la même façon à la diminution de la résolution. Dans certains cas, on diminue les différences d'uniformité entre l'objet et le fond, ou on fait ressortir des uniformités parasites, ce qui se traduit par une diminution des performances de la détection. Dans d'autres cas, on accentue ces différences, ce qui augmente la qualité de la détection par rapport à l'image initiale. Un autre effet du redimensionnement est de modifier le rapport entre l'aire et le périmètre des régions de l'image. En réduisant l'image, on augmente ainsi l'importance des contours par rapport aux régions. Ceci permet dans certains cas de détecter plus facilement des objets grâce à leurs contours. Les résultats de la détection peuvent dans certains cas dépendre du découpage initial. On voit donc l'intérêt de travailler en multirésolution et de faire plusieurs essais avec différentes résolutions d'images, afin de pouvoir choisir la bonne résolution qui permettra de détecter l'objet.

La méthode de détection a été essayée sur une base de test de 40 images, comprenant d'une part des photographies hautes résolution de mines antipersonnel et antichar, et d'autre part des images d'objets divers prises avec un appareil photo numérique.

La détection d'objets par la courbe de Zipf en utilisant le codage des rangs généraux donne des résultats assez moyens dans l'ensemble, l'objet n'a pu être détecté que dans 56 % des images testées. C'est pourquoi on va chercher à améliorer cette méthode. Plusieurs possibilités s'offrent à nous pour tenter d'améliorer les résultats, à savoir l'utilisation d'un prétraitement ou d'un autre codage de l'image.

2.5 Utilisation d'un prétraitement

Afin d'améliorer les résultats, on va maintenant appliquer un prétraitement à l'image à tester avant d'effectuer la détection d'objets par la courbe de Zipf. Le but de ce prétraitement est d'éliminer l'information inutile de l'image pour ne garder ce qui est pertinent pour la détection de l'objet. Les objets artificiels se caractérisent par des contours très nets et des variations de contraste beaucoup plus importantes que les objets naturels. C'est pourquoi on va utiliser un prétraitement qui permet de mettre en évidence les zones de contraste élevé tels que les contours des objets. On va donc utiliser un filtre de détection de contours.

On a ici choisi d'utiliser un filtrage par différence avec la moyenne. Par rapport à d'autres filtres de détection de contours, ce type de filtre a l'avantage d'être omnidirectionnel, c'est à dire qu'il détecte les contours dans toutes les directions, il n'a pas de direction privilégiée.

L'application d'un tel filtre consiste à remplacer chaque pixel $f(x,y)$ de l'image par la différence entre le niveau de gris du pixel considéré et la moyenne des niveaux de gris des pixels voisins. La fonction de transfert d'un tel filtre est donné par la formule :

$$g(x, y) = f(x, y) - \sum_{i=1}^n \sum_{j=1}^n f(i, j) h(x-i, y-j) \quad (2-5-1)$$

Le choix de la taille du voisinage se fait en fonction du niveau de détails voulu et des dimensions de l'image. Un filtre de taille minimale (3x3) ne détectera que les contours les plus fins alors qu'un filtre de dimension plus importante pourra mettre en évidence des zones plus larges. Il faut aussi noter que le temps de calcul augmente avec la taille du filtre. Pour des

images de grande taille comme celles utilisées ici, le choix d'une taille de 15x15 est un bon compromis. L'application du filtre sur une image donne le résultat suivant :

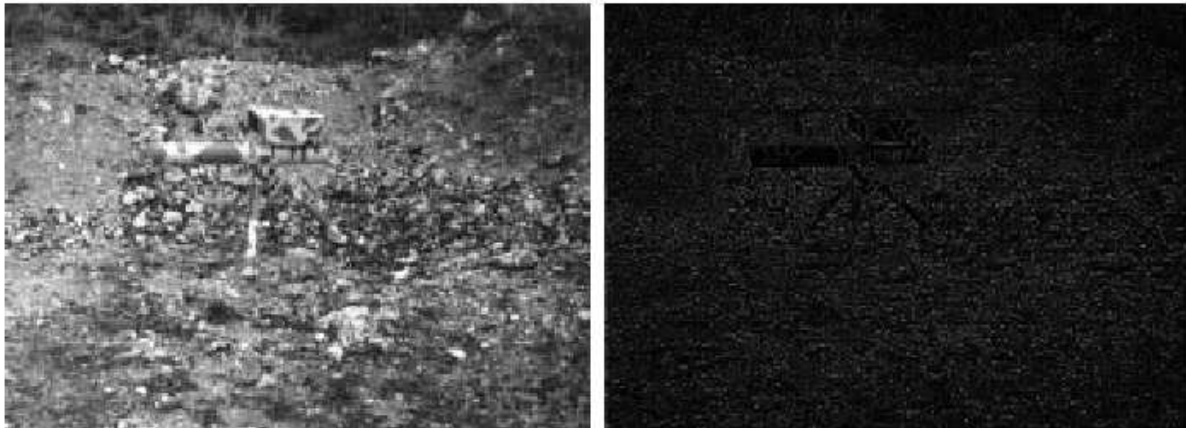


Fig .17 : Filtrage par différence avec la moyenne

L'utilisation du prétraitement pour la détection consiste à appliquer le filtrage sur l'image à tester, et ensuite à appliquer sur l'image filtrée la méthode de détection précédente, à savoir le découpage de l'image en imquettes, le calcul des courbes de Zipf et l'évaluation des critères de détection précédemment définis. La figure ci-dessous présente le résultat de la détection sur l'image précédente.

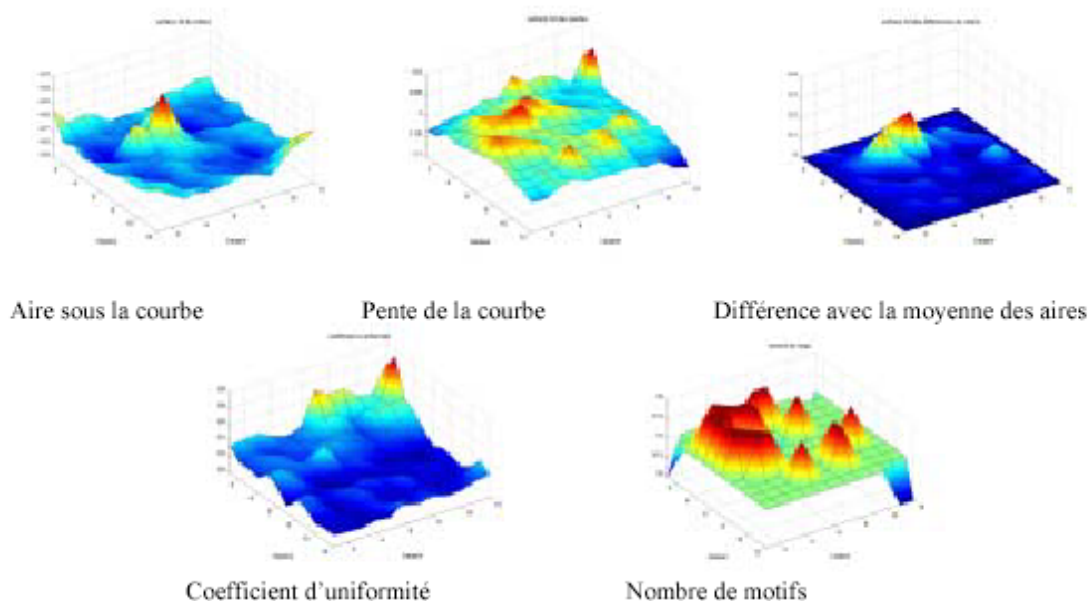


Fig. 18 : Résultat de la détection sur l'image filtrée de la Fig. 17

Sur l'image originale, sans prétraitement, on n'arrivait pas à distinguer l'objet du fond. L'application du prétraitement permet cette fois-ci de détecter l'objet avec les critères de l'aire sous la courbe et de la différence avec la moyenne. Le filtrage permet de mettre en évidence une distribution différente des contours sur les objets artificiels et sur les objets naturels. Sur l'image filtrée, l'intérieur de l'objet apparaît comme une surface parfaitement

uniforme, contrairement au fond sur lequel on distingue des contours fins dispersés à travers les textures naturelles de l'image. L'utilisation du filtrage a permis de mettre en évidence l'objet par rapport au reste de l'image, en accentuant les différences entre les objets naturels et artificiels. La figure 17 montre un autre exemple de l'efficacité de ce prétraitement. Sur cette image, l'objet est placé près d'un rocher et c'est ce dernier qui est détecté sur l'image non filtrée. Par contre, sur l'image filtrée, le rocher disparaît et l'objet est correctement détecté. L'utilisation du filtrage permet une détection de l'objet sur 68% des images testées.

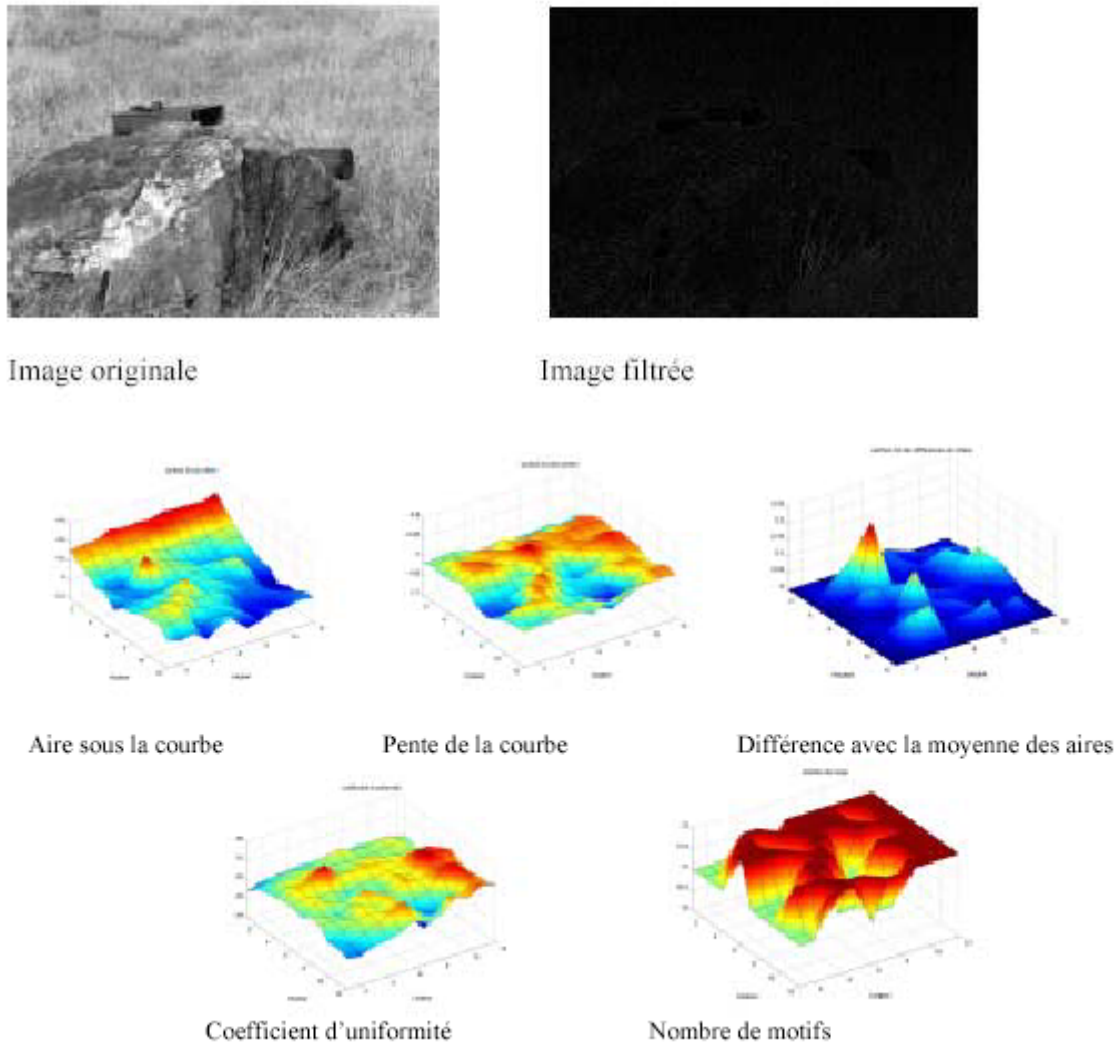


Fig. 19 : Autre exemple de détection sur une image filtrée.

Un autre type de prétraitement a également été testé, il s'agit de la méthode des orientations locales dominantes. Cette méthode permet de déterminer le degré d'artificialité de l'image en se basant sur l'orientation des contours. Les images qui représentent des objets artificiels se caractérisent par la présence de contours horizontaux et verticaux, alors que les images de milieux naturels ne présentent pas en général de direction privilégiée des contours. La méthode des orientations locales dominantes consiste à effectuer une détection de contours, puis à classifier les contours en fonction de leur orientation, et enfin à interpréter les résultats. La détection de contours consiste à calculer le gradient en chaque point de l'image en utilisant des filtres passe-haut directionnels. On utilisera ici des filtres de Sobel pour calculer le gradient horizontal et le gradient vertical en chaque point de l'image. Pour chaque pixel $P(i,j)$,

le gradient horizontal $S_x(i,j)$ et le gradient vertical $S_y(i,j)$ sont données par les produits de convolution suivants :

$$S_x(i,j)=h_x * P(i,j) \text{ et } S_y(i,j)=h_y * P(i,j) \quad (2-5-2)$$

$$\begin{array}{cc} \begin{array}{ccc} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{array} & \begin{array}{ccc} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{array} \\ \text{Avec } h_x = & \text{et } h_y = \end{array} \quad (2-5-3)$$

Les valeurs de la composante horizontale S_x et de la composante verticale S_y du gradient de l'image seront représentées par un vecteur dans un plan avec S_x en abscisse et S_y en ordonnée. On va ensuite attribuer à chaque point de l'image une classe selon sa position dans le plan (S_x, S_y) . Pour cela on peut utiliser deux méthodes : la méthode des k-means et la méthode directe. La méthode des k-means consiste à attribuer à chaque point une classe en fonction de la distance au centre de chacune des classes. Dans notre cas, on a défini 9 classes numérotées de 1 à 9 et la position initiale du centre de chaque classe a été fixée comme il est indiqué sur la figure 20.

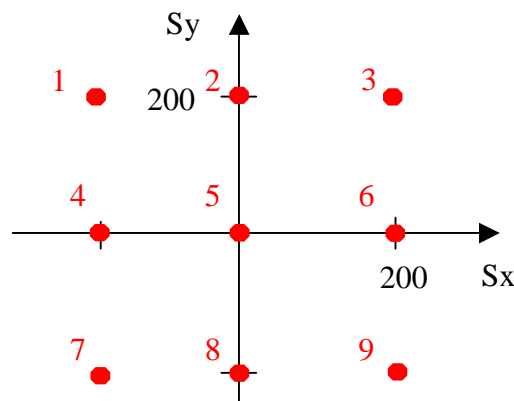


Fig. 20 : Disposition initiale du centre des classes

Chacune des classes représente une direction de contour. La classe centrale (classe 5) représente une absence de contour, c'est à dire une surface uniforme. A chaque pixel est associée une valeur de 1 à 9. Ces valeurs peuvent être représentées par une échelle de 9 niveaux de gris régulièrement espacés, dans laquelle la valeur 1 correspond au noir et la valeur 9 au blanc. On obtient ainsi une image pouvant être interprétée visuellement. L'utilisation de l'algorithme des k-means a cependant pour défaut d'avoir un temps de calcul assez long, c'est pourquoi on préfère utiliser la méthode directe de classification en classes fixes qui a l'avantage d'être plus rapide.

Cette méthode consiste à découper le plan (S_x, S_y) en secteurs angulaires correspondant chacun à une orientation de contour et à attribuer à chaque point un numéro de classe qui correspond au numéro du secteur dans lequel il se trouve. On remarque qu'ici les classes sont fixes, contrairement à la méthode précédente. L'opération de classification se fait en une seule itération, ce qui réduit considérablement le temps de calcul par rapport à la méthode précédente. On définira ici 9 classes, 8 qui correspondent aux 8 orientations possibles du

contour et une classe centrale qui contiendra les 20% de points de l'image les plus proches de l'origine dans le plan (S_x, S_y) . Ces points correspondent aux zones de l'image les plus uniformes. La figure suivante montre le découpage utilisé :

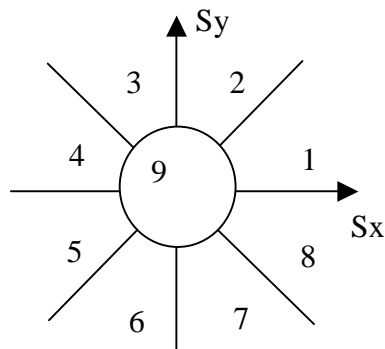


Fig. 21 : Partition en 9 classes du plan (S_x, S_y)

Comme précédemment, la classification sera représentée par une image à 9 niveaux de gris afin de pouvoir interpréter les résultats. La méthode de détection par courbe de Zipf est ensuite appliquée sur l'image codée par la méthode des orientations locales dominantes. La figure 22 montre un exemple de résultat obtenu avec cette méthode.

Les images codées par les orientations locales dominantes se caractérisent par une grande zone uniforme sur la surface de l'objet artificiel, qui est dépourvue de contours. On observe un minimum sur le critère de l'aire de la courbe qui se traduit par un pic sur la différence des aires. On observe également un minimum beaucoup plus important sur les critères de la pente de la courbe et du coefficient d'uniformité, qui est une manière différente de mesurer la pente de la courbe. D'une manière générale, sur les images traitées par la méthode des orientations locales dominantes, le critère le plus efficace pour la détection d'objets artificiels est la pente moyenne de la courbe, qui est inférieure pour les images qui contiennent un objet artificiel, en effet ces images apparaissent nettement plus uniformes.. En revanche, les critères de l'aire sous la courbe et de la différence des aires donnent des résultats moins satisfaisants. Par contre, les critères basés sur une mesure de la pente de la courbe permettent de détecter l'objet, qui correspond à un minimum de la pente moyenne de la courbe. Un critère de la différence des pentes, calculé comme le critère de la différence des aires mais à partir de la pente de la courbe de chaque image, permet d'assurer une bonne détection. Les performances sont nettement meilleures en utilisant le codage des orientations locales dominantes, en effet l'objet a pu être détecté sur 80% des images testées.



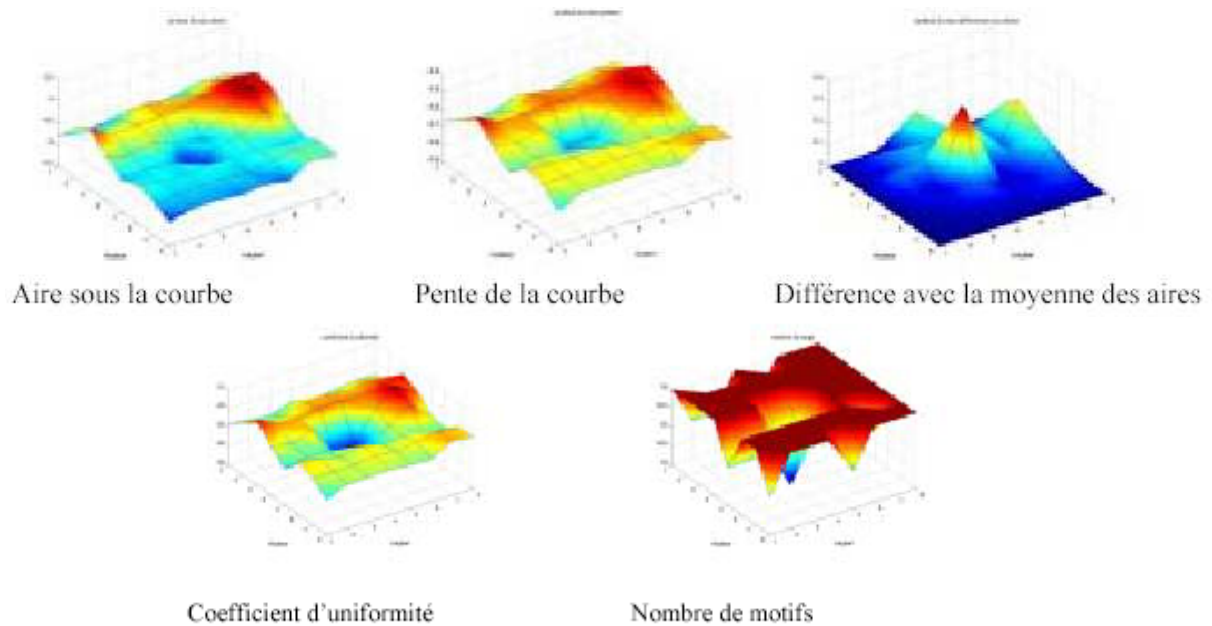


Fig. 22 : Résultat de la détection sur une image codée à l'aide des orientations locales dominantes.

2.6 Utilisation d'un codage à seuil

L'utilisation du codage des rangs généraux pour la détection d'objets ne donne pas, comme on l'a vu, des résultats totalement satisfaisants. Ce codage est en effet très sensible aux faibles variations de luminance à l'intérieur d'une surface, ce qui fait que les surfaces uniformes qui caractérisent les objets artificiels ne le sont pas en réalité. Cela rend alors plus difficile la différenciation entre les textures naturelles et artificielles et donc la détection d'un objet artificiel. C'est pourquoi on va utiliser un autre codage des motifs de l'image qui est plus robuste aux faibles variations des niveaux de gris, à savoir le codage des rangs généraux avec seuil. Comme on l'a vu précédemment, ce codage diffère du précédent par le fait que l'on applique la même valeur de rang à deux pixels si la différence entre les niveaux de gris est inférieure à un certain seuil. Dans notre cas, ce seuil a été fixé à 10, en effet les variations de niveaux de gris inférieures à 10 sont peu perceptibles par l'œil humain.

On applique la même méthode de détection que précédemment, à savoir le découpage de l'image en imagerie, le calcul des courbes de Zipf et l'évaluation des critères de détection, sur l'image codée à l'aide du codage des rangs généraux avec seuil. La figure ci-dessous montre un exemple d'application de cette méthode pour la détection d'un objet artificiel.

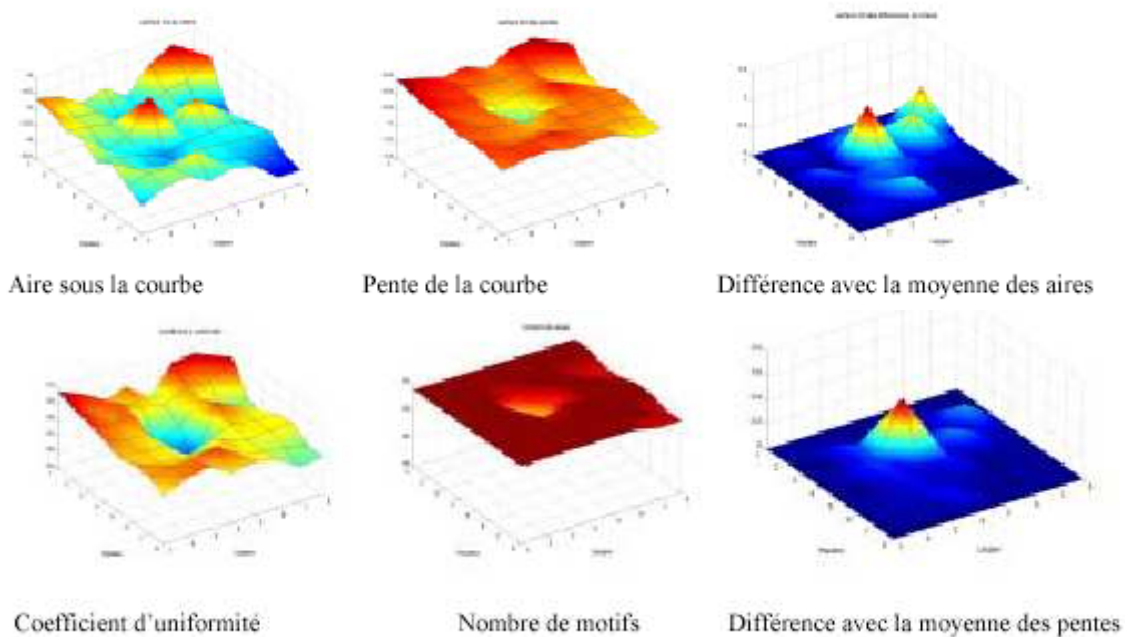
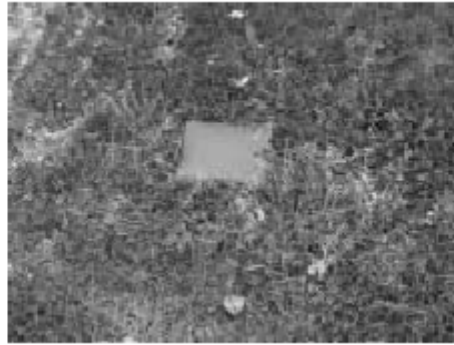


Fig. 23 : Détection d'un objet artificiel à l'aide d'un codage à seuil

Contrairement à ce qui se passe avec le codage précédent, c'est ici la pente de la courbe de Zipf qui est le critère le plus efficace pour la détection d'un objet artificiel. Cela s'explique par le fait que lorsqu'on utilise un codage avec seuil, le nombre total de motifs différents est moins grand, ce qui fait qu'on a davantage de motifs qui apparaissent plusieurs fois dans l'image. De ce fait, la distribution des fréquences des motifs est plus régulière, et les courbes de Zipf ont une allure plus linéaire, ce qui fait que la pente moyenne de la courbe caractérise mieux la courbe de Zipf. Cette pente moyenne est plus faible quand l'imagerie contient un objet artificiel, car le nombre de motifs différents apparaissant dans l'image est plus petit. Pour détecter la présence d'un objet artificiel dans l'image, un nouveau critère de détection a été défini, la différence avec la moyenne des pentes notée D_p . Ce critère se calcule suivant le même principe que la différence avec la moyenne des aires, suivant la formule :

$$D = \left| A_{ij} - \left(\frac{1}{9} \sum_{k=i-1}^{i+1} \sum_{l=j-1}^{j+1} A_{kl} \right) \right| \quad (2-6-1)$$

où P_{ij} représente la pente de la courbe de Zipf de l'imagerie d'indice (i,j) . La présence d'un objet artificiel dans l'image est détectée si la valeur maximum de ce critère est supérieure à 5 fois sa valeur moyenne.

Les performances de la détection avec ce nouveau codage ont été évaluées sur les mêmes images de test que précédemment. Les performances sont nettement meilleures que précédemment, l'objet a pu être détecté dans 76% des images testées à l'aide du critère de la différence des pentes. De plus, le temps de calcul est plus réduit qu'avec le codage précédent, du fait d'un nombre plus réduit de motifs différents dans l'image. On peut essayer de réduire encore ce temps de calcul en utilisant des motifs non recouvrants, c'est à dire que l'on balaie l'image de manière à ce que chaque pixel de l'image ne soit compté que dans un seul motif. De ce fait, on ne compte plus qu'un motif sur 9. Les essais ont montré que l'utilisation de motifs non recouvrants n'a qu'une faible incidence sur les pentes des courbes de Zipf des imagettes, comme on peut le voir sur la figure suivante.

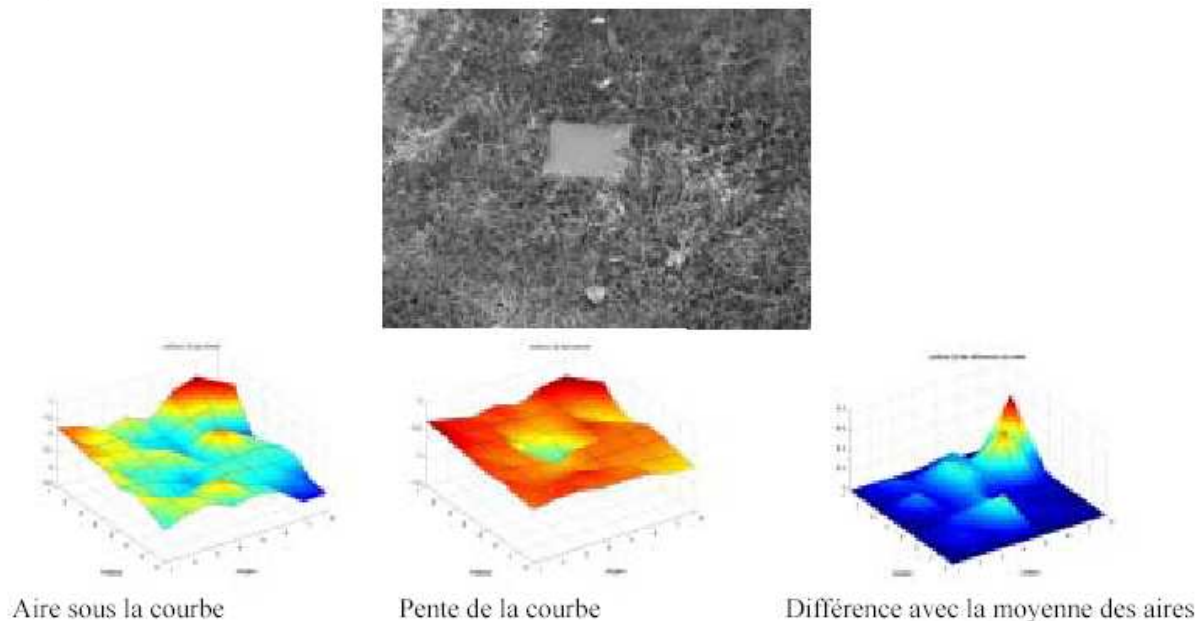


Fig. 24 : Détection d'un objet artificiel à l'aide de motifs non recouvrants

L'utilisation de motifs non recouvrants est particulièrement avantageuse car elle permet de diviser le temps de calcul par 8 pour une dégradation négligeable des performances : l'objet a pu être détecté dans 73,3 % des images testées.

L'utilisation du codage avec seuil permet donc de meilleures performances que le codage des rangs généraux, même sans utiliser de prétraitement. Il reste cependant des images pour lesquelles il n'est pas possible de détecter correctement l'objet quelque soit la méthode utilisée, notamment quand il n'est pas possible de séparer l'objet du fond, comme dans le cas d'un objet dans un arbre, ou quand le fond apparaît très uniforme à cause de la profondeur de champ ou de la saturation de l'image.

Le tableau ci-dessous présente un récapitulatif des résultats des différentes méthodes utilisées sur une même base de test de 40 images.

Méthode	Taux de réussite
Rangs généraux (motifs recouvrants)	56 %
Rangs généraux (motifs non recouvrants)	42,5 %
Filtrage par différence à la moyenne locale	68 %
Filtrage par les orientations locales dominantes	80 %
Rangs généraux avec seuil (motifs recouvrants)	76 %
Rangs généraux avec seuil (motifs non recouvrants)	73,3 %

2.7 Conclusion

Une méthode de détection d'objets basée sur la courbe de Zipf a été employée pour la détection d'objets artificiels dans une image d'un milieu naturel. Cette méthode présente l'avantage de ne pas nécessiter d'informations préalables sur la structure de la scène. Les objets naturels et artificiels se différencient entre eux par la répartition statistique des motifs de leur surface, ce qui se traduit par des caractéristiques différentes des courbes de Zipf. Sur une image codée à l'aide des rangs généraux, la différence entre l'aire sous la courbe de Zipf d'une imagerie par rapport aux imageries voisines permet de détecter la présence dans l'image d'un objet artificiel. Cependant, avec le codage des rangs généraux, cette méthode de détection est assez peu fiable, l'objet artificiel pouvant être confondu avec des éléments naturels de la scène. L'application d'un prétraitement comme un filtre de détection de contours ou la méthode des orientations locales dominantes permet une détection nettement plus efficace, les objets artificiels se distinguant des objets naturels par une distribution très différente de leurs contours. L'utilisation d'un codage à seuil permet également une meilleure détection, sans qu'il soit besoin d'utiliser un prétraitement. De plus, l'utilisation de ce codage permet également de raccourcir le temps de calcul qui reste l'un des défauts majeurs de notre méthode. Pour toutes ces raisons, c'est donc le codage des rangs généraux avec seuil qui est le plus intéressant pour la détection d'objets artificiels.

3 *Détection de zones d'intérêt à l'aide des lois puissance*

3.1 Introduction

Jusqu'à présent, nous avons utilisé la loi de Zipf pour détecter des objets dans un cas particulier, celui d'un objet artificiel dans un milieu naturel. Nous allons maintenant étendre notre approche à un cas plus général, celui de la détection d'une *zone d'intérêt* dans une image. Une des applications possibles est l'utilisation de la compression JPEG 2000, qui permet de spécifier une zone d'intérêt qui soit moins compressée que le reste de l'image. Il serait alors intéressant de pouvoir détecter automatiquement la zone d'intérêt à restituer avec une plus grande qualité d'image.

La difficulté est ici que l'on n'a pas véritablement d'information à priori sur le contenu de l'image. La notion de zone d'intérêt est en effet définie avant tout de manière subjective, par rapport à ce qui intéresse l'observateur humain qui regarde l'image. La zone d'intérêt d'une image, dans le cas général, peut représenter à peu près n'importe quoi, selon l'image considérée. Contrairement au cas particulier vu précédemment où l'on cherchait à détecter des objets artificiels dans des milieux naturels, l'on ne dispose pas ici d'hypothèses aussi fortes sur la nature de l'objet à détecter, c'est pourquoi il nous faut une méthode qui soit suffisamment générale pour pouvoir détecter une zone d'intérêt dans n'importe quelle image, sans connaissance a priori sur le contenu de l'image et la nature de la zone d'intérêt. Une telle tâche est évidemment très difficile à réaliser. Cependant, une zone d'intérêt présente des caractéristiques visuelles particulières qui permettent de la distinguer du reste de l'image. En général, la zone d'intérêt a une structure plus complexe que le reste de l'image, elle contient davantage de détails, elle est moins homogène out apparaît plus nette du fait de sa situation au premier plan de l'image. Elle contient donc davantage de motifs différents que l'arrière-plan de l'image. La répartition des fréquences d'apparition des différents motifs de l'image sera donc différente suivant la complexité structurelle de l'image. De ce fait, on peut utiliser les modèles de lois puissance afin de détecter la zone d'intérêt de l'image. En effet, il est possible de détecter par ce moyen une zone particulière de l'image qui se distingue du reste de l'image par sa complexité, si cette zone de l'image est suffisamment importante, elle a de fortes chances de constituer une zone d'intérêt pour celui qui observe l'image. Nous allons donc détailler diverses méthodes utilisant les modèles de lois puissance afin de détecter la zone d'intérêt. Nous allons d'abord utiliser la loi de Zipf, puis la loi de Zipf inverse afin de détecter des zones d'intérêt. Ensuite, nous utiliserons une mesure de l'entropie des fréquences d'apparition des motifs et la fusion des caractéristiques des courbes obtenues avec plusieurs codages des motifs de l'image. Enfin, nous étudierons l'utilisation d'une méthode de classification des caractéristiques des courbes de Zipf à l'aide d'un réseau de neurones.

Nous nous limitons à des zones qui ne seraient pas définies par un expert du domaine illustré par l'image mais par un observateur naïf qui regarderait globalement l'image. Nous supposerons également que la zone d'intérêt de l'image est unique et qu'elle occupe une surface relativement importante de l'image.

3.2 Utilisation de la loi de Zipf

Nous allons maintenant essayer d'utiliser la loi de Zipf afin de détecter des zones d'intérêt dans l'image. L'objectif est de détecter une zone particulière de l'image qui se distingue du reste de l'image par sa complexité structurelle, qui est ici modélisée par la répartition statistique de la fréquence d'apparition des motifs de l'image. La loi de Zipf permet de représenter cette répartition, mais en revanche elle ne donne aucune information spatiale sur la position d'une zone d'intérêt de l'image. C'est pourquoi il sera nécessaire de segmenter l'image et de calculer les courbes de Zipf associées à différentes régions de l'image afin de déterminer lesquelles correspondent à la zone d'intérêt. Comme on recherche les grandes zones d'intérêt de l'image, la segmentation n'a pas besoin d'être très précise. De plus, les caractéristiques de la courbe de Zipf associée à une image dépendent de la taille de celle-ci, il est préférable que les régions de l'image soient de même taille afin que la comparaison soit pertinente. C'est pourquoi la segmentation se fera en découpant les images en imasettes de même taille. Ensuite, on calculera les courbes de Zipf associées à chacune des imasettes et on classifiera les caractéristiques de ces courbes afin de déterminer quelles sont les imasettes qui appartiennent à la zone d'intérêt. L'organigramme ci-dessous présente le principe général de la méthode de détection de zones d'intérêt.

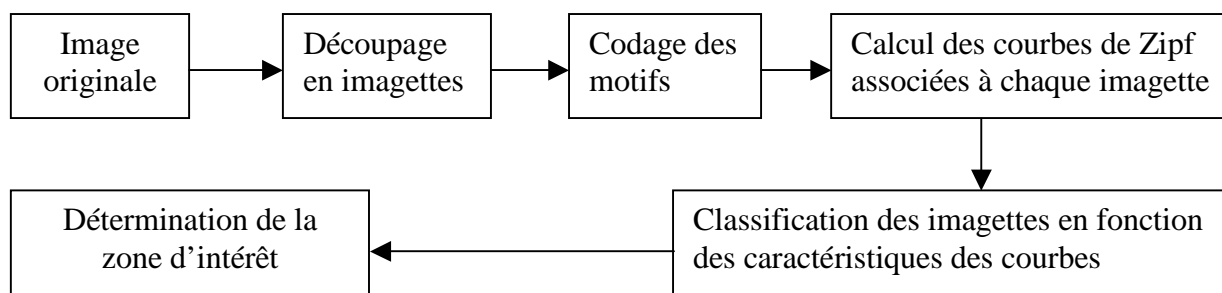


Fig. 25 : Principe général de la méthode de détection

Une première approche possible est d'utiliser la méthode déjà développée pour la détection d'objets artificiels dans des milieux naturels, qui utilise le codage des rangs généraux. La figure 26 montre des exemples de résultats obtenus avec cette méthode.

Cette méthode peut fonctionner sur certaines images, notamment celles qui se rapprochent d'un point de vue visuel des images d'objets artificiels, à savoir qui présentent un objet facilement séparable d'un fond relativement homogène sur l'ensemble de l'image. Cette méthode fonctionne relativement bien sur des scènes « fermées », qui présentent une faible profondeur de champ, comme c'est le cas pour l'image du haut, mais moins bien sur des scènes « ouvertes » comme des paysages ni les scènes d'intérieur, ou sur des fonds complexes comprenant plusieurs régions très différentes les unes des autres. Cette méthode a donc une efficacité limitée à certains types d'images, ce qui réduit son intérêt pratique. C'est pourquoi on va chercher autre chose. Jusqu'à présent, on avait surtout utilisé le codage des rangs généraux. Ce codage permet de caractériser la structuration fine de l'image, mais il ne donne pas de résultats conformes à la perception visuelle humaine car il ne donne aucune

information sur l'amplitude des variations des niveaux de gris à l'intérieur d'un motif. Or, l'objectif d'une détection de zones d'intérêt est en principe de détecter des zones d'intérêt qui soient conformes à la perception humaine. C'est pourquoi on va plutôt utiliser le codage des 9 classes. Ce codage ne prend pas en compte les faibles variations d'amplitude des niveaux de gris qui sont peu perceptibles par l'œil humain, en revanche il met en évidence les variations plus importantes. Il est donc davantage conforme à la perception visuelle humaine, et donc, en principe, mieux adapté à la détection d'une zone d'intérêt dans une image quelconque.

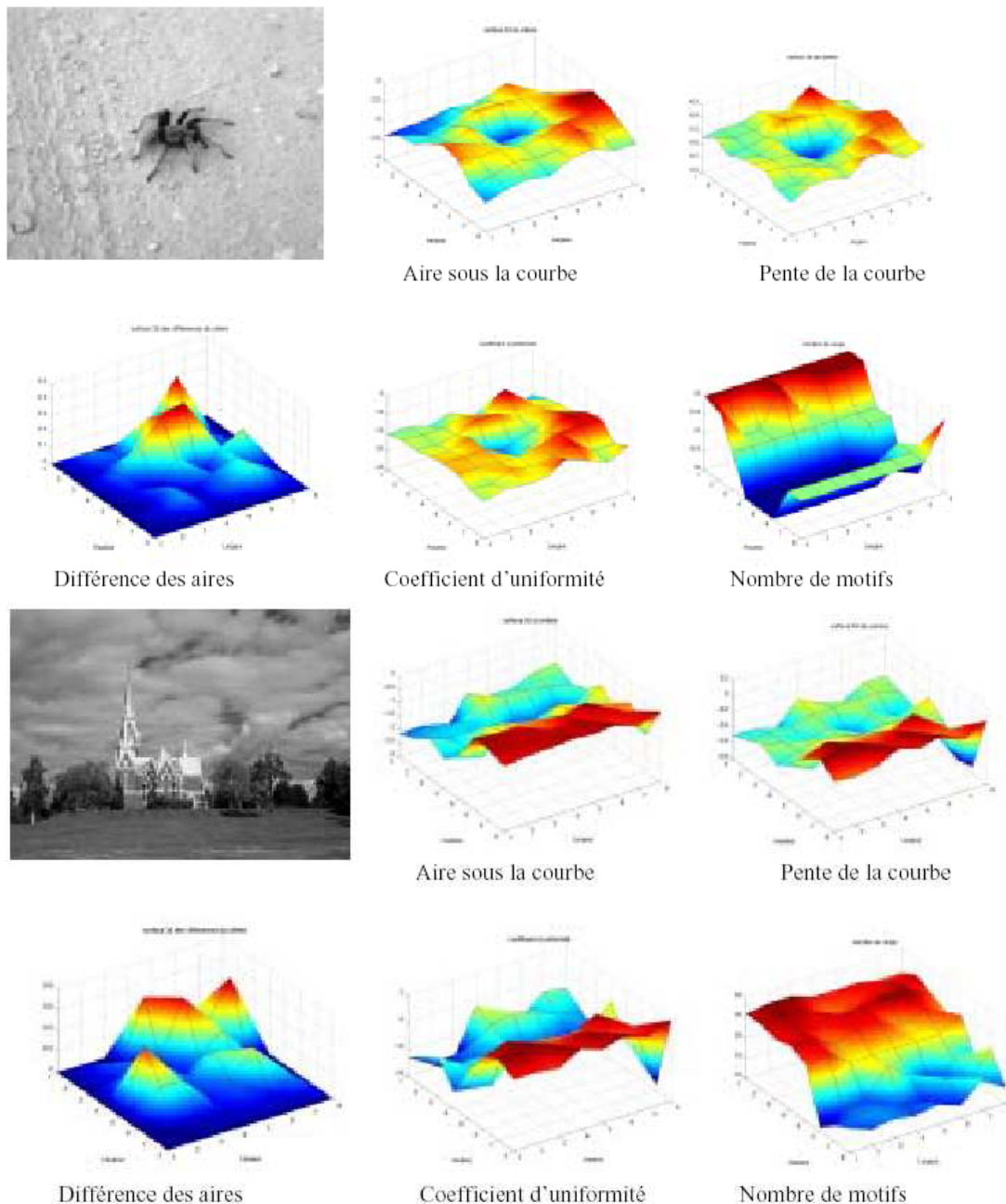


Fig.26 : Utilisation de la méthode de détection d'objets artificiels pour la détection de zones d'intérêt dans le cas général.

Avant d'élaborer une méthode de détection de zones d'intérêt basée sur la loi de Zipf et le codage des 9 classes, il convient d'abord de vérifier que la courbe de Zipf réalisée avec ce codage permet bien de caractériser la complexité du contenu structurel de celle-ci. Dans le cas des images entières, cela a déjà été vu au chapitre 2, nous allons maintenant voir ce qu'il en est de différentes parties d'une même image. La figure suivante montre les courbes de Zipf associées à deux zones d'une même image, une zone uniforme et une zone comportant de nombreux détails.

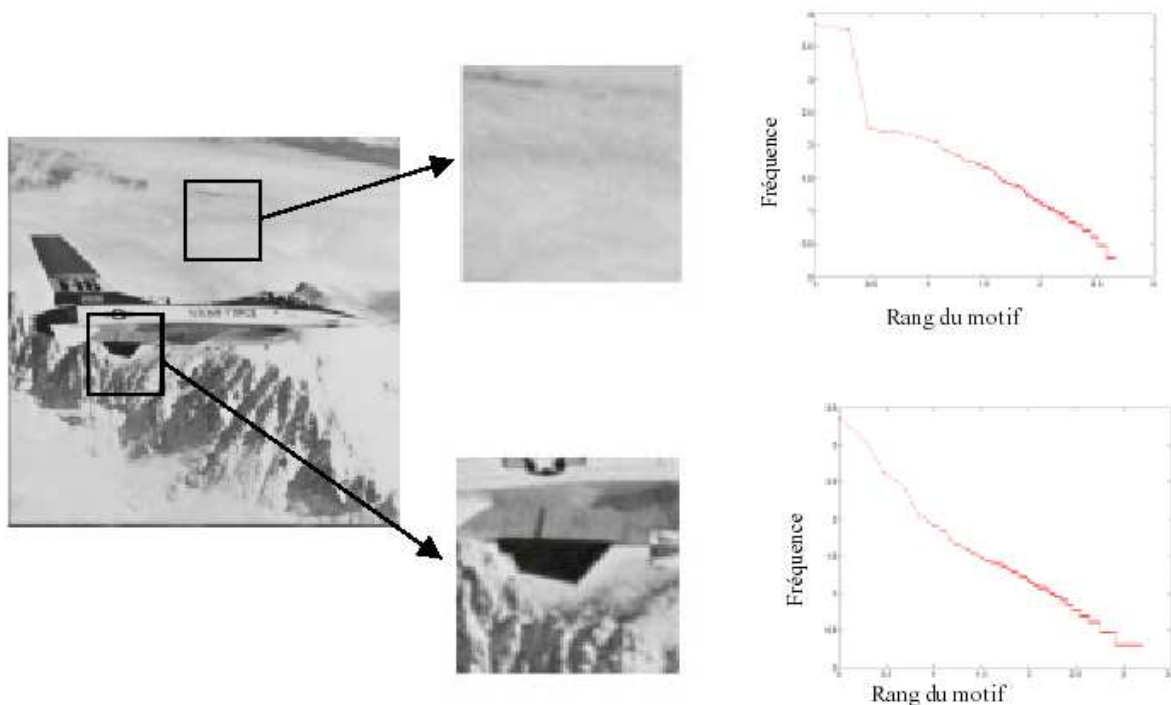


Fig. 27 : Courbes de Zipf d'une zone homogène (en haut) et d'une zone comportant de nombreux détails (en bas)

Sur la figure, on peut remarquer que les courbes de Zipf associées aux deux images sont différentes. La courbe de Zipf associée à l'image la plus uniforme comporte deux parties nettement distinctes, celle qui correspond aux motifs uniformes, qui sont ici particulièrement fréquents, et celle correspondant aux détails qui sont proportionnellement beaucoup moins nombreux. L'image contenant le plus de détails a par contre une répartition des fréquences des motifs plus régulière. La courbe de Zipf d'une zone de l'image ayant des caractéristiques différentes selon le contenu structurel de l'image, il est donc possible d'exploiter cette information afin de détecter des zones d'intérêt dans l'image qui se distinguent de leur environnement.

Nous avons vu que la courbe de Zipf d'une image codée avec la méthode des 9 classes peut être divisée en deux parties, celle correspondant aux zones uniformes et celle correspondant aux détails. On va donc pouvoir utiliser ces caractéristiques pour classifier les imagerie afin de détecter les zones d'intérêt. Les imagerie vont donc être classifiées selon les valeurs des pentes des deux parties de la courbe. La séparation entre les deux parties de la courbe peut être définie par le point le plus éloigné de la droite joignant les deux points extrêmes de la courbe. La courbe sera alors caractérisée par les deux pentes des droites de régression linéaire des deux parties de la courbe. Les caractéristiques des courbes de Zipf associées à chaque

imagerie peuvent être modélisées par un nuage de points, l'abscisse de chaque point représente la pente de la partie gauche de la courbe associée à chacune des imagerie et l'ordonnée correspond à la pente de la partie droite. La figure ci-dessous représente un exemple de nuage de points associé à une image découpée en 64 imagerie.

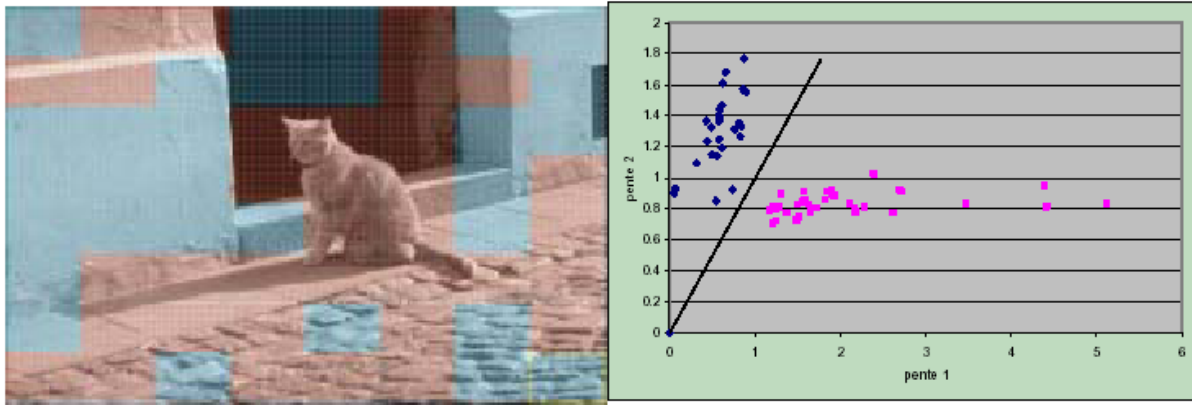


Fig. 28 : Classification des imagerie suivant les pentes des deux parties des courbes de Zipf

Sur cette image, le nuage de points est divisé en deux classes séparées par la droite d'équation $y=x$. Les imagerie correspondant aux points situés au-dessus de cette droite ont été représentées en bleu sur la figure et les imagerie correspondant aux points situés en-dessous ont été représentées en rouge. Nous voyons alors que le chat qui constitue dans cette image une région d'intérêt est entièrement contenu dans une unique classe : la classe rouge, c'est à dire celle qui est en dessous de la droite $y=x$. Si l'on étudie plus précisément les imagerie de cette classe, on peut également remarquer que les imagerie qui correspondent à la zone d'intérêt se situent dans la partie gauche de cette classe. Expérimentalement, la partie du nuage de points correspondant à la zone d'intérêt peut être délimitée par une droite verticale d'équation $x=1,2G_x$, où G_x représente l'abscisse du centre de gravité du nuage de points. La zone d'intérêt est alors constituée des imagerie qui correspondent aux points du nuage qui sont situés entre les droites d'équations $y=x$ et $x=1,2G_x$, comme le montre la figure 29.

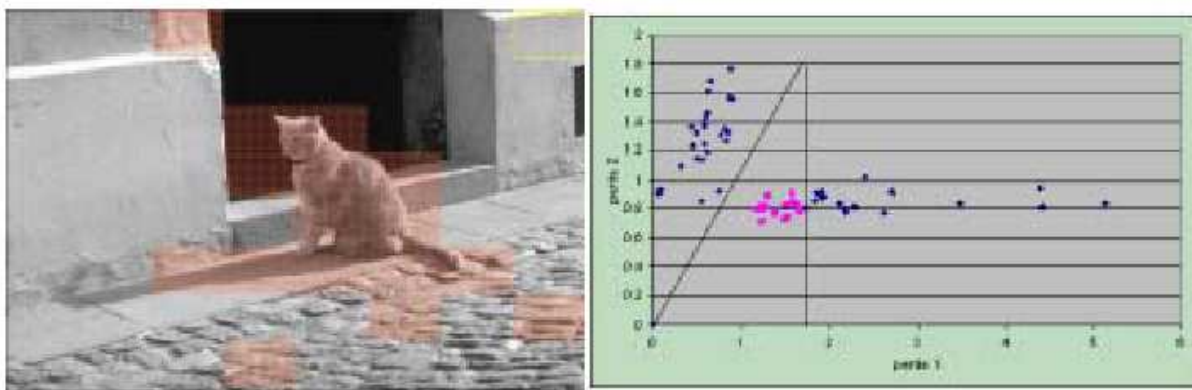


Fig. 29 : Détection d'une zone d'intérêt à l'aide de la loi de Zipf.

Les imagerie correspondant à la partie du nuage comprise entre les deux droites comprend d'une part la zone d'intérêt principale de l'image, c'est à dire le chat et d'autre part des imagerie qui correspondent à des détails isolés disséminés dans l'arrière-plan de l'image. Afin d'éliminer ces détails parasites, on définit comme zone d'intérêt la plus grande

composante connexe d'imagettes appartenant à la zone du nuage de points situé entre les deux droites. Sur certaines images, il se peut que la zone connexe comporte des trous. Dans ce cas, il est possible de incorporer dans la zone d'intérêt. Pour cela, on utilise un algorithme qui ajoute à la zone d'intérêt les imagettes dont les 8 imagettes voisines appartiennent déjà à la zone d'intérêt.

La composante connexe ayant une forme quelconque, si l'application envisagée nécessite une zone d'intérêt rectangulaire, on utilisera le rectangle englobant la zone d'intérêt détectée.

Le découpage de l'image en imagettes peut être fait de deux manières différentes. Il est possible d'utiliser un découpage en un nombre fixe d'imagettes indépendamment de la taille de l'image, mais on peut également utiliser un découpage adaptatif de l'image en fonction de la taille de celle-ci. En effet, l'efficacité de la méthode de détection dépend de la taille des imagettes, si les imagettes sont trop petites, les motifs ne seront pas assez nombreux pour que la répartition statistique des motifs de l'image suive une répartition en loi puissance, la courbe de Zipf obtenue ne sera donc pas significative. En revanche, si les imagettes sont trop grandes, la segmentation obtenue ne sera pas suffisamment précise pour détecter la zone d'intérêt effective de l'image. La taille optimale des imagettes a été déterminée expérimentalement, elle est de l'ordre de 5000 pixels. L'image sera donc partitionnée en imagettes rectangulaires de même dimension dont le nombre sera choisi automatiquement de manière à ce que la surface de l'imagette soit la plus proche possible de 5000 pixels.

La méthode a été évaluée sur une base de test de 100 images photographiques représentant des sujets divers. Ces images sont en niveaux de gris, leur taille varie de 256x256 à 2048x1536 pixels. Certaines de ces images ne sont pas compressées, d'autres ont subi une compression de type JPEG. Ces images ont été choisies de manière à représenter un éventail de cas typiques que l'on peut retrouver dans la pratique. Chacune de ces images comporte une zone d'intérêt nettement visible qui a été identifiée comme telle par plusieurs observateurs humains. Lors de l'évaluation de la méthode de détection, on considère que la zone d'intérêt est détectée si elle correspond à celle qui a été déterminée par l'observation humaine.

La méthode de détection utilisant les pentes des deux parties de la courbe de Zipf donne des résultats assez moyens. Ainsi, sur une base de test de 100 images, la zone d'intérêt a été détectée correctement sur seulement 56 % des images testées. C'est pourquoi on va chercher une autre méthode qui se base sur l'utilisation d'un autre modèle de loi puissance : la loi de Zipf inverse.

3.3 Utilisation de la loi de Zipf inverse

On va maintenant utiliser la loi de Zipf inverse afin de déterminer la zone d'intérêt de l'image. Comme celle qui a été utilisée avec la loi de Zipf, la méthode de détection sera basée sur une classification d'imagettes en fonction des caractéristiques des courbes de Zipf inverse. Le codage utilisé sera également celui des 9 classes.

Nous allons d'abord vérifier quelles sont les caractéristiques de la courbe de Zipf inverse qui peuvent permettre de détecter une zone d'intérêt. La figure ci-dessous montre les courbes de Zipf inverse associées à deux parties différentes de la même image.

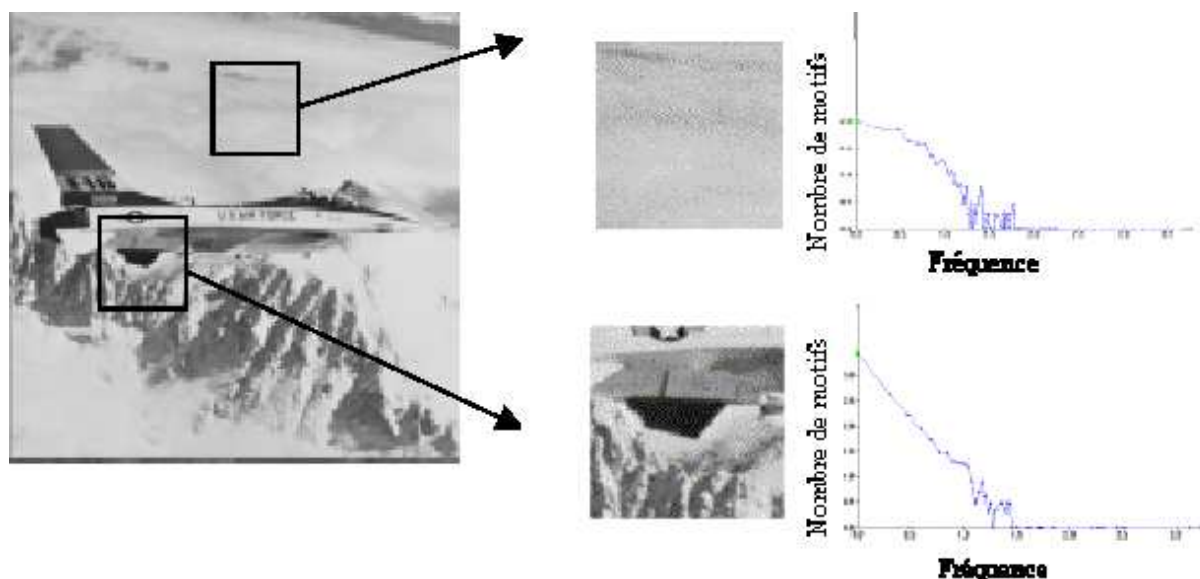


Fig.30 : Courbes de Zipf inverse associées à différentes parties d'une même image codées par la méthode des 9 classes.

On remarque ici que la pente de la courbe et l'ordonnée à l'origine sont plus élevées dans la courbe associée à la zone de l'image qui comporte le plus de détails. Les deux caractéristiques peuvent donc être utilisées pour détecter la zone d'intérêt. Comme avec la loi de Zipf, on calculera la loi de Zipf inverse associée à chaque imagerie et les caractéristiques des courbes seront représentées par un nuage de points. Les caractéristiques qui seront utilisées ici pour déterminer la zone d'intérêt seront la pente de la droite de régression des n premiers points de la courbe de Zipf inverse et l'ordonnée à l'origine. La pente de la droite de régression sera calculée à l'aide de la méthode des moindres carrés. Chaque point du nuage correspond à une imagerie, l'abscisse du point représentant la pente et l'ordonnée du point représentant l'ordonnée à l'origine de la courbe de Zipf inverse, c'est à dire le logarithme du nombre de motifs n'apparaissant qu'une seule fois dans l'image. Les imageries qui ont l'ordonnée à l'origine la plus élevée sont celles qui ont le plus de motifs n'apparaissant qu'une seule fois dans l'image, c'est à dire celles qui sont les moins uniformes et comportent donc le plus de détails. Ces imageries, qui se trouvent en haut du nuage de points, sont celles qui correspondent dans la plupart des cas à la zone d'intérêt. Etant donné que l'ordonnée à l'origine de la courbe de Zipf dépend de la taille de l'imagerie, les imageries devront toutes être de même taille afin que la comparaison soit possible. Comme le montre la figure ci-dessous, les points du nuage sont proches les uns des autres et forment un ensemble compact, c'est pourquoi il serait difficile de les séparer en utilisant une méthode de classification non supervisée telle que celle des k -means. C'est pourquoi on préfère utiliser une autre méthode qui consiste en un partitionnement dynamique du nuage de points en fonction de la taille souhaitée de la zone d'intérêt. Une zone d'intérêt initiale est constituée des imageries qui sont représentées par les points dont l'ordonnée à l'origine est supérieure à celle du centre de gravité du nuage de points. Afin d'éliminer les détails isolés, on ne conserve que la plus grande composante connexe de ces imageries. La zone d'intérêt déterminée avec cette méthode comportant éventuellement des trous, on bouche les trous en incluant dans la zone d'intérêt toutes les imageries qui se trouvent à l'intérieur d'une région fermée constituée d'imageries déjà étiquetées comme appartenant à la zone d'intérêt. La figure 32 montre une

zone d'intérêt avant et après le rebouchage des trous. Pour qu'une région de l'image soit considérée comme une zone d'intérêt par un observateur humain, elle doit être d'une taille suffisamment importante pour être visible, sans toutefois occuper une trop grande portion de l'image. C'est pourquoi on ajuste la taille de la zone d'intérêt afin que sa surface soit comprise entre deux valeurs limites que l'on a fixées à 20% et 50% de la surface totale de l'image. Pour cela, si la région d'intérêt initialement trouvée est inférieure à 20% de la surface de l'image, on abaisse l'ordonnée de la droite de séparation entre deux classes afin de prendre en compte davantage d'images, et on recalcule la zone d'intérêt jusqu'à ce que la surface totale de la zone d'intérêt, après extraction de la plus grande composante connexe et rebouchage éventuel des trous soit au moins égale à 20% de la surface de l'image. Inversement, si la zone d'intérêt trouvée a une surface supérieure à 50% de la surface totale, on augmente l'ordonnée de la droite de séparation entre les deux classes jusqu'à ce que la zone d'intérêt occupe moins de 50% de l'image. Selon l'application envisagée et le type d'images que l'on veut traiter, ces valeurs limites peuvent éventuellement être modifiées. La figure 31 montre quelques exemples de détection de zones d'intérêt à l'aide de cette méthode.

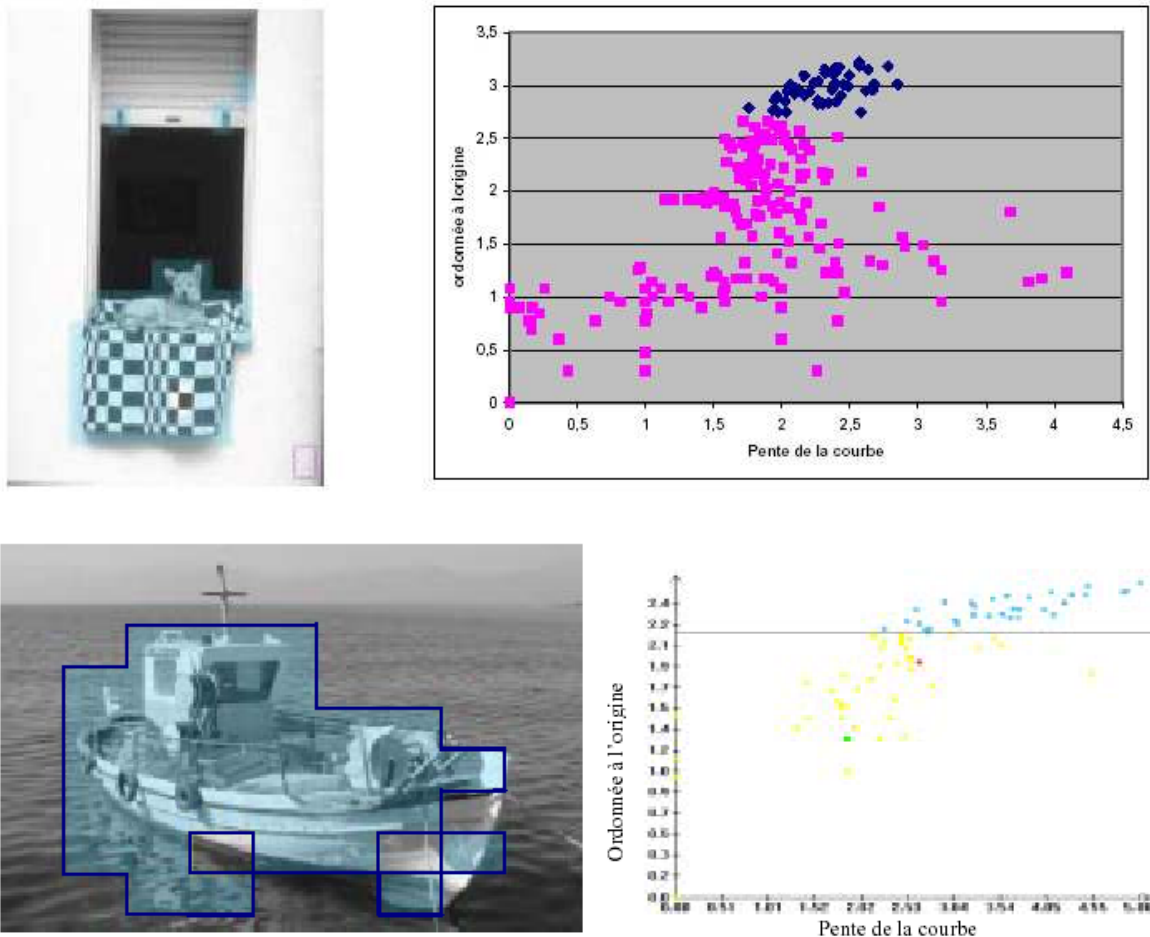


Fig. 31 : Détection de zones d'intérêt à l'aide de la loi de Zipf inverse.

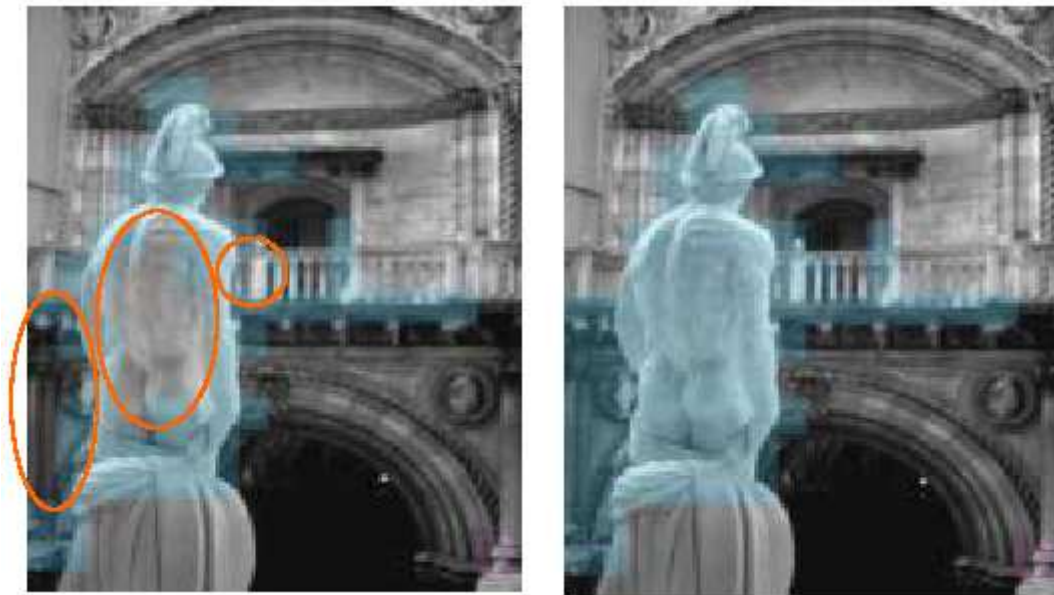


Fig. 32 : Bouchage des trous dans une zone d'intérêt

La méthode de détection basée sur la loi de Zipf inverse, telle qu'elle a été présentée, permet de détecter des zones d'intérêt qui contiennent des textures comportant plus de détails que le reste de l'image. Cependant, dans certains cas, la zone d'intérêt que l'on cherche à détecter est au contraire plus uniforme que le reste de l'image. C'est notamment le cas pour les images qui représentent des objets artificiels dans les milieux naturels, où les textures naturelles du fond comportent de nombreux détails fins. Dans ce cas, il est également possible d'utiliser la loi de Zipf inverse pour détecter la zone d'intérêt de l'image, mais à la différence du cas précédent, la région d'intérêt sera constituée par la partie inférieure du nuage de points, située en dessous du centre de gravité de celui-ci. Mis à part cette différence, on utilise la même méthode que dans le cas précédent pour l'extraction de la plus grande composante connexe, le rebouchage des trous et l'ajustement de la taille de la zone d'intérêt. La figure 33 montre le résultat de la détection dans le cas où l'objet à détecter est plus uniforme que le fond. La distinction entre les deux catégories d'images se fait en fonction automatiquement du nombre de motifs n'apparaissant qu'une seule fois dans l'image, si plus de 50% des motifs n'apparaissent qu'une seule fois sur l'ensemble de l'image, on recherchera une zone d'intérêt moins uniforme que l'arrière-plan de l'image.

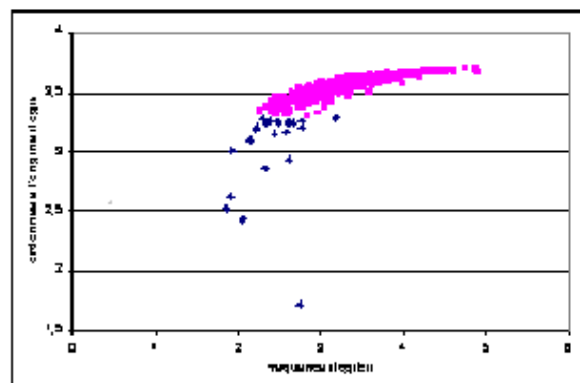
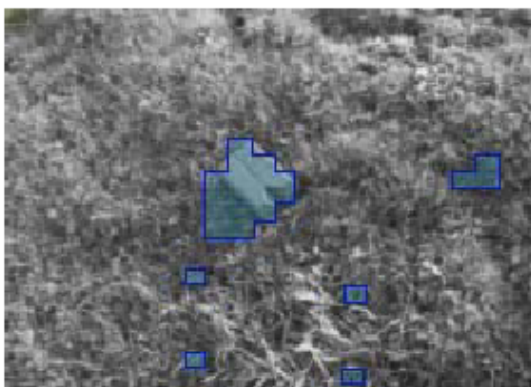


Fig. 33 : Détection d'une zone d'intérêt plus uniforme que l'arrière-plan

La méthode de détection utilisant la loi de Zipf inverse donne de bien meilleurs résultats que celle qui utilise la loi de Zipf. Sur la base de test de 100 images évoquée précédemment, la zone d'intérêt a été détectée sur 80% des images testées, contre 56% pour la loi de Zipf. De plus, la zone d'intérêt détectée avec la loi de Zipf inverse est en général plus précise que celle détectée avec la loi de Zipf, comme on peut le voir sur la figure 34. Sur cette image, la méthode de Zipf ne détecte pas le moteur de la moto, mais détecte une partie du mur qui ne fait pas véritablement partie de la zone d'intérêt.

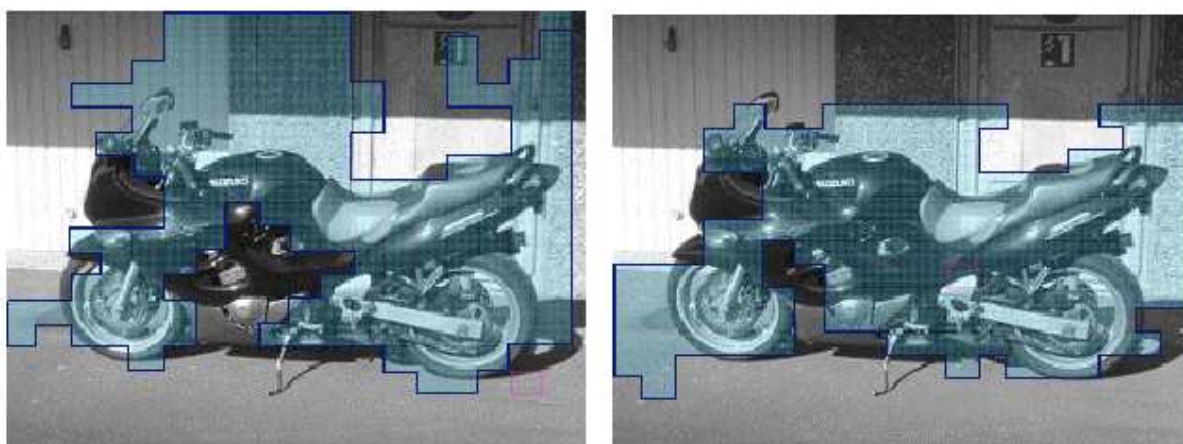


Fig. 34 : Comparaison des zones d'intérêt détectées avec la loi de Zipf (à gauche) et la loi de Zipf inverse (à droite)

Cependant, la zone d'intérêt trouvée avec cette méthode est souvent plus grande que la véritable région d'intérêt de l'image, en effet il est fréquent que des imagerie appartenant à l'arrière-plan soient incluses dans la zone d'intérêt, comme on peut le voir dans la figure ci-dessous.

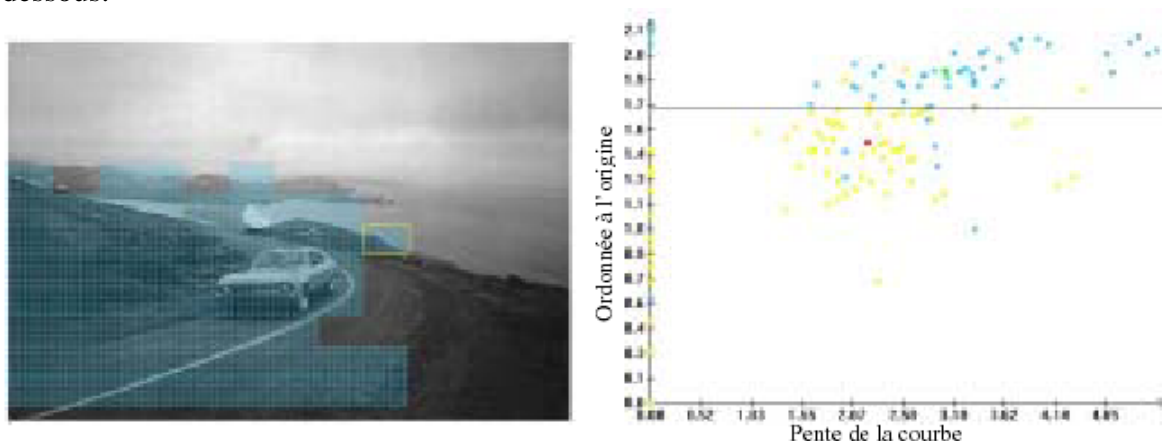


Fig. 35 : Détection d'une zone d'intérêt incluant une partie de l'arrière-plan

Jusqu'à présent, on a utilisé l'ordonnée à l'origine des courbes de Zipf inverse des imagerie pour séparer la zone d'intérêt du reste du nuage de points. Il est également possible d'utiliser la pente de la droite de régression des courbes de Zipf inverse. Dans ce cas, la zone d'intérêt est constituée de la plus grande composante connexe des imagerie dont la pente est supérieure à celle du centre de gravité du nuage de points. Sur la plupart des images, les

résultats obtenus avec la pente sont comparables à ceux que l'on obtient avec l'ordonnée à l'origine des courbes de Zipf inverse, comme le montre la figure 36. Il existe cependant des images sur lesquelles seule l'une des deux caractéristiques permet de détecter la zone d'intérêt.

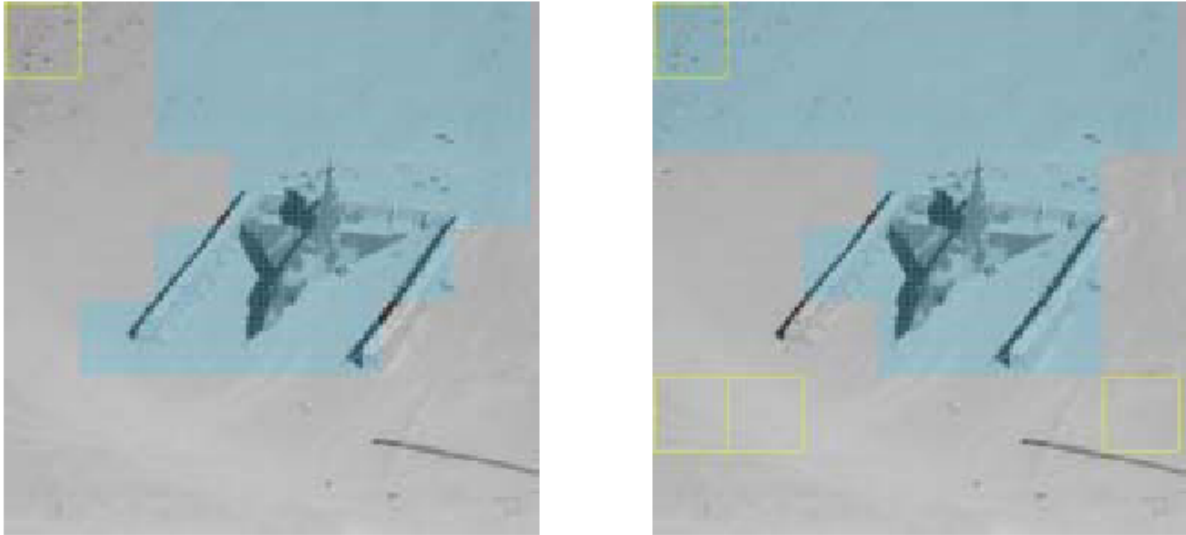


Fig. 36 : Zones d'intérêt obtenues en utilisant la pente (à gauche) et l'ordonnée à l'origine (à droite) des courbes de Zipf inverse.

Afin d'améliorer la détection, il est possible de combiner les deux caractéristiques et de ne prendre en compte que les imagerie qui sont représentées par les points qui correspondent aux courbes dont la pente et l'ordonnée à l'origine sont toutes deux supérieures à celles du centre de gravité du nuage de points. La zone d'intérêt est alors constituée de la plus grande composante connexe des imagerie représentées par les points situés en haut et à droite du nuage, comme le montre la figure ci-dessous. L'utilisation des deux caractéristiques donne de meilleurs résultats que lorsqu'on n'en utilise qu'une seule. La zone d'intérêt a été détectée dans 84% des images de la base de test. Les zones d'intérêt sont en général plus précises, la zone détectée correspond mieux à la zone d'intérêt réelle de l'image et elle comprend moins d'imagerie d'arrière-plan, comme on peut le voir sur l'image représentée sur la figure 37.

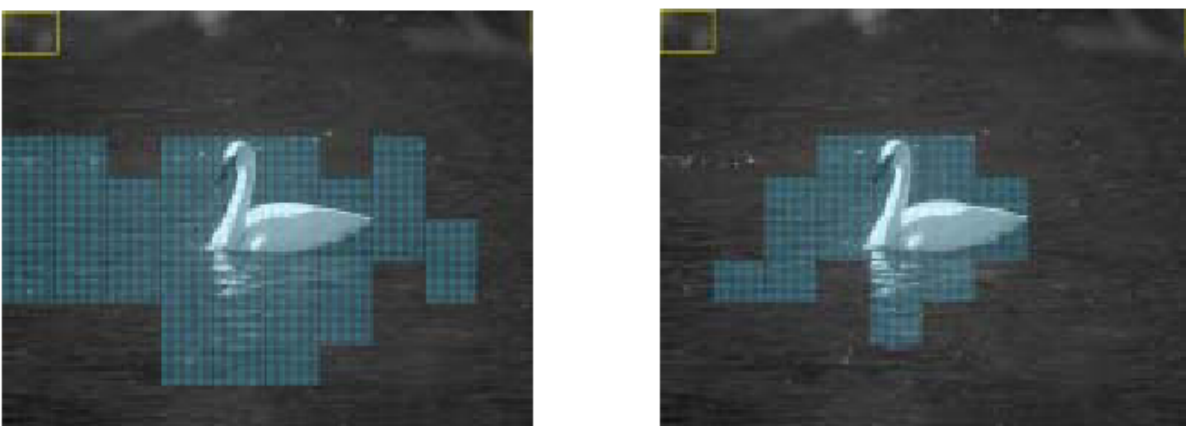


Fig. 37 : Zone d'intérêt détectée avec l'ordonnée à l'origine seule (à gauche) et avec les deux caractéristiques (à droite)

Comme pour la méthode utilisant la loi de Zipf, le résultat final dépend du découpage initial de l'image. Les imagerie ne doivent pas être trop petites pour que la répartition statistique de l'image ait une influence significative. A l'inverse, des imagerie trop grandes ne permettraient pas de définir des régions d'intérêt suffisamment précises. Il faut donc déterminer quelle est la taille optimale des imagerie, c'est à dire celle qui donne la zone d'intérêt la plus proche de celle que voit un observateur humain. La figure 38 montre le résultat de la détection de zones d'intérêt à l'aide de la loi de Zipf inverse, avec un découpage en 8x8 imagerie, l'objet n'est pas détecté, la zone d'intérêt détectée est ici le sol au premier plan de l'image. Avec un découpage en 19x19 imagerie, la zone d'intérêt, c'est à dire la moto, est correctement détectée. Par contre, si l'on utilise un découpage en 32x32 imagerie, la zone d'intérêt ne correspond plus que partiellement à l'objet à détecter, certaines parties uniformes de l'objet étant considérées comme faisant partie du fond et des zones texturées du fond comme appartenant à la zone d'intérêt. Dans le cas de la méthode de détection utilisant la loi de Zipf inverse et le codage des 9 classes, la taille optimale des imagerie est, dans le cas des photographies numériques haute résolution, de l'ordre de 5000 pixels. Cependant, cette valeur optimale peut varier suivant le type d'images que l'on veut traiter. Par exemple, dans le cas d'images de reproductions de tableaux impressionnistes, les meilleurs résultats ont été obtenus avec des imagerie d'une taille de 2000 pixels.

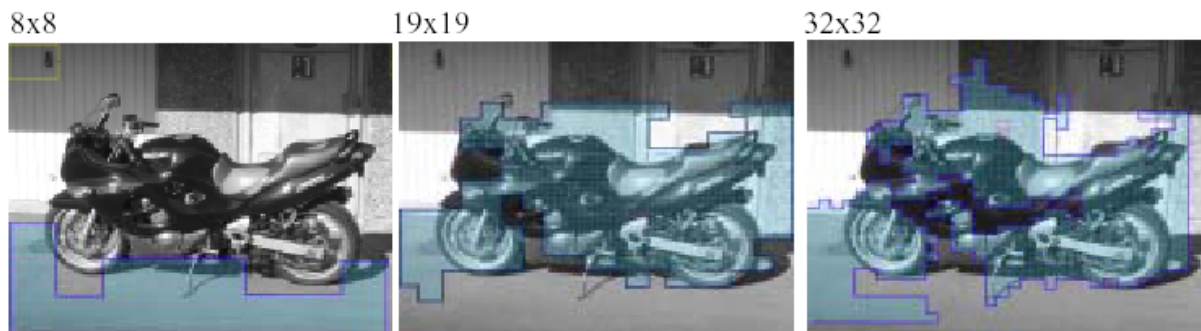


Fig. 38 : Zone d'intérêt détectée sur une même image avec un découpage en 8x8,19x19 et 32x32 imagerie

3.4 Entropie des fréquences d'apparition des motifs

Nous avons déjà vu qu'il est possible de définir une mesure d'entropie à partir des fréquences d'apparition des motifs de l'image. Deux formulations de l'entropie sont possibles, l'entropie relative aux motifs et l'entropie relative aux fréquences d'apparition. L'entropie relative aux motifs est définie par la formule suivante pour un ensemble de R motifs distincts :

$$\tilde{H}_w = - \sum_{r=1}^R \frac{f(r)}{T} \log_R \frac{f(r)}{T} \quad (3-4-1)$$

Dans cette formule, $f(r)$ représente la fréquence du motif de rang r , T représente le nombre total de motifs différents ou non, et on utilise un logarithme de base R. Cette entropie atteint son maximum quand tous les motifs de l'image ont la même fréquence d'apparition, et son minimum quand la fréquence relative d'un des différents motifs atteint 1. Ainsi, l'entropie donne une mesure de l'uniformité de la distribution des différents motifs de l'image. L'entropie relative à la fréquence est définie par la formule suivante :

$$\tilde{H}_f = - \sum_{r=1}^F \frac{I(f)}{R} \log_F \frac{I(f)}{R} \quad (3-4-2)$$

$I(f)$ représente le nombre de motifs distincts de fréquence f et F le nombre total d'occurrences des motifs dans l'image. On utilise un logarithme à base F afin que l'entropie soit comprise entre 0 et 1. L'avantage de cette formulation est qu'elle donne un poids plus important aux motifs dont la fréquence d'apparition est le plus souvent trouvée, ce sont les motifs les moins fréquents. Or, on a vu que les zones d'intérêt d'une image se caractérisent le plus souvent par de nombreux motifs n'apparaissant qu'une seule fois. Donc, l'entropie relative aux fréquences est a priori la mieux adaptée pour caractériser les zones d'intérêt de l'image. L'utilisation de l'entropie pour la détection des zones d'intérêt d'une image consiste, après avoir découpé l'image en imagerie, à calculer les fréquences d'apparition des motifs et à en déduire la valeur de l'entropie \tilde{H}_f pour chacune des imagerie. En théorie, la zone d'intérêt, c'est à dire la moins homogène, doit correspondre à un maximum d'entropie. L'entropie \tilde{H}_f peut être visualisée sur l'image par une échelle de couleurs, la plus claire correspondant à une entropie maximale et la plus foncée à l'entropie minimale. La figure 39 montre un exemple de détection de zone d'intérêt à l'aide de l'entropie relative aux fréquences d'apparition des motifs.

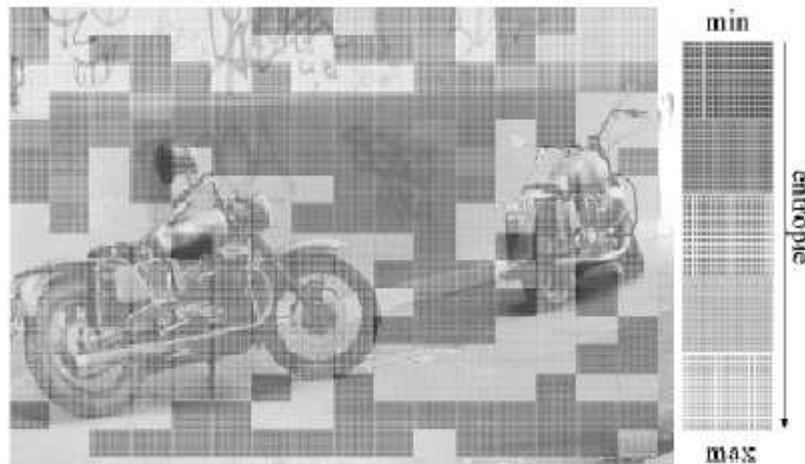


Fig. 39 : Détection de zones d'intérêt à l'aide de l'entropie

On remarque que la méthode ne donne pas des résultats aussi satisfaisants que ceux attendus. En effet, les zones de plus forte entropie correspondent en fait à des détails isolés de l'image plutôt qu'à la zone d'intérêt, qui se retrouve en fait mêlée à des imagerie d'arrière-plan dans une zone d'entropie moyenne. L'entropie n'est donc pas un critère très efficace pour la détection de zones d'intérêt, il vaut mieux utiliser les pentes et les ordonnées à l'origine des courbes.

3.5 Utilisation d'un codage basé sur l'écart type

La loi de Zipf peut également être utilisée avec un codage basé sur l'écart type afin de détecter la zone d'intérêt d'une image. Ce nouveau codage consiste à partitionner en 9 classes non plus les valeurs des niveaux de gris de l'image, mais les valeurs de l'écart type des niveaux de gris au voisinage de chaque pixel. Le codage de l'image à l'aide de cette méthode consiste à balayer l'image avec un masque de taille 3x3, à calculer la valeur de l'écart type

des niveaux de gris des pixels du masque et à associer cette valeur au pixel central du masque. Ensuite, on partitionne en 9 intervalles de largeur égale, l'échelle des valeurs de l'écart type obtenues sur l'image et on affecte à chaque pixel l'indice de la classe à laquelle il appartient. On obtient ainsi une image à 9 niveaux de gris qui représentent les valeurs de l'écart type des niveaux de gris de l'image originale. On découpe ensuite cette image en imagerie de taille 32x32 et l'on calcule la courbe de Zipf associée à chaque imagerie, de la même manière qu'avec le codage des 9 classes de niveaux de gris, en utilisant des motifs de taille 3x3. On calcule ensuite la pente de la droite de régression de chacune des courbes de Zipf et on classe les imageries suivant la valeur de cette pente. On considère que l'imagerie fait partie de la zone d'intérêt si la pente de sa courbe de Zipf est supérieure à 0,8 fois la valeur maximale de la pente sur l'ensemble des imageries. Cette méthode présente le défaut d'être très sensible à la profondeur de champ, c'est pourquoi elle est surtout efficace pour les cas où la profondeur de champ n'intervient pas, c'est à dire des prises de vues verticales, comme c'est le cas pour la détection d'objets artificiels au niveau du sol et pour la détection de structures urbaines dans les images aériennes et satellites, comme le montrent les images ci-dessous

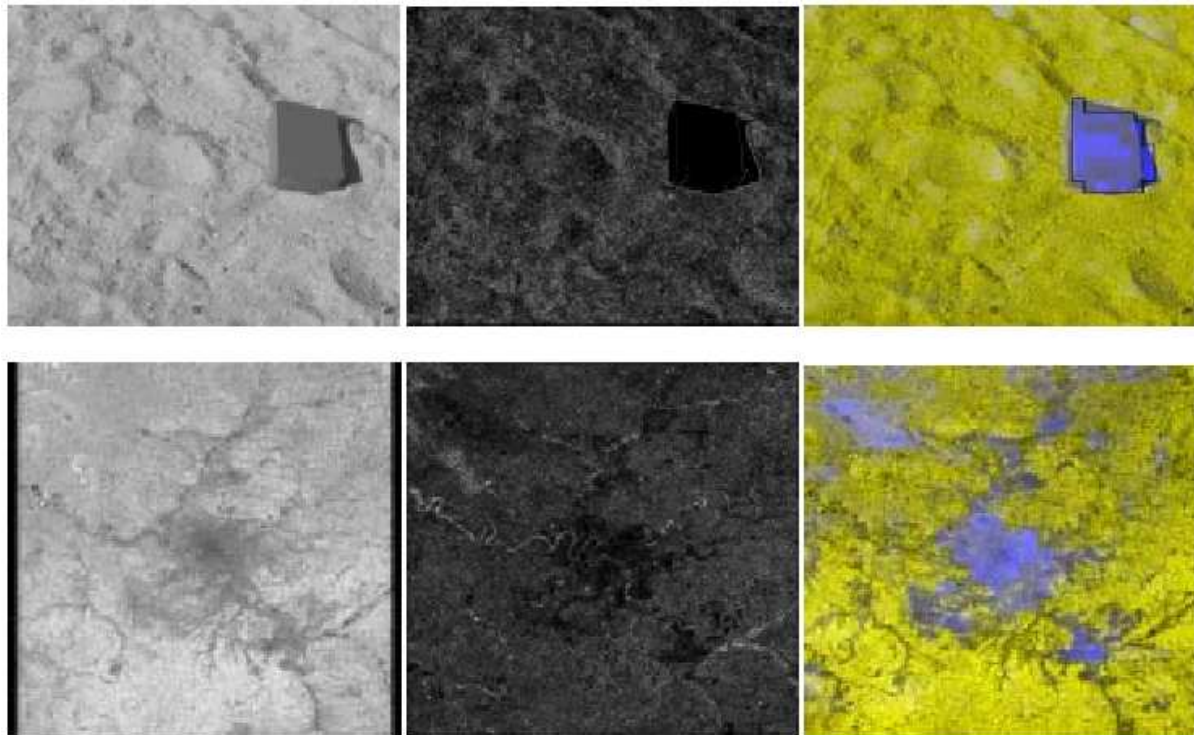


Fig. 40 : Détection de zones d'intérêt à l'aide de l'écart type. De gauche à droite : image originale, image de l'écart type des niveaux de gris et zone d'intérêt détectée.

3.6 Fusion de plusieurs codages

Jusqu'à présent, la loi de Zipf a été utilisée pour la détection de zones d'intérêt en utilisant les résultats obtenus à partir des caractéristiques des courbes de Zipf utilisant un seul codage des motifs de l'image. Il est également possible de fusionner des caractéristiques obtenues à partir de plusieurs courbes de Zipf réalisées avec des codages différents de l'image. Les valeurs différentes caractéristiques des courbes pour chaque imagerie sont alors représentées par un nuage de points dans un plan, si l'on utilise deux caractéristiques ou dans un hyperplan à n dimensions si l'on utilise n caractéristiques. Le problème est qu'il est difficile de déterminer a priori quelles sont les caractéristiques que l'on doit employer et surtout quels sont les critères à utiliser pour assurer une bonne détection de la zone d'intérêt. Pour cela, on n'a pas véritablement d'autre solution dans la pratique que de procéder empiriquement. La méthode consiste à essayer différentes combinaisons de caractéristiques sur différentes images et de repérer à quelles valeurs des caractéristiques correspond la zone d'intérêt.

a) Zipf et Zipf inverse

Il est possible d'utiliser simultanément les lois de Zipf et de Zipf inverse pour détecter une zone d'intérêt. Pour cela, on calcule les deux courbes pour chacune des images et on représente dans le nuage de points une caractéristique de chaque courbe. L'abscisse du point représente la caractéristique de la courbe de Zipf et l'ordonnée la caractéristique de la courbe de Zipf inverse. Si l'on utilise les pentes des courbes, la zone d'intérêt est constituée des images dont la pente de la courbe de Zipf et celle de la courbe de Zipf inverse sont toutes deux supérieures à celles du centre de gravité du nuage de points, c'est à dire qu'une image appartient à la zone d'intérêt si elle est représentée par un point (x,y) tel que $x > G_x$ et $y > G_y$. Cette combinaison de caractéristiques détecte des zones d'intérêt très précises sur certaines images mais la détection n'est réussie que sur 52% des images testées, des zones d'intérêt très apparentes n'étant pas détectées. On peut également utiliser les ordonnées à l'origine des deux courbes, le critère de détection restant le même, dans ce cas, le taux de réussite est de 68% mais les zones d'intérêt sont moins précises. L'utilisation de la pente de la courbe de Zipf et de l'ordonnée à l'origine de la courbe de Zipf inverse donne un taux de réussite de 67% et celle de la pente de la courbe de Zipf inverse et de l'ordonnée à l'origine de la courbe de Zipf donne un taux de réussite de 48%. Dans ces deux derniers cas, l'image est considérée comme appartenant à la zone d'intérêt si elle est représentée par un point tel que $x > 0,5G_x$ et $y > G_y$.

b) Entropie, Zipf et Zipf inverse

On peut également utiliser la mesure de l'entropie définie par la formule (3-4-2) en association avec les caractéristiques des courbes de Zipf et de Zipf inverse. On représente alors l'entropie en ordonnée et la pente ou l'ordonnée à l'origine de la courbe en abscisse. Dans ces conditions, quelle que soit la caractéristique utilisée, la zone d'intérêt est constituée par les images représentées par les points dont l'abscisse est supérieure à l'abscisse G_x du centre de gravité du nuage de points et dont l'ordonnée est inférieure à G_y . Dans ces conditions, le taux de réussite varie de 61% si l'on utilise la pente de la courbe de Zipf à 75% pour la pente de la courbe de Zipf inverse, en passant par 74% pour l'ordonnée à l'origine de la loi de Zipf et 69% pour l'ordonnée à l'origine de la loi de Zipf inverse. La figure ci-dessous

donne un exemple de zone d'intérêt détectée avec cette méthode, pour les différentes caractéristiques utilisées.

Pente de la courbe de Zipf



Ordonnée à l'origine de la courbe de Zipf



Pente de la courbe de Zipf inverse



Ordonnée à l'origine de la courbe de Zipf inverse



Fig. 41 : Exemple de zones d'intérêt détectées avec différentes caractéristiques associées à l'entropie

c) Ecart type et codage des 9 classes

Une de ces combinaisons est celle qui utilise le codage des 9 classes et une mesure locale de l'écart type. On calcule pour chaque imagerie deux courbes de Zipf. La première courbe est calculée à l'aide de la méthode des 9 classes déjà utilisée. Pour la seconde, le codage utilisé est une mesure de l'écart type des valeurs des pixels codés par la méthode des 9 classes sur un motif de taille 3x3. On calcule ensuite la pente de chaque courbe de Zipf et pour chacune des imageries, on représente en abscisse la pente de la courbe obtenue à l'aide de la méthode

des 9 classes et en ordonnée la pente de la courbe obtenue avec l'écart type. On peut remarquer que les imagerie contenant des zones totalement uniformes se caractérisent par une pente nulle pour la courbe obtenue avec la méthode des 9 classes. Les autres imagerie contenant davantage de détails se concentrent autour du centre de gravité du nuage de points. Par conséquent, les points représentant les imagerie de la zone d'intérêt seront celles qui seront les plus proches du centre de gravité du nuage. Sur la plupart des images, la zone d'intérêt est constituée par la plus grande composante connexe des imagerie représentées par des points dont la distance au centre de gravité est inférieure à 0,8 fois l'abscisse de celui-ci. La figure 42 montre un exemple de résultat obtenu avec cette méthode. Cette méthode donne de meilleurs résultats avec des imagerie dont la taille est de l'ordre de 100 pixels, elle permet donc de travailler avec des imagerie plus petites, ce qui permet de détecter des zones d'intérêt plus précises. Cependant, si on la compare au Zipf inverse, elle ne parvient pas à détecter de zones d'intérêt dans autant d'images, avec ce codage la zone d'intérêt n'est détectée que sur 60% des 100 images de la base de test.

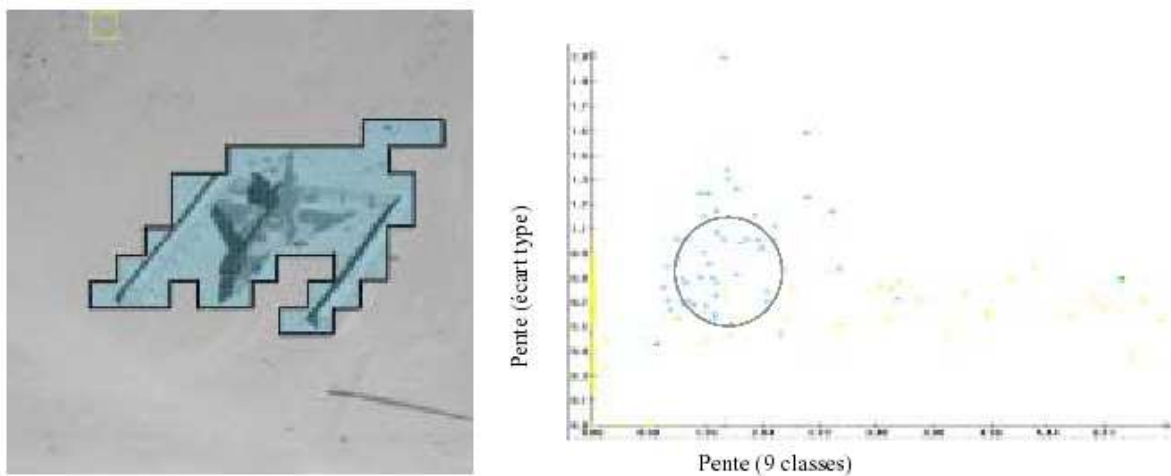


Fig. 42 : Détection de zones d'intérêt avec les 9 classes et l'écart type

d) Ecart type, rangs généraux et codage des 9 classes

Une autre méthode testée utilise un nouveau codage qui consiste à appliquer le codage des rangs généraux non pas sur les niveaux de gris de l'image d'origine, mais sur des motifs déjà préalablement codés sur la méthode des 9 classes. Par rapport à la méthode des rangs généraux classiques, ce codage permet d'éliminer l'influence des faibles variations de niveaux de gris pour ne conserver que les plus importantes et de mettre ainsi en valeur les zones les moins uniformes de l'image. La figure 43 montre un exemple de motif codé à l'aide de ce nouveau codage.

255	210	210
25	2	34
40	2	40

a

8	7	7
0	0	1
1	0	1

b

3	2	2
0	0	1
1	0	1

c

Fig. 43 : Motif original (a), codé par la méthode des 9 classes (b), par la méthode des rangs généraux appliquée sur le motif codé par les 9 classes (c)

Ce nouveau codage est ici associé au codage de l'écart des 9 classes défini au paragraphe précédent. On construit donc deux courbes de Zipf pour chacune des imasettes, et on représente chaque imasette dans le nuage de points par un point dont les coordonnées représentent les pentes des courbes obtenus avec les deux codages.

L'abscisse représente la pente de la courbe obtenue à l'aide du nouveau codage présenté à la figure 43 et l'ordonnée, la pente de la courbe de Zipf obtenue à partir des valeurs de l'écart type. On constate alors que les imasettes qui représentent l'arrière-plan de l'image tendent à se regrouper autour de la première bissectrice du graphe, tandis que les imasettes comportant des détails plus complexes, en particulier celles correspondant à la zone d'intérêt sont situées sous celle-ci. Avec cette méthode, la zone d'intérêt a été définie comme la plus grande composante connexe des imasettes situées sous la première bissectrice et dont l'abscisse est inférieure à 2,2 fois celle du centre de gravité du nuage de points. Cette méthode donne des résultats satisfaisants surtout pour des images dont la zone d'intérêt présente de nombreux détails fins. La figure 44 montre un exemple de zone d'intérêt obtenue avec cette méthode. La taille optimale des imasettes est de 1000 pixels. On peut remarquer que cette méthode détecte des zones d'intérêt plus petites et donc plus précises que la précédente. Cependant, sur certaines images, cette méthode peut ne pas détecter l'ensemble de la zone d'intérêt de l'image.

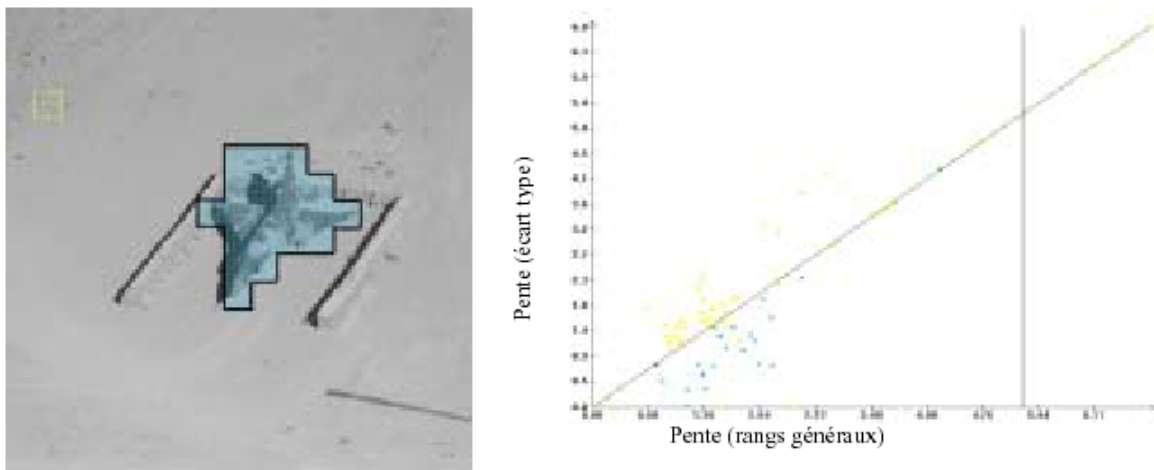


Fig. 44 : Détection de zones d'intérêt avec les rangs généraux, les 9 classes et l'écart type

3.7 Comparaisons des différents modèles de lois puissance utilisés

Nous allons maintenant comparer les résultats obtenus avec les différents modèles de lois puissance utilisés sur une même base de test de 100 images. Les taux de réussite des différentes méthodes utilisées sont présentés dans le tableau ci-dessous. La détection est

considérée comme réussie si la zone d'intérêt qui a été déterminée par l'observation humaine est incluse dans celle qui a été détectée. Dans le tableau, x et y représentent l'abscisse et l'ordonnée du point correspondant à l'imagette testée et Gx et Gy l'abscisse et l'ordonnée du centre de gravité du nuage de points. Pour chacune des différentes méthodes, la zone d'intérêt est constituée de la plus grande composante connexe des imagettes représentées par des points qui satisfont au critère de détection.

Caractéristique représentée en abscisse	Caractéristique représentée en ordonnée	Critère de détection	Taux de détection réussie
Pente de la 1 ^{ère} partie de la courbe de Zipf	Pente de la 2 ^{ème} partie de la courbe de Zipf	$y < x$ et $x < 1,2 G_x$	56 %
Pente de la courbe de Zipf inverse	Ordonnée à l'origine de la courbe de Zipf inverse	$x > G_x$ et $y > G_y$	84 %
-	Entropie	$y > 0,6 \cdot \max(y)$	38 %
Pente de la courbe de Zipf	Pente de la courbe de Zipf inverse	$x > G_x$ et $y > G_y$	52 %
Pente de la courbe de Zipf	Ordonnée à l'origine de la courbe de Zipf inverse	$x > 0,5 G_x$ et $y > G_y$	67 %
Ordonnée à l'origine de courbe de Zipf	Ordonnée à l'origine de la courbe de Zipf inverse	$x > G_x$ et $y > G_y$	68 %
Ordonnée à l'origine de courbe de Zipf	Pente de la courbe de Zipf inverse	$x > 0,5 G_x$ et $y > G_y$	48 %
Pente de la courbe de Zipf	Entropie	$x > G_x$ et $y < G_y$	61 %
Pente de la courbe de Zipf inverse	Entropie	$x > G_x$ et $y < G_y$	75 %
Ordonnée à l'origine de courbe de Zipf	Entropie	$x > G_x$ et $y < G_y$	74 %
Ordonnée à l'origine de courbe de Zipf inverse	Entropie	$x > G_x$ et $y < G_y$	69 %
Pente de la courbe avec le codage des 9 classes	Pente de la courbe avec l'écart type	$d[(x,y),(G_x,G_y)] < 0,8 G_x$	60 %
Pente de la courbe avec le codage des rangs généraux	Pente de la courbe avec l'écart type	$y < x$ et $x < 2,2 G_x$	69 %

D'après ce tableau, on remarque que la loi de Zipf inverse est celle qui a les meilleures performances en ce qui concerne la détection de zones d'intérêt. Les meilleurs résultats sont obtenus en utilisant les deux paramètres de la courbe, à savoir la pente et l'ordonnée à l'origine. L'utilisation d'un seul de ces paramètres donne des résultats à peine moins bons. Les autres méthodes utilisant la loi de Zipf sont comparativement beaucoup moins efficaces, et l'entropie encore moins. Par contre, les méthodes utilisant la loi Zipf inverse ont tendance à détecter des zones d'intérêt trop grandes, les méthodes utilisant une fusion de plusieurs codages permettant une meilleure précision. Cependant, il faut noter que toutes les méthodes ne réussissent pas avec les mêmes images, et qu'une méthode ayant un moins bon taux de réussite peut donner de meilleurs résultats sur une image en particulier, sans qu'il soit possible de déterminer a priori quelle méthode fonctionnera mieux avec quelle image.

3.8 Classification d'images par réseaux de neurones

Jusqu'à présent, la détection de zones d'intérêt à l'aide des lois puissances a été effectuée en utilisant une méthode de classification par partition d'un nuage de points représentant les caractéristiques des courbes de Zipf ou de Zipf inverse associées aux images. Si cette méthode de classification permet d'obtenir des résultats satisfaisants. Néanmoins, on réalise une simple partition linéaire. On va utiliser un outil de classification plus élaboré, à savoir un réseau de neurones, pour détecter les zones d'intérêt en utilisant les caractéristiques des courbes de Zipf.

Dans un premier temps, on utilise un unique réseau de neurones pour analyser l'ensemble de l'image. Le réseau de neurones utilisé est un réseau à couches composé d'une couche d'entrée, d'une couche cachée et d'une couche de sortie. L'image est partitionnée en un nombre fixe d'images, par exemple 8 x 8. Le réseau de neurones reçoit en entrée les caractéristiques des courbes de Zipf ou de Zipf inverse calculées pour chacune des images et il renvoie en sortie, pour chaque image, une valeur égale à 0 ou à 1 selon que l'image correspondante appartient ou non à la zone d'intérêt.

Comme la loi de Zipf inverse a une meilleure efficacité que la loi de Zipf pour la détection de zones d'intérêt, ce sont les caractéristiques des courbes de Zipf inverse qui vont être utilisées comme variables d'entrée du réseau de neurones. Le vecteur d'entrée du réseau de neurones sera constitué des valeurs de la pente et de l'ordonnée à l'origine de chaque image, on aura donc au total 128 entrées. La couche cachée est constituée de 128 neurones, et la couche de sortie comprend 64 neurones, chaque neurone correspondant à une image. Les neurones des couches cachées et de sortie sont réalisés avec des fonctions de transfert de type log-sigmoïde. La fonction log-sigmoïde est une fonction de la forme :

$$y(x) = \frac{1}{1+e^{-x}} \text{ La valeur de } y \text{ est comprise entre } 0 \text{ et } 1.$$

Les images d'apprentissage sont étiquetées manuellement : après découpage en images, on attribue une valeur 0 ou 1 à chaque image suivant que l'image appartient ou non à la zone d'intérêt déterminée par l'observateur. Pour chaque image on calcule la pente et l'ordonnée à l'origine de la courbe de Zipf inverse en utilisant le codage des 9 classes. Le vecteur d'entrée E est constitué des valeurs de ces deux caractéristiques pour chaque image, ces valeurs sont représentées sous la forme suivante :

$$E = \left\{ \begin{array}{l} \text{Pente(Imagette1)} \\ \text{Ordonnée(Imagette1)} \\ \text{Pente(Imagette2)} \\ \text{Ordonnée(Imagette2)} \\ \vdots \\ \text{Pente(Imagette64)} \\ \text{Ordonnée(Imagette64)} \end{array} \right\}$$

L'apprentissage se fait en utilisant un algorithme de rétropropagation du gradient. La fonction d'erreur que l'on cherche à minimiser étant l'erreur quadratique moyenne.

La phase de test consiste à calculer les caractéristiques des courbes de Zipf inverse associées à chacune des 64 imagerie, et à les présenter aux entrées du réseau de neurones. Le réseau de neurones fournit pour chaque sortie correspondant à une imagerie, un nombre réel compris entre 0 et 1, qui représente le degré d'appartenance à la zone d'intérêt. Ces valeurs sont ensuite binarisées, l'imagerie appartenant à la zone d'intérêt si la sortie correspondante est supérieure à un seuil de 0,6.

Le réseau de neurones a été développé en utilisant une base d'apprentissage constituée de 40 images extraites de la base de 100 images utilisée précédemment, et une base de test constitué des 60 autres images. La figure 45 montre quelques exemples de zones d'intérêt détectées avec le réseau de neurones sur des images de test.

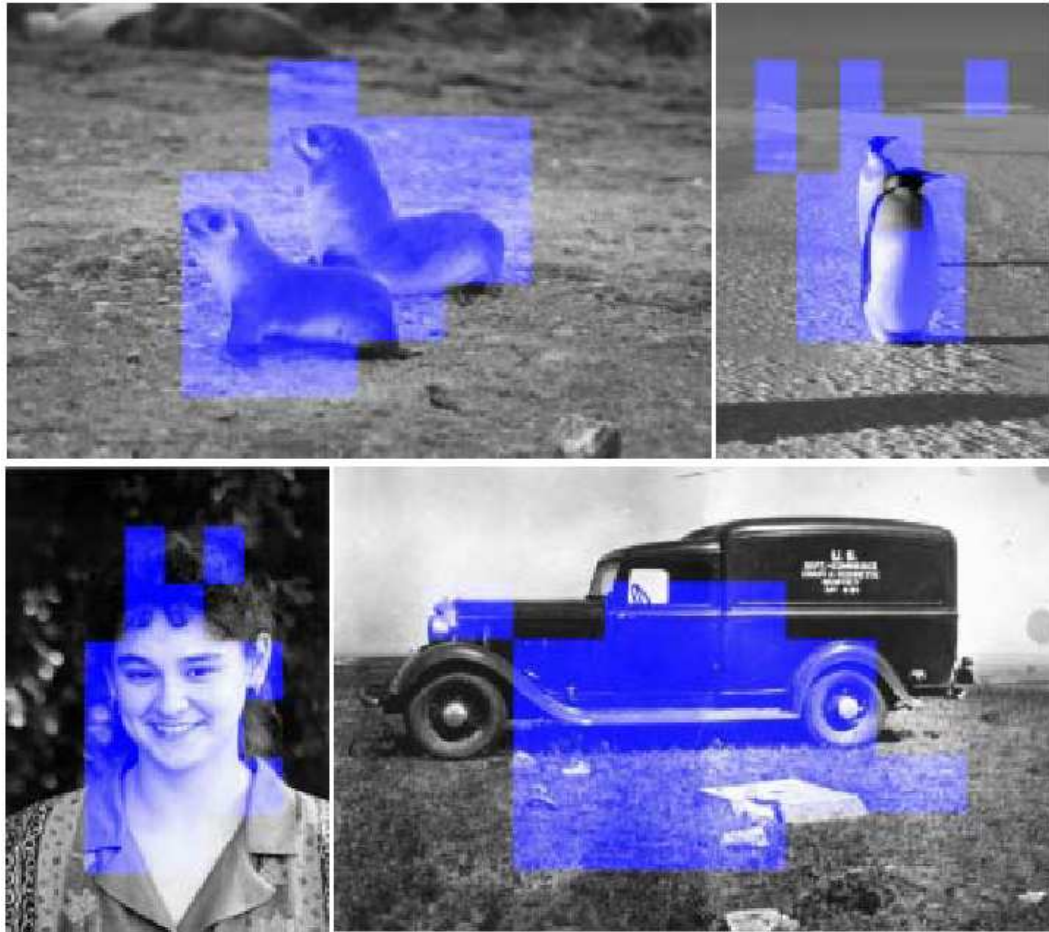


Fig .45 : Exemples de zones d'intérêt détectées avec un réseau de neurones

Les résultats de ces tests sont moyennement satisfaisants, la zone d'intérêt est détectée seulement sur 66% des images testées. Les performances obtenues avec le réseau de neurones sont donc inférieures à celles obtenues avec la classification par partition d'un nuage de points. Souvent, la région d'intérêt détectée par la méthode ne correspond que partiellement à l'objet recherché, la position de la région d'intérêt détectée se trouvant le plus souvent proche de la position moyenne des zones d'intérêt des images d'apprentissage, c'est à dire au voisinage du centre de l'image. En fait, le réseau de neurones a tendance à apprendre davantage la position des régions d'intérêt sur les images d'apprentissage que les caractéristiques des courbes de Zipf inverse. De ce fait, si les images d'apprentissage ont majoritairement une zone d'intérêt située au centre de l'image, la région détectée par le réseau

de neurones se trouvera au centre de l'image même si la véritable zone d'intérêt se trouve ailleurs.

Par contre, si l'on empêche le réseau de neurones d'apprendre la position des zones d'intérêt, en mélangeant aléatoirement les imagerie lors de la phase d'apprentissage, le réseau de neurones ne détecte plus rien : la zone étiquetée comme zone d'intérêt est une petite zone située n'importe où sur l'image et presque toujours au même endroit quelle que soit l'image testée. Le réseau de neurones a donc besoin d'une information sur la position de la zone d'intérêt pour pouvoir la détecter correctement, les caractéristiques des courbes de Zipf n'étant pas assez discriminantes à elles seules pour la détection de la zone d'intérêt.

Afin de réduire l'influence de la position des imagerie sans l'éliminer totalement, on a utilisé une autre architecture de réseaux de neurones qui découple les informations des différentes imagerie. On construit un réseau de neurones différent pour chaque imagerie. Chacun des réseaux de neurones comprend alors deux neurones d'entrée, qui reçoivent les valeurs de la pente et de l'ordonnée à l'origine de la courbe de Zipf inverse, une couche cachée à 3 neurones et une couche de sortie qui comprend un unique neurone qui indique si l'imagerie appartient ou non à la zone d'intérêt. Les résultats sont peu différents de ceux que l'on obtient avec un seul réseau de neurones, le système apprenant toujours la position des zones d'intérêt de préférence aux caractéristiques des courbes de Zipf inverse. Le taux de détection réussie est de 68%, soit une différence faible par rapport au cas précédent.

On peut également fusionner les caractéristiques des courbes de Zipf et de Zipf inverse à l'aide des réseaux de neurones. Le réseau de neurones associé à chaque imagerie comprend alors 4 entrées correspondant aux pentes et aux ordonnées à l'origine des deux courbes. Dans ce cas, l'erreur d'apprentissage du réseau augmente et le taux de réussite tombe à 57%. On n'améliore donc pas ici les performances en utilisant davantage de paramètres.

Les résultats obtenus s'expliquent par le fait que les images utilisées, ainsi que la nature des zones d'intérêt à détecter sont très différentes les unes des autres. Or, les méthodes d'apprentissage comme les réseaux de neurones sont davantage conçues pour détecter un type d'objet particulier sur des images semblables les unes aux autres. Dans notre cas, les images sont trop différentes entre elles pour que le réseau de neurones puisse effectuer une généralisation à partir des caractéristiques des courbes de Zipf inverse des images de l'ensemble d'apprentissage qu'on lui présente, la seule information qui est relativement commune à toutes les images étant la position de la zone d'intérêt dans l'image. La classification par réseau de neurones est donc mal adaptée à un problème aussi général que le nôtre, qui est la détection d'une zone d'intérêt perceptuellement saillante dans une image quelconque.

3.9 Conclusion

Les lois puissance peuvent être utilisées pour la détection de zones d'intérêt dans les images, car elles permettent de caractériser la complexité du contenu structurel d'une image. De ce fait, il est possible d'utiliser la loi de Zipf ou de Zipf inverse pour la détection d'une zone d'intérêt dans l'image, en découpant l'image en imagerie et en calculant la courbe de Zipf ou de Zipf inverse associée à chacune des imagerie. Le codage le mieux adapté à cette application est le codage des 9 classes, le codage des rangs généraux ne donnant pas une information sur l'image qui corresponde à la perception humaine. L'utilisation des pentes des courbes de Zipf comme critère de détection de la zone d'intérêt donne des résultats assez peu

convaincants. La loi de Zipf inverse permet d'obtenir des performances nettement meilleures que la loi de Zipf, même en n'utilisant qu'une seule des caractéristiques des courbes. La mesure de l'entropie des fréquences d'apparition des motifs, utilisée seule, n'est pas une méthode très efficace pour la détection des zones d'intérêt. L'utilisation d'une mesure de l'écart type n'est efficace que pour certains types d'images particuliers tels que les images aériennes. La fusion des informations issues de plusieurs codages, ou de la loi de Zipf et de la loi de Zipf inverse permet une légère amélioration par rapport aux résultats obtenus avec le codage des 9 classes, sans pour autant atteindre les performances obtenues par la loi de Zipf inverse avec un seul codage des motifs. La loi de Zipf inverse est donc le modèle de loi puissance de loin le plus efficace pour la détection des zones d'intérêt dans les images, même si la méthode utilisée a souvent tendance à détecter des zones d'intérêt trop grandes. Ce défaut peut cependant être réduit quand on utilise les deux caractéristiques de la courbe. La méthode de classification par partition linéaire d'un nuage de points donne de meilleurs résultats que des classifieurs plus élaborés comme les réseaux de neurones, en effet elle est mieux adaptée à la grande variété des images à analyser que les méthodes basées sur un apprentissage. De plus, son temps de calcul est relativement réduit.

4 Comparaison avec d'autres méthodes

Afin d'évaluer l'efficacité des modèles de loi puissance pour la détection des zones d'intérêt, nous allons maintenant appliquer d'autres méthodes de détection et en évaluer les performances. Parmi les méthodes existantes nous en avons choisies deux qui paraissent bien adaptées au problème posé, à savoir la détection de zones d'intérêt dans des images quelconques en niveaux de gris, sans faire d'hypothèses préalables quant au contenu de l'image. Il s'agit d'une méthode fractale par box-counting 3D et d'une méthode de construction de carte d'importance basée sur les caractéristiques perceptuelles de l'image.

4.1 Détection par calcul d'une dimension fractale

Cette méthode s'inspire de celle proposée par Carlotto et Stein [Car-] pour la détection d'objets sur des images de surfaces planétaire. Elle consiste à découper l'image en imasettes, et à calculer la dimension fractale de la surface de l'enveloppe des niveaux de gris de chacune des imasettes par la méthode du box-counting 3D.

On découpe tout d'abord l'image rectangulaire en imasettes carrées, dont la taille T est variable, mais la longueur de chaque côté de l'imasette est comprise dans l'intervalle $[2..64]$ dimension fractale par la méthode du box counting en 3D. Une imasette est vue comme une surface en 3 dimensions, dont les coordonnées horizontales représentent les positions en x et en y de chacun des pixels et l'altitude est le niveau de gris des pixels. On définit une séquence décroissante de tailles de boîte $\{\varepsilon_k\}$ comprise dans l'intervalle $[2 ; T_{imasette}]$ et de valeur 2^n avec $n \in \mathbb{N}^*$, et telle que ε_0 soit la plus grande valeur de taille de boîte.

Connaissant cette taille de boîte ε_k et la longueur d'un côté d'une imasette $L_{côté_imasette}$, on en déduit le nombre de boîtes $Nb_{boîte_côté_imasette}$ que peut contenir une imasette sur son arête.

La valeur de niveau de gris d'un pixel se trouve dans l'intervalle [0 ; 255] et on découpe la hauteur de la boîte en $Nb_{bo\grave{t}e_c\^o\grave{t}\acute{e}_im\grave{a}gette}$ portions. Il en résulte que les cubes qui composent chaque imagerie auront pour taille $\epsilon_k \times \epsilon_k \times \epsilon'_k$ avec

$$\epsilon'_k = \frac{256}{Nb_{bo\grave{t}e_c\^o\grave{t}\acute{e}_im\grave{a}gette}} \quad (4-1-1)$$

Pour chaque valeur de ϵ_k , on réalise plusieurs manipulations pour une imagerie l . On calcule d'abord, pour chaque colonne $c_{i,j}(\epsilon_k)$ de coordonnées (i ; j), le nombre de cubes $n_{i,j}(\epsilon_k)$ correspondant à la différence entre la valeur minimale et la valeur maximale de niveau de gris des pixels contenus dans le carré de l'imagerie :

$$n_{i,j}(\epsilon_k) = \text{int} \left(\frac{(\max_valeur_NDG - \min_valeur_NDG)}{\epsilon'_k} \right) + 1 \quad (4-1-2)$$

Ensuite, on fait la somme des $n_{i,j}(\epsilon_k)$, ce qui donne le nombre de cubes qu'il faut considérer pour couvrir toute la surface 3D contenant les niveaux de gris de l'imagerie :

$$N_l(\epsilon_k) = \sum_{i=1}^{Nb_{bo\grave{t}e_c\^o\grave{t}\acute{e}_im\grave{a}gette}} \sum_{j=1}^{Nb_{bo\grave{t}e_c\^o\grave{t}\acute{e}_im\grave{a}gette}} n_{i,j}(\epsilon_k) \quad (4-1-3)$$

On calcule alors, pour chaque imagerie l , le coefficient directeur de la droite de régression linéaire des points de coordonnées :

$$\left(\ln(N_l(\epsilon_k)), \ln\left(\frac{1}{\epsilon_k}\right) \right) \quad (4-1-4)$$

Ce coefficient directeur nous donne donc une approximation de la dimension fractale de l'imagerie l , qui sera comprise dans l'intervalle [2..3]. La dimension fractale d'une imagerie se trouve à partir du coefficient directeur de la droite de régression linéaire dont le tracé est lié à la séquence de tailles de cubes choisie. Pour une même imagerie, le coefficient de cette droite ne sera pas le même d'une séquence à une autre. Par conséquent, le choix d'une séquence affecte la qualité de l'estimation. Le choix de la séquence $\{\epsilon_k\}$ de taille des cubes affecte l'estimation de la dimension fractale D d'une imagerie, car ce choix est lié à la régression linéaire qui va permettre d'estimer D . Le problème est que l'on ne connaît pas de séquence « optimale ». Pour obtenir une estimation plus correcte, il suffit de calculer les dimensions fractales DF_k d'une imagerie pour différentes séquences k de tailles de cubes, et ensuite, on donnera pour dimension fractale finale de l'imagerie la moyenne des DF_k .

Pour sélectionner les images appartenant à la zone d'intérêt, on utilise une méthode basée sur le calcul de l'écart type. Après avoir calculé la dimension fractale de chaque imagerie, on calcule l'écart type σ_D par rapport à la moyenne \bar{D} des dimension fractales toutes les images. En fait, on ne conserve que les images dont la dimension fractale est inférieure à $\bar{D} - \sigma_D$ ou supérieure à $\bar{D} + \sigma_D$. On extrait ensuite la plus grande composante connexe de ces images qui constitue la zone d'intérêt de l'image. La figure ci-dessous montre un exemple de zone d'intérêt obtenue avec cette méthode.

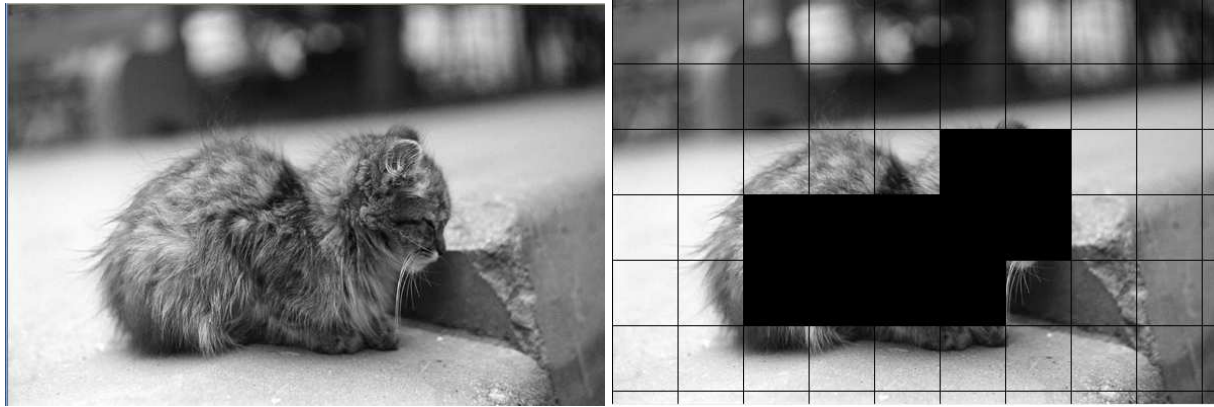


Fig . 46 : Détection de zone d'intérêt à l'aide de la dimension fractale. Image originale (à gauche) et zone d'intérêt détectée (à droite)

Cette méthode fonctionne bien sur certains types d'images comme par exemple une image composée d'un arrière plan assez uniforme avec des objets contrastés posés dessus. Cependant, la méthode n'arrive pas à détecter correctement les objets si l'image comporte beaucoup de détails. De plus, si l'image est de petite taille, alors les imageriettes qui auront une taille maximale faible et donc l'estimation fractale des imageriettes ne sera pas convenablement estimée. Les zones d'intérêt ont été correctement détectées sur 58% des images testées, avec des imageriettes de taille 32x32. Les résultats obtenus avec la dimension fractale sont donc moins bons que ce que l'on peut obtenir avec les lois de Zipf et de Zipf inverse.

4.2 Détection de zones d'intérêt par carte d'importance

Cette méthode est celle proposée par Osberger et Maeder [Os98] qui a déjà été évoquée dans le paragraphe 1.3. Elle consiste à segmenter l'image en régions et à calculer pour chaque région un ensemble de 5 facteurs d'importance qui sont ensuite combinés pour construire une carte d'importance qui indique l'importance perceptuelle de chaque région de l'image. Nous allons rappeler la définition de ces 5 facteurs d'importance :

- Le contraste de la région avec le fond. L'importance du contraste de la région R_i est calculée par cette formule :

$$I_{contrast}(R_i) = gl(R_i) - gl(R_{i-voisins}) \quad (4-2-1)$$

où $gl(R_i)$ est la moyenne des niveaux de gris de la région R_i et $gl(R_{i-voisins})$ est la moyenne des niveaux de gris de toutes les régions voisines de R_i . $I_{contrast}$ est ensuite normalisée pour appartenir à l'intervalle [0 1]. Si $I_{contrast}(R_i) = 1$ ceci correspond à un degré d'importance maximal du contraste de la région R_i . Si $I_{contrast}(R_i) = 0$ alors la région R_i n'est pas du tout importante du point de vue contraste.

- La taille de la région. Le degré d'importance de la taille d'une région est calculé comme ceci :

$$I_{size}(R_i) = \max\left(\frac{A(R_i)}{A_{max}}, 1\right) \quad (4-2-2)$$

où $A(R_i)$ est la taille de la région R_i en pixels, et A_{max} est une constante utilisée pour éviter d'accorder une importance excessive aux régions de très grande taille. Donc, c'est un point de saturation au-delà duquel toutes les régions auront la même importance telle quelle soit leur taille. A_{max} a été fixée à 1% de la taille de l'image. I_{size} est ensuite normalisée pour appartenir à l'intervalle [0..1]

– La forme de la région. L'importance due à la forme d'une région est calculée par :

$$I_{forme}(R_i) = \frac{bp(R_i)^{sp}}{A(R_i)} \quad (4-2-3)$$

où $bp(R_i)$ est le nombre de pixels de la région R_i qui sont voisins d'autres régions, et sp est une constante. On fixe sp à 1.75 pour avoir une bonne discrimination des formes. Ainsi, les régions longues et minces auront une plus grande importance due à la forme, tandis que les régions arrondies en auront moins. Ce facteur d'importance est ensuite normalisé afin d'appartenir à l'intervalle [0..1].

– La position de la région. L'importance due à la position de la région R_i est calculée par la formule :

$$I_{position}(R_i) = \frac{centre(R_i)}{A(R_i)} \quad (4-2-4)$$

où $centre(R_i)$ est le nombre de pixels de la région R_i qui se trouvent dans le 25% du centre de l'image. Ainsi, les régions localisées entièrement dans le quart central de l'image auront $I_{position}(R_i) = 1$, alors que les régions qui n'ont aucun pixel central auront $I_{position}(R_i) = 0$. Ce facteur d'importance est ensuite normalisé dans l'intervalle [0..1].

– Le fond de la région. Pour détecter le fond d'une région on calcule la proportion des bordures de l'image qui sont contenues dans chaque région. Les régions ayant un grand nombre de pixels en bordure de l'image seront classifiées comme appartenant au fond de l'image. Ces régions se verront attribuer un facteur d'importance faible. L'importance du fond est calculée par la formule :

$$I_{fond}(R_i) = 1 - \min\left(\frac{borderpix(R_i)}{0,5 \times tot_borderpix}, 1\right) \quad (4-2-5)$$

où $borderpix(R_i)$ est le nombre de pixels de la région R_i qui appartiennent à la bordure de l'image, et $tot_borderpix$ est le nombre total des pixels en bordure de l'image.

Ensuite on combine ces 5 facteurs pour chaque région afin de produire une carte d'importance (IM) globale pour l'image. Il y a peu de données quantitatives qui indiquent l'importance relative de ces différents facteurs, et ces données, même si elles existent, peuvent changer d'une image à une autre. Par conséquent, il faut traiter tous les facteurs comme étant d'importance égale. Toutefois, si on sait a priori qu'un facteur particulier est plus important qu'un autre, cette information peut être exploitée lors du calcul de l'IM. Les régions qui sont

très bien classées pour certains facteurs doivent avoir une importance plus élevée. Si on fait une simple moyenne des facteurs d'importance, ceci ne donnera pas le résultat escompté. Ainsi, pour calculer l'IM d'une région, il est préférable de faire la somme des facteurs élevés au carré comme indiqué par cette formule :

$$IM(R_i) = \sum_{k=1}^5 (I_k(R_i))^2 \quad (4-2-6)$$

La carte d'importance finale est obtenue par la normalisation de ce résultat. Ainsi, la région dont l'importance est la plus élevée aura une importance de 1.0.

La segmentation de l'image peut se faire avec différentes méthodes. Deux méthodes ont été testées, la méthode Split and Merge et la méthode CSC (Color Structure Code).

L'algorithme Split and Merge a été conçu en 1974 par Pavlidis et Horowitz. [Pav-74] Il est composé d'une phase de découpage (Split) et d'une phase de regroupement (Merge).

La méthode de découpage de l'image utilisée dans cet algorithme est basée sur la notion de quadtree. Cette structure de données est un arbre quaternaire qui permet de stocker l'image à plusieurs niveaux de résolution. On part d'une région initiale qui est l'image tout entière. Si cette image vérifie un critère d'homogénéité de couleur ou de niveau de gris, l'algorithme s'arrête. Sinon, on découpe cette région en quatre parties de même taille et on lance la procédure récursivement dans ces quatre parties. La région initiale va être stockée comme un nœud dans un graphe et les sous-parties comme des fils de ce nœud.

La phase de découpage aboutit à un nombre de régions trop élevé. La cause fondamentale de cette sur-segmentation est que l'algorithme découpe les régions de manière arbitraire. C'est pourquoi il est nécessaire de procéder à une fusion de régions après le découpage. Pour cela, on cherche tous les couples de régions adjacentes dans l'arbre issu du découpage et on les fusionne si leur couleur ou leur niveau de gris est assez proche. Pour réaliser cette fusion, il faut d'abord tenir à jour une liste des contacts entre régions. On obtient ainsi un graphe d'adjacence de régions. Ensuite, l'algorithme va marquer toutes les régions comme non traitées et choisir la première région R non traitée disponible. Les régions en contact avec R sont empilées et sont examinées les unes après les autres pour savoir si elles doivent fusionner avec R. Si c'est le cas, la couleur moyenne de R est mise à jour et les régions en contact avec la région fusionnée sont ajoutées à la pile des régions à comparer avec R. La région fusionnée est marquée traitée. Une fois la pile vide, l'algorithme choisit la prochaine région non traitée et recommence, jusqu'à ce que toutes les régions soient traitées.

L'algorithme CSC (Color Structure Code) a été présenté en 1993 par Rehrmann et Priesse [Reh-93]. Il repose sur l'utilisation d'une structure hexagonale hiérarchique de codage de l'image. L'une des propriétés de cet algorithme est qu'il est complètement parallélisable et qu'il ne dépend pas d'un ordre de traitement de l'image comme les algorithmes d'accroissement de région. L'algorithme CSC utilise à la fois des informations locales pour la phase de regroupement et globales pour la phase de découpage.

L'algorithme repose sur l'utilisation d'une topologie hiérarchique particulière des pixels qui est constituée d'îlots hexagonaux disposés à différents niveaux. Les îlots de niveau 0 sont des groupements hexagonaux de sept pixels. La disposition légèrement décalée des pixels de rang pair et de rang impair peut être soit interpolée soit simulée par l'intermédiaire d'une topologie légèrement différente plaquée sur le réseau orthogonal de l'image comme on peut le voir sur la figure ci-dessous. Cette topologie est récursive, c'est à dire qu'un îlot de niveau n+1 est constitué de sept îlots de niveau n, disposés hexagonalement de la même manière que les pixels constituant un îlot de niveau 0. Les îlots d'un niveau donné se recouvrent partiellement. Cette propriété est particulièrement importante pour l'algorithme puisqu'elle permet des simplifications et des optimisations significatives. La construction récursive des îlots s'arrête

lorsqu'un îlot recouvre toute l'image. Le nombre d'îlots décroît d'un facteur 4 d'un niveau à l'autre. La première phase de l'algorithme consiste à appliquer un simple algorithme d'accroissement de région localement dans chaque îlot de niveau 0. Cette première étape construit entre 1 et 7 régions de niveau 0, contenues dans chaque îlot. On appelle ces régions des éléments.

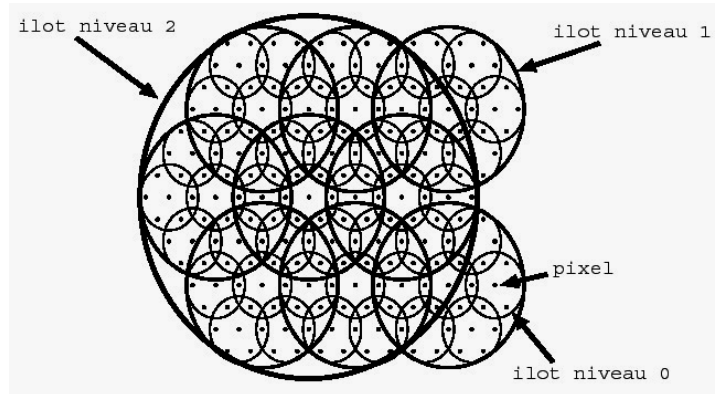


Fig. 47 : Topologie hiérarchique de la segmentation CSC

Dans la phase de regroupement de l'algorithme, les éléments de niveau n sont regroupés en éléments de niveau $n + 1$. Cette opération s'effectue à l'intérieur d'un îlot de niveau $n + 1$. Les éléments de niveau n sont regroupés s'ils sont connectés et de couleur ou de niveau de gris proche. L'opération est répétée de niveau en niveau jusqu'au niveau recouvrant toute l'image. Lorsqu'un élément de niveau n ne trouve aucun autre élément auquel se lier, il devient la racine d'un arbre correspondant à une région terminée de l'image et peut être stocké.

Chaque îlot de niveau n contient un ensemble d'éléments de niveau n , issus du traitement récursif à l'étape précédente. Les éléments de niveau n sont ensuite considérés par l'îlot de niveau $n + 1$ auquel ils appartiennent. Finalement, les éléments de niveau n sont regroupés au sein d'éléments de niveau $n + 1$. On obtient ainsi une structure arborescente.

La phase de découpage a lieu simultanément à la phase de regroupement. Lorsque deux éléments de niveau n sont connectés et de couleur similaire, ils sont regroupés dans un élément de niveau $n+1$. Dans le cas où ils sont connectés mais de couleur très différente ils ne vont pas être regroupés, bien que tous les éléments de niveau $n - 1$ les constituant soient de couleurs homogènes et proches. Lorsque deux éléments connectés ne sont pas regroupés parce que leurs couleurs sont trop éloignées, ils partagent cependant un ensemble de sous éléments puisqu'ils sont connectés. Ceci est dû à la structure partiellement recouvrante des îlots. Il est donc nécessaire de les séparer pour obtenir deux éléments sans recouvrement et partager la région commune.

La méthode a été testée sur la même base d'images qui a été utilisée pour tester les méthodes basées sur les lois puissance. Pour des images simples comme un visage ou un objet unique, la méthode identifie systématiquement les régions qui attirent notre attention, avec plus ou moins de précision selon la méthode de segmentation utilisée. Dans la carte d'importance finale (Final IM), les zones d'intérêt seront plus claires que les autres régions présentant un intérêt visuel moins important. Les résultats pour des images plus complexes comportant de nombreuses textures différentes sont conditionnés par le résultat de la phase de segmentation. Généralement, la segmentation CSC sur des images bruitées produit beaucoup de régions. Les régions produites par la segmentation CSC sont en général plus précises car elles suivent les contours des formes. La détection des zones d'intérêt avec la segmentation CSC est

légèrement plus rapide qu'avec Split and Merge. La figure 48 montre les régions obtenues par les deux types de segmentation et la figure 49 montre les régions d'intérêt correspondantes sur l'image.

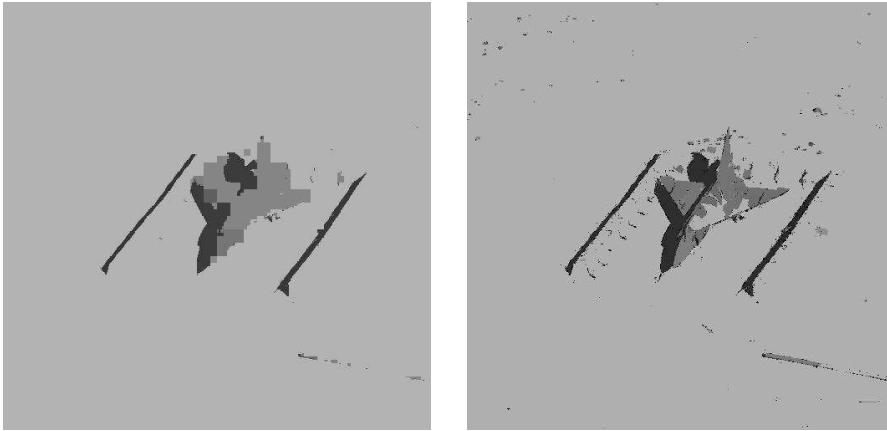


Fig. 48 : Segmentation d'une même image avec l'algorithme Split and Merge (à gauche) et avec l'algorithme CSC (à droite)

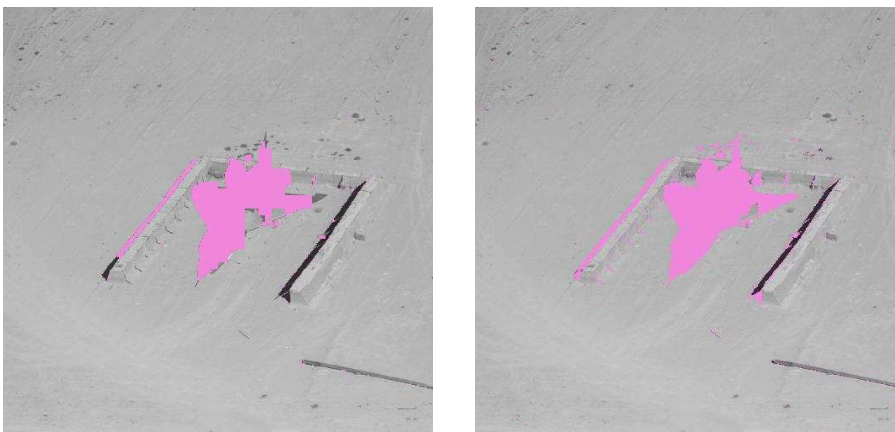


Fig . 49 : Zones d'intérêt détectées avec la méthode Split and Merge (à gauche) et avec la méthode CSC (à droite)

Cette méthode a de très bonnes performances pour la détection des zones d'intérêt. Sur la base de test utilisée, la zone d'intérêt a été détectée sur 83% des images testées avec l'algorithme Split and Merge et 86% avec l'algorithme CSC. La méthode permet de détecter les zones d'intérêt sur la plupart des types d'images, y compris celles avec lesquelles on avait du mal à détecter les zones d'intérêt avec les lois puissance. Cependant cette méthode a tendance à détecter des zones d'intérêt trop petites sur certaines images, si la zone d'intérêt est constituée de plusieurs régions, on risque de n'en prendre en compte qu'une seule et de ne détecter qu'une partie de la véritable zone d'intérêt de l'image. Ce phénomène est plus sensible avec la segmentation par l'algorithme CSC. Les performances de la méthode de Osberger et Maeder sont comparables à celles de la méthode de la loi de Zipf inverse en ce qui concerne les taux de détection réussie, la différence étant la taille des régions d'intérêt détectées par les deux méthodes, plus grandes avec Zipf inverse. La méthode de Zipf inverse présente également l'avantage d'avoir un temps de calcul plus réduit.

Conclusion

La détection de zones d'intérêt est un problème particulièrement difficile à résoudre lorsque l'on n'a aucune connaissance a priori sur le contenu de la scène. Cette difficulté tient en partie au caractère subjectif de la notion de région d'intérêt. Certaines Diverses méthodes ont été proposées pour tenter de résoudre ce problème, basées sur des modélisations de la vision humaine ou sur l'étude de caractéristiques structurelles de l'image. L'utilisation des modèles de lois puissance telles que la loi de Zipf et la loi de Zipf inverse permettent de définir une nouvelle approche de ce problème. La loi de Zipf appliquée avec le codage des rangs généraux permet de détecter des objets artificiels dans des milieux naturels. La mesure de la différence des aires sous les courbes ou des pentes des courbes de Zipf d'une imagerie par rapport aux imageries voisines permet de détecter la position d'un objet artificiel dans l'image. Un prétraitement basé sur un calcul de gradient permet d'améliorer les résultats, ainsi que l'utilisation d'un codage à seuil.

Les lois de Zipf et de Zipf inverse peuvent être utilisées pour détecter des zones d'intérêt dans des images quelconques, en utilisant une méthode de classification par partition d'un nuage de points représentant les caractéristiques des courbes. L'utilisation de la loi de Zipf inverse avec le codage des 9 classes donne les meilleurs résultats. Elle permet de détecter les zones d'intérêt dans les images les plus diverses. Cependant, les zones d'intérêt détectées sont souvent plus grandes que la zone d'intérêt réelle de l'image, même si ce n'est pas nécessairement un défaut dans certaines applications telle que la définition automatique de la zone d'intérêt pour la compression JPEG2000. En effet, dans ce cas il est préférable que la zone d'intérêt soit trop grande que trop petite, en effet la différence de taux de compression entre la zone d'intérêt et le fond sera moins visible si la séparation entre les deux zones n'affecte que l'arrière-plan de l'image. Une classification par réseaux de neurones est moins intéressante que la classification sans apprentissage par partition d'un nuage de points, du fait de la grande diversité des images à traiter.

La méthode de détection de zone d'intérêt basée sur la loi de Zipf inverse a de meilleures performances que le calcul de la dimension fractale par box-counting 3D et des performances équivalentes à la méthode de Osberger et Maeder, qui est basée sur une segmentation de l'image et l'extraction de caractéristiques visuelles.

Conclusion et perspectives

Dans cette thèse, nous avons montré qu'il est possible d'utiliser la loi de Zipf en analyse d'images. Ce modèle issu de l'analyse linguistique avait déjà été utilisé dans d'autres domaines pour modéliser différents phénomènes naturels ou issus des productions humaines mais jamais encore auparavant dans le domaine des images, et son application dans ce nouveau domaine a donné des résultats intéressants. Deux modèles de lois puissance peuvent être utilisés pour décrire la fréquence d'appartition des motifs de l'image, la loi de Zipf, qui donne la fréquence d'apparition de chacun des motifs en fonction de leur rang dans une séquence, et la loi de Zipf inverse, qui donne le nombre de motifs distincts ayant une fréquence d'appartition donnée. Plusieurs codages peuvent être utilisés pour représenter les motifs de l'image, notamment la partition en n classes de l'échelle des niveaux de gris et le codage des rangs généraux avec ou sans seuils. L'utilisation de ces différents codage permet de caractériser le contenu structurel de l'image, suivant le codage utilisé les caractéristiques mises en évidence ne seront pas les mêmes. Anisi, le codage des n classes permet de caractériser les zones homogènes de l'image et le codage des rangs généraux permet de mettre en évidence les détails et les contours. L'application des modèles de lois puissance permet de donner une estimation de la complexité de l'image.

Plusieurs applications de la loi de Zipf en analyse d'image peuvent être envisagées, notamment la mesure de la dégradation d'une image compressée, et la détection de zones d'intérêt dans les images. En ce qui concerne la compression, différents critères de mesure de la perte de qualité d'une image compressée par rapport à l'original ont été définis, ces critères se basent sur des mesure de différence entre les courbes de Zipf associées aux deux images. Plusieurs critères ont été définis et chacun est adapté à un type particulier de compression, mais il n'y a pas de critère unique permettant d'estimer de façon correcte la dégradation sur tous les types de compression. Ces critères de qualité peuvent être utilisés pour une optimisation automatique de la compression d'une image, afin de minimiser la taille du fichier tout en préservant une qualité visuelle acceptable.

Pour la détection d'objets et de zones d'intérêt dans les images, deux approches ont été envisagées. La première approche concerne la détection d'un objet artificiel dans un milieu naturel, une méthode basée sur la segmentation de l'image en imagerie et le calcul des courbes de Zipf à l'aide du codage des rangs généraux. Avec cette méthode, il est possible de détecter la position d'un objet artificiel dans l'image en recherchant le maximum d'une mesure de différence entre les caractéristiques de la courbe de Zipf associée à une imagerie et celle des imagerie voisines.

La seconde approche est beaucoup plus générale, il s'agit de détecter la zone d'intérêt correspondant à la perception qu'en aurait un être humain dans une image photographique quelconque. La méthode de détection consiste en un découpage de l'image en imagerie et en une classification des imagerie en fonction de différentes caractéristiques des courbes de Zipf et/ou de Zipf inverse. La méthode de classification employée, qui consiste en un partitionnement de nuage de points, présente l'avantage de ne pas nécessiter de phase d'apprentissage explicite, elle permet de meilleures performances que les méthodes utilisant un apprentissage telles que les réseaux de neurones. Le modèle de la loi de Zipf inverse est celui qui donne les meilleurs résultats dans cette application, ses performances sont au moins aussi bonnes que celles des méthodes de détection basées sur un modèle de la vision humaine.

D'autres applications de la loi de Zipf en analyse d'image que celles qui ont été présentées ici sont envisageables. Dans le domaine des images fixes, la loi de Zipf pourrait être également utilisée pour l'indexation et la recherche par le contenu dans une base d'images. Mais il est également possible d'étendre son utilisation aux cas des séquences vidéo. Dans ce cas, on peut définir des codages des motifs qui soient spécifiques à ces problèmes, comme ceux qui décrivent l'évolution des pixels entre les images successives d'une séquence. Les applications possibles de la loi de Zipf dans le domaine de la vidéo concernent en particulier la détection des changements de plan, le suivi d'un objet dans la séquence ou la détection d'images subliminales.

Bibliographie

- [Aue-13] F. Auerbach, Das Gesetz der Bevölkerungskonzentration, Petermanns Geographische Mitteilungen, LIX:73-76, 1913
- [Bi-97] D. Bi, Segmentation d'images basée sur les statistiques de rangs des niveaux de gris. Thèse de doctorat, Université de Tours, 1997
- [Bil-69] G. Billmeier, Worthauffigkeiten vom Zipfschen Typ, überprüft an deutschem Textmaterial, Hamburg: Buske, 1969
- [Bre-99] L. Breslau, P. Cao, L. Fan, G. Phillips, S. Shenker "Web caching and Zipf-like distributions: evidence and implications", Proceedings of INFOCOM'99 (IEEE Press), 1999
- [Car-90] M. J. Carlotto, M. C. Stein, A Method for Searching Artificial Objects on Planetary Surfaces, Journal of the British Interplanetary Society, vol.43, 1990
- [CCI-82] CCIR .Rec. 500-2, Method for the subjective assessment of the quality of television pictures, Recommendations and reports of the CCIR, Geneva, 1982.
- [Cha-53] D. Champernowne, A Model of Income Distribution, Economic Journal, LXIII, pp. 318-351, 1953
- [Coh-97] A. Cohen, R.N. Mantegna, S. Havlin, Numerical Analysis of Word Frequencies in Artificial and natural Language Texts, Fractals, Vol.5 No.1, 95-104, 1997
- [Cun-95] C. Cunha, A. Bestavros, and M. Crovella Characteristics of WWW client-based traces. Technical Report TR-95-010, Boston University, Computer Science Dept., Boston, MA 02215, USA, April 1995
- [Dah-79] H. Dahl, Word Frequencies of Spoken American (Verbatim), 1979
- [Dal-93] S. Daly, The Visible Differences Predictor, in Digital Images and Human Vision, edited by A.B. Watson, A Bradford Book, MIT Press, Cambridge, 1993
- [Dar-53] J.H. Darwin, Population differences between species growing according to simple birth and death process, Biometrika, 40, 370, 1953
- [Dau-92] I. Daubechies, Ten Lectures on Wavelets. SIAM, Philadelphia, Pennsylvania, 1992
- [Del-02] E. Dellandrea, P. Makris, M. Boiron, N. Vincent A medical acoustic signal analysis method based on Zipf law. International Conference on Digital Signal Processing (DSP 2002). Santorini, Greece, Vol. 2. p. 615-618., 2002
- [DiG-97] V. Di Gesu, C. Valenti, L. Strinati, Local operators to detect regions of interest, Pattern Recognition Letters vol. 18, pp.1077-1081, 1997

- [Esk-00] A.M Eskicioglu, Quality Measurement for Monochrome Compressed Images in the Past 25 Years Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) Conference, Vol. 4, pp.1907-1910, Istanbul, Turkey, June 5-9, 2000
- [Esk-95] A. M. Eskicioglu, P. S. Fisher and S. Chen, "Image Quality Measures and Their Performance," IEEE Transactions on Communications, Vol. 43, No. 12, pp. 2959-2965, December 1995
- [Esk-96] A. M. Eskicioglu, Application of Multi-Dimensional Measures to Reconstructed Medical Images, Optical Engineering, Vol. 35, No. 3, pp. 778-785, March 1996
- [Est-16] J.B. Estoup, Gammes Sténographiques, Institut Sténographique de France, Paris, 1916
- [Fan-00] G. Fan and X.-G. Xia, Multiscale Texture Segmentation Using Hybrid Contextual Labeling Tree, in Proc. of the IEEE International Conference on Image Processing (ICIP2000), Vancouver, Canada, Sept. 2000.
- [Fin-80] J. Findlay. The visual stimulus for saccadic eye movement in human observers. Perception, 9:7–21, Sept. 1980
- [Fra-98] P. Fränti: Blockwise Distortion Measure for Statistical and Structural Errors in Digital Images, Signal Processing: Image Communication, 13 (2), 89-98, August 1998
- [Fre-91] W.T. Freeman, E.H. Adelson, The Design and Use Of Steerable Filters, IEEE PAMI, vol. 13 n° 9 ,pp.891-906, 1991
- [Gab-99] X. Gabaix , Zipf's law for cities: an explanation, Quarterly Journal of Economics, 114:739-767, 1999
- [Gai-02] J.Gaite, S.Manrubia, Scaling of void sizes and fractality in the galaxy distribution, Monthly Notice of the Royal Astronomical Society, Vol.335, No.4, pp. 977-983, 2002
- [Gib-31] R.Gibrat, Les Inégalites Economiques. Librairie du Recueil, Paris ,1931
- [Gla-94] S. Glassman, A Caching Relay for the World Wide Web. Computer Networks and ISDN systems, First International Conference on the World-Wide Web, Elsevier Science BV, 1994
- [Gue-01] A.Guérin-Dugué, P. Ladret, J. Herault, Catégorisation et Ordonnancement de Scènes, Forme de Spectre, Statistique d'Orientations et de Couleur, ASTI' 2001, Paris, 2001
- [Hal-81] C. F. Hall, Subjective Evaluation of a Perceptual Quality Metric, Proceedings of SPIE, Vol. 310, pp.200-204, 1981.
- [Har-85] D. Harwood, M. Subbarao, and L.S. Davis, Texture classification by local rank correlation, CVGIP 32, 1985, 404-411
- [Hav-95] S. Havlin, The Distance between Zipf Plots, Physica A216, 148-150, 1995

[Her-66] G. Herdan, The Advanced theory of Language as Choice and Chance, Springer-Verlag, New York, 1966

[Hos-86] K. Hosaka, A New Picture Quality Evaluation Method, Proceedings of the International Picture Coding Symposium PCS '86, Tokyo, Japan, pp. 17-18, April 1986

[Hub-88] D.H. Hubel, Eye, Brain and Vision, W.H. Freeman & Co., New York, 1988

[Huf-52] D. A. Huffman, A method for the construction of minimum-redundancy codes, Proceedings of the IRE, vol. 40 no. 9, pp. 1098-1101, 1952

[Imb-83] M. Imbert, La neurobiologie de l'image, La Recherche, Mai 1983

[Ioa-01] Y.M. Ioannides, H. G. Overman Zipf's Law for Cities: An Empirical

[ISO-00] ISO/IEC JTC1/SC29 WG1, JPEG2000 Part 1 Final Comitee Draft Version 1.0, March 2000.

[Itt-98] L. Itti, C. Koch, E. Niebur, A model of saliency-based visual attention for rapid scene analysis, IEEE Trans Patt Anal Mach Intell. 20(11), pp. 1254-1259, 1998

[Jai-89] A. K. Jain, Fundamentals of Digital Image Processing, Prentice-Hall, Inc., Saddle River, NJ, USA, 1989.

[Kad-01] T.Kadir, M.Brady, Saliency, Scale and image Description, International Journal of Computer Vision, Volume 45 Issue 2 November 2001

[Kad-03] T.Kadir, M. Brady, Scale Saliency : A Novel Approach to Salient Feature and Scale Selection. International Conference Visual Information Engineering 2003. Pages 25-28, 2003

[Kal-64] V.M. Kalinin, Razvitie schemy Puassona i ee primenenie dlja statisticeskich svojstv reci, Leningrad, 1964

[Kuc-67] Henry Kucera, W Nelsen Francis, Computational Analysis of Present-Day American English (Brown Univ Press), 1967

[Lai-97] Y.-K. Lai ,C.-C. J. Kuo, "Image quality measurement using the Haar wavelet," in SPIE: Wavelet Applications in Signal and Image Processing V, June 1997

[Lef-02] S. Lefèvre, Détection d'événements dans une séquence vidéo, Thèse de Doctorat. Université de Tours. Décembre 2002.

[LeQ-02] Le Quan Ha, E. Sicilia, J. Ming and F. J. Smith, "Extension of Zipf's law to words and phrases," International Conference on Computational Linguistics (COLING'2002), pp. 315-320, Taipei, Taiwan, Aug 2002.

[Li-92] W. Li "Random texts exhibit Zipf's-law-like word frequency distribution", IEEE Transactions on Information Theory , 38(6):1842-1845, 1992

- [Li-98] B.Li, G.W.Meyer, R. V. Klassen, "A Comparison of Two Image Quality Models," SPIE Conf. on Human Vision and Electronic Imaging III, Vol. 3299, San Jose, January 1998
- [Loy-03] Gareth Loy, Alexander Zelinsky: Fast Radial Symmetry for Detecting Points of Interest. IEEE Transactions on Pattern Analysis and Machine Intelligence 25(8), 959-973, 2003
- [Lua-94] K.T. Lua, Frequency-rank curves and entropy for Chinese characters and words. Computer Processing of Chinese & Oriental Languages, vol.8, no.1: 37-52, 1994
- [Lub-95] J. Lubin, "A visual discrimination model for imaging system design and evaluation", in Vision Models for Target Detection, E. Peli, Ed. pp. 245-283, World Scientific, Singapore, 1995.
- [Luo-01] J. Luo, A. Singhal: On measuring low-level self and relative saliency in photographic images. Pattern Recognition Letters 22(2): 157-169, 2001
- [Man-01] B. Manaris, C. McCormick, and T. Purewal, "Searching for Beauty in Music-Applications of Zipf's Law in MIDI-Encoded Music," 2001 Sigma Xi Forum, "Science, the Arts and the Humanities: Connections and Collisions", Raleigh, NC, November 8-9, 2001
- [Man-53] B.B. Mandelbrot, An informational theory of the statistical structure of languages, in Communication Theory, ed. W. Jackson, Butterworth, 1953 , pp. 486-502.
- [Man-74] J. L. Mannos and D. L. Sakrison, The Effects of a Visual Fidelity Criterion on the Encoding of Images, IEEE Transactions on Information Theory, Vol. 20, No.4, pp. 525-536, July 1974
- [Man-83] B.B. Mandelbrot, The Fractal Geometry of Nature, Freeman, New York, 1983.
- [Mar-83] D. Marr, Vision, W.H. Freeman and Co., New York, 1982
- [Mar-84] D.M. Mark, P.B. Aronson, Scale-Dependent Fractal Dimensions of Topographic Surfaces : an Empirical Investigation with Applications in Geomorphology and Computer Mapping, Mathematical Geology, vol .16 no. 7, 1984
- [Mey-93] Y.Meyer, Les ondelettes, Algorithmes et applications, Editions Armand Colin, Paris, 1993.
- [Mil-57] G.A. Miller, Some effects of intermittent silence, American Journal of Psychology, 70:311-314, 1957
- [Mil-58] G.A. Miller, E.B. Newman, Tests of a statistical explanation of the rank-frequency relation for words in written English, American Journal of Psychology, 71, 209-218, 1958
- [Mit-01] M. Mitzenmacher, A brief history of generative models for power law and lognormal distributions. In Proceedings of the Thirty-Ninth Annual Allerton Conference on Communication, Control, and Computing, pages 182-191, 2001

- [Miy-98] M. Miyahara, K. Kotani, and V. R. Algazi. Objective Picture Quality Scale (PQS) For Image Coding. *IEEE Trans. on Communications*, September 1998
- [Nil-92] N. B. Nill, B. H. Bouzas, "Objective Image Quality Measure Derived From Digital Image Power Spectra, *Optical Engineering*, Vol. 31, No. 4, pp. 813-825, 1992
- [Osb-98a] W. Osberger, N.W. Bergmann, A.J. Maeder: An Automatic Image Quality Assessment Technique Incorporating Higher Level Perceptual Factors. *Proceedings of the 1998 IEEE International Conference on Image Processing (ICIP-98)*, Chicago, Illinois, October 4-7, 1998
- [Osb-98b] W. Osberger and A.J. Maeder, "Automatic identification of perceptually important regions in an image," *11th international conference on Pattern Recognition*, no. 1, pp. 701-704, 1998
- [Par-97] V. Pareto, *Cours d'économie politique*, Rouge, Lausanne et Paris, 1897
- [Pat-92] D. Patel, T.J Stoneham, Texture Image classification and segmentation using rank-order clustering, *Proc. 11th TCPR*, Den Haag, 1992
- [Pav-74] T. Pavlidis, L. Horowitz, L. Segmentation of plane curves, *IEEE Transactions on Computers*, C-23(8), pp. 860-869, 1974
- [Pen-93] W. Pennebaker and J. Mitchell, *JPEG Still Image Data Compression Standard*, Van Nostrand Reinhold, New York, 1993
- [Pra-78] W. K. Pratt, *Digital Image Processing*, John Wiley and Sons, Inc., USA, 1978
- [Pri-98] C. Privitera, L.W. Stark: Evaluating image processing algorithms that predict regions of interest. *Pattern Recognition Letters* 19(11): 1037-1043 ,1998
- [Ree-01] W.J Reed, The Pareto, Zipf and other power laws, *Economics Letters* 74:15-19, 2001
- [Ree-02] W.J. Reed, B. D. Hughes. From gene families and genera to incomes and internet file sizes: why power-laws are so common in nature, *Phys. Rev. E* 66 067103, 2002
- [Reh-93] L. Priese, V. Rehrmann, A Fast Hybrid Color Segmentation Method, In *Mustererkennung 93*, 15. DAGM-Symposium Lübeck, pp. 297-304. Springer-Verlag, Berlin, 1993
- [Rei-97] P. Reinagel, A.M. Zador, The Effect of Gaze on Natural Scene Statistics, *Neural information and coding workshop*, Snowbird, Utah, 16-20 Mar. 1997
- [Sai-96] A. Said, W. A. Pearlman, "A new fast and efficient image codec based on set partitioning in hierarchical trees," *IEEE. Trans. Circ. Syst. Video Tech.* 6, pp. 243--250, 1996
- [Sha-48] C.E. Shannon. *A Mathematical Theory of Communication*. *Bell System Tech. Journal*, 27 : 379-423, 1948

- [She-02] Z. Wang, H. R. Sheikh and A. C. Bovik, "No-reference perceptual quality assessment of JPEG compressed images," IEEE International Conference on Image Processing, Sept. 2002
- [Sie-56] S.Siegel, Nonparametric Statistics for the Behavioral Sciences, Mc Graw Hill, New York, 1956
- [Sim-55] H.A .Simon, "On a class of skew distribution functions", Biometrika, 42:425-440,1955
- [Ste-01] F. Stentiford, An Estimator for Visual Attention through Competitive Novelty with Application to Image Compression, Picture Coding Symposium, Seoul, Korea, pp. 25-27, April 2001
- [Sye-97] T.F. Syeda-Mahmood, Data and model-driven selection using color regions, International Journal of Computer Vision (IJCV), vol. 21, no.1, pp.9-36, 1997
- [Tam-03] A. Tamtaoui, D. Aboutajdine, Mesure de la qualité visuelle des images compressées, Traitement du signal, vol.20 n°1, pp.43-53, 2003
- [Tau-01] D. Taubman, M. Marcellin, JPEG2000: Image compression fundamentals, standards and practice, Kluwer Academic Publishers, Boston, MA, 2001
- [Vos-75] R.F. Voss, J. Clarke. 1/f Noise in music and Speech, Nature 258, pp.317-318, 1975
- [Wal-91] G. K. Wallace. The JPEG still picture compression standard. Communications of the ACM, 34(4):30--44, Apr. 1991.
- [Wan-01] J. Z. Wang, J. Li, R. M. Gray, G. Wiederhold, "Unsupervised multiresolution segmentation for images with low depth of field," IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI), vol. 23, no. 1, pp. 85-90, 2001.
- [Wan-02a] Z. Wang, H. R. Sheikh, A. C. Bovik, No-reference perceptual quality assessment of JPEG compressed images, IEEE International Conference on Image Processing, Sept. 2002.
- [Wan-02b] Z. Wang, and A. C. Bovik, "A universal image quality index," IEEE Signal Processing Letters, vol. 9, no. 3, pp. 81-84, 2002
- [Wan-90] L.Wang, D.C. He, Texture classification using texture spectrum, Pattern Recognition, Vol.23, No 8, pp. 905-910, 1990
- [Wat-97] A.B Watson, R.Borthwick, M.Taylor, Image quality and entropy masking, Proceedings of the SPIE Human Vision and Electronic Imaging II, vol. 3016, pp.2-12, 1997
- [Yar-67] A. Yarbus. Eye Movements and Vision. Plenum Press, NewYork NY, 1967
- [Ye -02] Zhen Ye, Cheng-Chang Lu, Unsupervised Multiscale Focused Objects Detection Using Hidden Markov Tree, Proceeding of the 4th International Conference on Computer Vision, Pattern Recognition & Image Processing, pp. 812-815, 2002

[Yul-25] GU Yule, "A mathematical theory of evolution based on the conclusions of Dr. J.C. Willis, F.R.S. ", *Philosophical Transactions of the Royal Society of London (Series B)*, 213:21-87 ,1925

[Zek-92] S. Zeki , *Les images visuelles. Pour la science*. 181, p. 60-67., nov. 1992

[Zip-35] G.K. Zipf, *Psycho-Biology of Languages*, Houghton-Mifflin, 1935.

[Zip-49] G.K. Zipf, *Human Behavior and the Principle of Least Effort*, Addison-Wesley, 1949