

THÈSE

présentée en vue de
l'obtention du titre de

DOCTEUR

de

**L'ÉCOLE NATIONALE SUPÉRIEURE
DE L'AÉRONAUTIQUE ET DE L'ESPACE**

SPÉCIALITÉ : SIGNAUX, IMAGES ET COMMUNICATIONS

par

Mathieu JUNG

ÉVALUATION UNIVARIANTE DE LA QUALITÉ DES IMAGES PAR UNE APPROCHE NEURONALE

Soutenue le 23 octobre 2000 devant la Commission d'Examen :

| | | |
|------------|-------------------|---------------------------|
| MM. | P. BOURRET | Président |
| | D. BARBA | Rapporteur |
| | M. GAZALET | Directeur de thèse |
| | D. LEGER | |
| | A. TREMEAU | Rapporteur |
| | C. VALORGE | |

Thèse préparée au sein du département d'Optique Théorique et Appliquée de l'ONERA - Centre de Toulouse.

Résumé

Disposer d'un outil universel d'évaluation automatique de la qualité des images est très utile lorsque le nombre d'images à traiter est important. C'est le cas quand il s'agit d'optimiser les paramètres caractéristiques d'un système imageur. La plupart des méthodes d'évaluation de la qualité des images sont des méthodes bivariantes, c'est-à-dire reposant sur une comparaison entre une image dégradée et la même image parfaite. Très souvent, cependant, la référence n'existe pas. Quelques méthodes univariantes, c'est-à-dire sans image de référence, ont été développées mais les résultats sont encore peu probants.

Cette thèse propose une nouvelle approche pour évaluer de façon univariante la qualité d'une image. Cette approche repose sur l'utilisation de réseaux de neurones. La démarche proposée peut être appliquée à tout type de dégradation. Elle comporte trois étapes. Il faut tout d'abord trouver sur l'image dégradée les caractéristiques qui permettent d'évaluer sa qualité. Il faut ensuite étalonner un modèle permettant d'associer ces caractéristiques et la note de qualité attendue. Enfin, il faut vérifier la stabilité du modèle sur un grand nombre d'images.

Les caractéristiques de la dégradation dépendent non seulement de son importance, mais aussi du type d'image. Ces informations sont extraites de l'image par des traitements et calculs systématiques. Elles sont regroupées sous la forme d'un *vecteur caractéristique*. Ce vecteur sert d'entrée au modèle univariant. Le modèle choisi est un réseau de neurones (RN). L'étalonnage des paramètres du RN se fait sur un grand nombre d'exemples connus que l'on appelle *base d'apprentissage*. Ils sont choisis de manière à obtenir un échantillon représentatif de tous les types d'images que l'on peut rencontrer dans l'application désirée.

La méthode est appliquée à différents contextes. Entre autres, elle permet d'estimer la qualité visuelle d'images comprimées JPEG, d'estimer en vol le défaut de mise au point d'un instrument satellitaire ou de modéliser la détection visuelle de défauts ponctuels sur un image. La précision des résultats est comparée à celle obtenue dans le même contexte avec des critères de qualité bivariants.

Mots clés : Qualité image, évaluation univariante de qualité, vision, réseaux de neurones

Summary

Assessing automatically the quality of images becomes very useful when the number of images to be processed is large. It happens, for instance, when optimizing the characteristics of an imaging system. To compute the quality of images, most methods compare a degraded image to a perfect reference. However, very often, the reference image is unavailable. Some univariant methods, i. e. those not requiring the reference image, have been developed, but the results are not yet very convincing.

The object of this thesis is to present an original univariant method based on the use of artificial neural networks. This method can be applied to any types of defect. It allows to reproduce the comportment of any bivariant method. It involves three steps. First, characterization of the defect is extracted mathematically from the image. Then the neural network is taught how to establish a relation between the mathematical characterization of the defect and the bivariant mark. Finally, the quality of unknown images has to be assessed to check the reliability of the model.

The characteristics of the defect depend not only on its importance but also on the type of image. The characteristics of the defect are extracted from the image very systematically and using any mathematical or signal processing tool. Those pieces of information constitute the "characteristic vector" of the defect for the image. This vector is used as input for the univariant model. As a model, a neural network (a very powerful non linear model) was used. The calibration of the model is done on a large pool of familiar examples. They are chosen in order to get a representative sample of all the types of images that may be found in the expected application.

The method can be applied in different contexts : quality assessment of JPEG compressed images, in-flight estimation of a focus problem on satellites, modelisation of the visual detection of small defects on images. Accuracy of the results is compared to the one that is obtained in the same context with the reference bivariant method.

Keywords : Image quality, univariant quality assessment, vision, neural networks

Remerciements

Je tiens à remercier René Jalin, directeur du DOTA, ainsi que Jean-Michel Maisonneuve, adjoint, qui m'ont accueilli dans leur département. Je tiens également à remercier Marc Gzalet, Dominique Léger, Paul Bourret et Jean Loup Bussenot pour leur encadrement et leurs conseils qui m'ont été précieux. Sans oublier Bernard Petit (SPOTIMAGE) et Aimé Meygret (CNES) qui m'ont gracieusement donné plusieurs dizaines d'images SPOT au cours de la thèse.

Un merci de plus pour Dominique, le meilleur correcteur orthographique et grammatical sur le marché.

Il y a aussi tous ceux que j'ai embêtés, plus ou moins régulièrement, pour des problèmes parfois futiles : Xavier, Françoise, Christophe, Sophie, Laurent, Laurent, Laurent, Alain, Véronique.

Merci aussi à tous les autres pour leur soutien et leur bonne humeur : Geneviève, Re Xavier pour ses blagues à deux balles, Joël, Philippe, Philippe, Bruno, Yvan et j'en oublie.

Et puis enfin, merci à ma famille et à mes proches pour leurs encouragements dans les moments difficiles.

Table des matières

| | |
|---|-----------|
| <i>Remerciements</i> | <i>1</i> |
| <i>Table des matières</i> | <i>3</i> |
| <i>Introduction</i> | <i>5</i> |
| Chapitre 1 : Etat de l'art des critères de qualité image | 9 |
| I - Introduction – Notion de qualité d'une image | 10 |
| II - Critères non pondérés bivariants | 11 |
| A - Mesure de distance | 11 |
| B - Critères issus du traitement du signal | 13 |
| C - Les critères de Linfoot généralisés | 13 |
| D - Conclusion | 14 |
| III - Critères pondérés bivariants | 14 |
| A - Modèles simple canal | 15 |
| B - Modèles multicanaux | 18 |
| C - Conclusion | 25 |
| IV - Critères univariants | 25 |
| V - Conclusion | 26 |
| Chapitre 2 : Une nouvelle approche | 27 |
| I - Introduction | 28 |
| II - Le modèle neuronal | 28 |
| A - Introduction | 28 |
| B - Principe de fonctionnement | 29 |
| III - Mise en œuvre | 34 |
| A - Etape fondamentale : caractérisation de la dégradation | 35 |
| B - Apprentissage – validation | 35 |
| C - Etude de sensibilité | 39 |
| IV - Conclusion | 39 |
| Chapitre 3 : Application à une image dégradée uniformément par un défaut unique | 41 |
| I - Evaluation de la qualité des images JPEG | 42 |
| A - Introduction | 42 |
| B - Dégradations introduites par le codage | 42 |
| C - Extraction du vecteur caractéristique | 43 |
| D - Résultats | 49 |
| E - Conclusion | 61 |
| II - Etude de la qualité des images floues – Application à l'évaluation du défaut de mise au point du satellite SPOT 1 | 62 |
| A - Introduction | 62 |
| B - Une première approche : utilisation d'un modèle de flou | 62 |
| C - Application : étude de la défocalisation du satellite SPOT 1 | 76 |
| Chapitre 4 : Application à une image dégradée par des défauts ponctuels | 85 |
| I - Introduction | 86 |
| II - Etude de la détection : tests visuels | 86 |
| A - Tests visuels | 86 |
| B - Recherche des paramètres significatifs | 87 |
| III - Adaptation du VDP pour la détection de singularités | 91 |

Table des matières

| | |
|--|------------|
| A - Principe | 91 |
| B - Résultats | 91 |
| C - Limitations | 94 |
| IV - Les réseaux de neurones | 94 |
| A - Le vecteur caractéristique | 95 |
| B - Résultats | 95 |
| C - Etude de sensibilité | 97 |
| V - Conclusion | 99 |
| Chapitre 5 : Application à une image dégradée par un mélange de deux défauts | 101 |
| I - Introduction | 102 |
| II - Cas où le système dégradant est connu | 103 |
| A - Une première approche | 103 |
| B - Réajustement des notes par un nouvel apprentissage : cas où le système dégradant est connu | 105 |
| C - Etude de sensibilité | 106 |
| III - Cas où le système dégradant n'est pas connu | 108 |
| A - Notation visuelle, modélisée par le VDP | 108 |
| B - Notation par calcul de distance à l'image parfaite | 109 |
| C - Etude de sensibilité | 110 |
| IV - Conclusion | 112 |
| Conclusion - Perspectives | 113 |
| Bibliographie | 115 |
| ANNEXE A - Base d'apprentissage générale | 121 |
| ANNEXE B - Le système visuel humain (SVH) | 129 |
| A1 - La perception visuelle | 129 |
| A2 - Anatomie du système visuel humain | 129 |
| A3 - Propriétés de la perception visuelle | 133 |
| ANNEXE C - Quelques statistiques de l'image et leur interprétation | 137 |
| A1 - Histogramme | 137 |
| A2 - Moments | 138 |
| A3 - Le variogramme | 138 |
| ANNEXE D - Le codage JPEG | 143 |
| A1 - Etape 1: Transformée discrète en cosinus (DCT) | 143 |
| A2 - Etape 2 : quantification | 143 |
| A3 - Etape 3 : codage de Huffman | 144 |
| ANNEXE E - Le système SPOT | 145 |
| A1 - Présentation de l'instrument | 145 |
| A2 - Modèle radiométrique | 151 |
| A3 - Prétraitement des images | 151 |
| A4 - Qualité des images | 152 |

Introduction

La notion de qualité d'une image est plus difficile à appréhender qu'il n'y paraît. Qu'entend-on par qualité d'une image ? Est-ce une qualité purement esthétique, répondant à des critères subjectifs très personnels, ou est-ce plutôt une qualité de service, permettant de définir si oui ou non l'image est adaptée à l'utilisation qui doit en être faite ? Une image dans laquelle se trouve toute l'information que l'on cherche a-t-elle besoin d'être parfaite par ailleurs ?

Avant d'essayer de déterminer comment évaluer la qualité d'une image, il faut d'abord savoir précisément ce que l'on veut en faire. Si l'objectif recherché est la vente au grand public, par exemple, le critère de qualité est purement esthétique. L'image doit être belle à regarder, sans défaut. Si, en revanche, elle est destinée à une application physique, mesure de niveaux radiométriques, analyse de texture, reconnaissance d'objets, le critère de qualité peut être défini comme la pertinence ou la fiabilité avec laquelle l'image peut être utilisée pour l'application à laquelle elle est destinée.

Pour la plupart des applications, l'homme est le dernier maillon de la chaîne. C'est lui qui décide si la qualité est recevable ou non. Les images sont donc traitées, ou plutôt évaluées, visuellement. Cette tâche peut rapidement devenir très fastidieuse lorsque le nombre d'images à traiter devient important. Les agents du service qualité de SPOTIMAGE, l'organisme qui diffuse les images du satellite SPOT, scrutent méticuleusement les images (qui ont une taille de 3000x3000 ou 6000x6000 pixels) pour rechercher et corriger d'éventuels défauts.

Le même problème se pose lors de la conception par simulation d'un instrument de prise de vue ou d'un nouvel algorithme de compression, par exemple. Des centaines, voire des milliers d'images sont générées, pour différentes configurations de l'outil de simulation. A chaque étape, la qualité de l'image est évaluée, le paramétrage de l'outil est corrigé et le processus est réitéré jusqu'à obtenir la qualité désirée. Il n'est pas pensable qu'un observateur en bout de chaîne analyse les images une à une.

Avec l'augmentation du volume de données à traiter, il est rapidement devenu nécessaire de fabriquer des outils pour évaluer de manière automatique la qualité des images. Sans entrer dans des considérations philosophiques hors sujet ici, il paraît évident que la possibilité pour un observateur de juger de la qualité esthétique d'une image, jugement d'ailleurs très personnel dès que l'on s'écarte un tant soit peu de la norme, a nécessité un apprentissage de plusieurs années. Comment intégrer dans un outil informatique une notion qui peut être aussi subjective que la notion de qualité ?

Ce problème a dans un premier temps été contourné grâce à la création d'outils, ou de critères d'évaluation de qualité que nous appellerons critères « bivariants ». Ces outils permettent de déterminer la qualité d'une image dégradée par comparaison avec une image de référence, généralement parfaite. C'est en quelque sorte le jeu des différences. Ils prennent cependant en compte les caractéristiques principales du système visuel humain qui est, comme on le sait, assez tolérant. Cette famille d'outils convient parfaitement pour certaines des applications citées précédemment : lorsqu'on cherche à évaluer la qualité d'un outil de compression ou lors de la simulation d'un instrument de prise de vue, une image parfaite est injectée en entrée et elle sert aussi de référence pour évaluer la qualité de l'image de sortie. Un algorithme itératif cherche à optimiser les paramètres du système tant que l'image de sortie, comparée à la référence, ne répond pas à certains critères de qualité. De tels outils existent et sont plus ou moins spécifiques et efficaces selon l'application visée.

Il reste néanmoins un très grand nombre de situations dans lesquelles on ne dispose pas d'image de référence. C'est par exemple le cas de la plupart des images satellitaires. Les outils bivariants n'étant pas adaptables à ces situations, il apparaît nécessaire de créer une nouvelle famille d'outils, « univariants », capables d'évaluer la qualité d'une image sans utiliser d'image de référence.

Actuellement, des recherches sont entreprises dans plusieurs laboratoires pour essayer de développer des modèles capables d'évaluer la qualité d'une image de manière univariante. Certains modélisent tous les niveaux de la vision, de la reconnaissance du contexte et des formes sur une image jusqu'à son interprétation en termes de qualité. Une telle modélisation, extrêmement performante si elle fonctionne, doit néanmoins être adaptée à l'application visée. Cette thèse propose une approche beaucoup plus pragmatique pour évaluer de manière univariante la qualité d'une image. Etant donné un contexte applicatif, nous utilisons un modèle paramétrique permettant de décider avec quelle pertinence une image donnée peut être utilisée. Ce modèle est préalablement étalonné, soit sur le jugement d'observateurs qualifiés, soit sur un autre type d'évaluation de qualité, bivariant par exemple. Nous avons de plus simplifié le problème en considérant que, dans un cadre applicatif donné, le système dégradant est souvent connu. On sait donc quels types de dégradations rechercher sur l'image. Notre approche nécessite néanmoins d'étalonner un modèle différent pour chaque type de dégradation, ou pour chaque type d'application.

Etant donné la complexité du problème, nous avons utilisé un réseau de neurones comme modèle. Il a l'avantage d'être constitué d'un très grand nombre de paramètres, donc de permettre la modélisation de phénomènes très complexes, tout en proposant une méthode d'étalonnage des paramètres très simple à mettre en œuvre. La difficulté d'une telle approche consiste maintenant à réduire la quantité d'informations contenue dans l'image de manière à pouvoir les exploiter avec le réseau de neurones.

L'évaluation de la qualité nécessite une phase préliminaire d'apprentissage destinée à calibrer le réseau de neurones. Cette phase est réalisée sur un groupe d'images dont la qualité a préalablement été déterminée, soit par des observateurs qualifiés, soit par le modèle bivariant dont on désire reproduire le comportement. La qualité peut aussi être donnée par le paramétrage du système dégradant : taux de compression, défaut de mise au point...

Dans un premier temps, il faut extraire des images une information physique ou mathématique, sous la forme de quelques paramètres susceptibles de décrire la dégradation. C'est l'étape la plus délicate mais aussi la plus importante. Dans un deuxième temps, le réseau de neurones apprend à faire la correspondance entre la modélisation mathématique et la note de qualité (ou les paramètres du système dégradant). Finalement, après cette phase d'apprentissage, le réseau de neurones peut être utilisé pour déterminer la note de qualité d'une image quelconque, à partir de la caractérisation mathématique qui en est extraite : il procède à une interpolation sur ce qu'il a appris pendant la phase précédente.

L'avantage d'une telle méthode est qu'elle permet, entre autre, de reproduire un comportement visuel sans modèle de vision explicite.

Nous l'avons mise en œuvre pour plusieurs types de dégradations. Pour simplifier le problème, nous avons dans un premier temps considéré des dégradations simples, non mélangées les unes avec les autres : flou, artefacts de compression, dégradations locales... Pour certaines de ces dégradations, nous avons modélisé le comportement visuel dans l'estimation de la qualité, pour d'autres nous avons montré qu'il était possible de caractériser le système dégradant.

Ce travail a été essentiellement réalisé dans le contexte proposé par les images du satellite SPOT. Ces images, brutes ou après corrections radiométriques, comportent des défauts plus ou moins flagrants qui nécessitent d'être détectés et corrigés avant toute commercialisation. Ces défauts peuvent être très localisés, voire ponctuels, et gênants ou non selon la zone de l'image dans laquelle ils se trouvent. Il n'est pas utile de corriger un défaut non gênant... Nous montrerons que notre méthode permet de différencier un défaut ponctuel gênant d'un défaut ponctuel non gênant. D'autres défauts peuvent être uniformes sur l'image. Le satellite SPOT1, lancé en 1986, avait un défaut de mise au point imperceptible à l'œil. Nous montrerons que notre méthode permet, grâce à l'étude des images floues, de déterminer quantitativement ce défaut de mise au point.

Indépendamment du satellite SPOT, nous étudierons aussi les artefacts de compression JPEG, défaut structuré, pour essayer soit de donner une note de qualité visuelle à l'image comprimée, soit de déterminer le paramètre réglable lors de la compression et dont toute trace disparaît sur l'image dégradée.

Le premier chapitre de ce manuscrit est consacré à un état de l'art des différentes méthodes d'évaluation de la qualité des images. Ces méthodes, quasiment toutes bivariantes, prennent pour la plupart en compte les principales caractéristiques visuelles.

Le deuxième chapitre présente le principe général de la méthode univariante que nous avons développée ainsi que le modèle utilisé, un réseau de neurones. Ce chapitre est fondamental pour une bonne compréhension des chapitres suivants.

Le chapitre trois présente une application de la méthode à des dégradations uniformément réparties sur l'image. Pour la dégradation structurée occasionnée par la compression JPEG, nous montrons qu'il est possible de donner une évaluation de la qualité visuelle de l'image, mais aussi de déterminer le facteur de qualité de la compression. Pour la dégradation provoquée par un défaut de mise au point du système imageur, nous montrons qu'il est possible de quantifier ce défaut, en terme d'erreur de mise au point, par exemple.

Le chapitre quatre présente une application de la méthode pour simuler la détection visuelle de défauts ponctuels dans une image. Nous montrons qu'elle permet de respecter les principales caractéristiques visuelles pour réaliser une détection ni meilleure ni pire que celle de l'œil.

Finalement, le chapitre cinq étudie le cas d'images dégradées par un mélange de deux défauts. Nous montrons qu'il est parfois possible de retrouver les paramètres du système dégradant pour chacun de ces défauts, mais aussi d'évaluer la qualité visuelle globale de l'image, par exemple.

Chapitre 1 :

Etat de l'art des critères de qualité image

I - Introduction – Notion de qualité d'une image

Les années 1980 et 1990 ont connu un essor très important des techniques d'information et de communication numériques grand public : ordinateurs personnels, internet, TV numériques... Dans des domaines plus pointus, comme l'observation de la Terre depuis l'espace (SPOT, Hélios), l'observation de l'espace (Hubble), l'image numérique est devenu le support incontournable de l'information. Ses avantages sont nombreux : non seulement l'information peut être stockée et reproduite sans dégradation, mais elle peut être comprimée pour une transmission plus rapide, elle peut être améliorée par de simples traitements informatiques (déconvolution, débruitage) ou tout simplement présentée sur internet, par exemple, à des millions de personnes spécialistes ou non.

Toutes ces images ne sont pas forcément de bonne qualité, un observateur peut s'en rendre compte parfois très facilement. Sur internet, par exemple, beaucoup de sites qui utilisent un très grand nombre d'images, les compriment systématiquement et assez fortement pour gagner de la place, au détriment de la qualité. La quantité importante d'images rend impossible une vérification visuelle. D'une manière générale, la qualité d'une image dépend de l'utilisation que l'on doit en faire : des images militaires de reconnaissance aérienne peuvent paraître très intéressantes mais si l'on ne reconnaît pas les objets (chars, bases) que l'on y cherche, elle sera de mauvaise qualité pour cette mission. Inversement, une image d'apparence médiocre peut parfaitement convenir à l'application pour laquelle elle est destinée.

La notion de qualité d'une image apparaît donc étroitement liée à l'usage qui doit être fait de l'image : visualisation grand public, détection d'objets, reconnaissance de surfaces... Les personnes concernées ont, pour chacune de leur mission, tenté de normaliser l'étude de qualité des images. Le CCITT (Comité Consultatif International Télégraphique et Téléphonique) propose une méthode d'évaluation subjective de la qualité des images de télévision [Cci90]. Les militaires ont aussi proposé différents protocoles d'observation des images [Bes93, DCK97]. Cependant, le très grand nombre d'images à traiter rend très contraignante une vérification visuelle et systématique de qualité et il est devenu indispensable de disposer d'outils permettant d'évaluer de façon automatique la qualité d'une image pour une mission donnée.

Des mesures objectives, c'est-à-dire des algorithmes mathématiques, pour évaluer la qualité des images ont peu à peu été développées. Ces critères de qualité sont de deux types (Figure 1) :

- les critères *non pondérés* fournissent une estimation de la qualité de l'images d'après ses propriétés statistiques, spatiales fréquentielles, sans tenir compte des propriétés du système visuel humain (SVH). Ces critères sont décrits dans la partie II - de ce chapitre ;
- les critères *pondérés* utilisent en plus une modélisation du SVH, que se soit une simple fonction de transfert (modèle simple canal), ou une combinaison de fonctions de transferts (modèle multicanal). Ces critères sont décrits dans la partie III - de ce chapitre.

La plupart de ces modèles de qualité sont des modèles *bivariants*. Leur particularité est d'estimer la qualité d'une image imparfaite par comparaison à l'image parfaite servant de référence. C'est en fait un calcul de distance mathématique amélioré. Ces modèles servent, par exemple, à concevoir ou optimiser un système de prise vue (Figure 2) [Oli98].

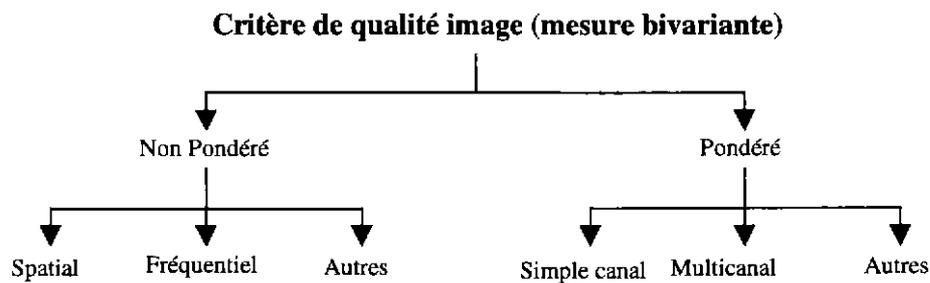


Figure 1 : Classification des critères de qualité image.

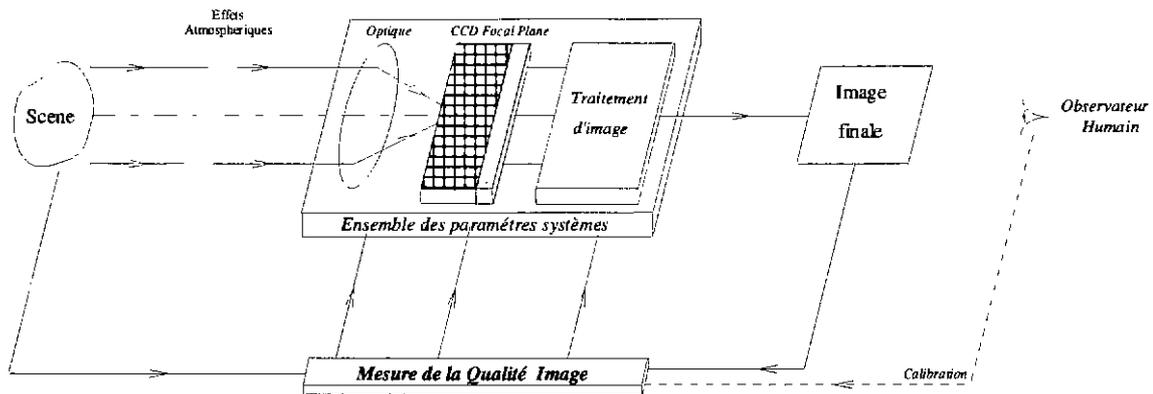


Figure 2 : Méthodologie de conception d'un système imageur.

Cependant dans beaucoup de situations, on peut avoir besoin d'estimer la qualité d'une image pour laquelle on ne dispose d'aucune référence. Peu de critères *univariants* existent actuellement, et leur fiabilité est souvent assez faible.

La suite de ce chapitre présente rapidement les différents critères objectifs de qualité des images utilisés actuellement.

II - Critères non pondérés bivariants

Ces critères sont essentiellement des mesures de distance mathématique, classiques en traitement d'images [RoK76, GoW77, Pra78, Jai89, Kun93].

A - Mesure de distance

La qualité d'une image est déterminée par la distance entre l'image dégradée et l'image parfaite. Pour une image dégradée \tilde{F} , de taille $N \times M$ pixels, dont on veut évaluer la qualité par rapport à la référence F , on peut définir plusieurs distances.

1) Famille des distances euclidiennes

Elle est définie par :

$$d\{F, \tilde{F}; p\} = \left\{ \sum_{j=1}^M \sum_{k=1}^N |F(j, k) - \tilde{F}(j, k)|^p \right\}^{1/p} \quad \text{pour } p \geq 1$$

Il est évident que la distance est minimale et égale à 0 si $\forall (i, j) F(i, j) = \tilde{F}(i, j)$; c'est-à-dire \tilde{F} n'est pas dégradée.

Remarquons que quand $p \rightarrow \infty$

$$\lim_{p \rightarrow \infty} (d\{F, \tilde{F}; p\}) = \max_{j,k} |F(j,k) - \tilde{F}(j,k)|$$

qui est l'erreur maximale de reconstruction.

Si $p_1 > p_2 \geq 1$ alors :

$$d\{F, \tilde{F}; p_1\} \leq d\{F, \tilde{F}; p_2\}$$

Cette distance est une fonction décroissante de p .

Cela signifie que quand p devient grand, on néglige les petites différences entre les pixels de \tilde{F} et ceux de F pour ne conserver que les plus importantes, le cas extrême étant $p = \infty$, pour lequel seul l'écart maximal intervient.

Un des problèmes posés par cette distance est qu'elle dépend du nombre de pixels de l'image (sauf pour $p = \infty$). On ne peut donc pas comparer ses valeurs pour deux images de tailles différentes mais dégradées de la même manière. On utilisera plutôt une distance normalisée par rapport au nombre de pixels.

2) Famille des normes L_p

$$L_p\{F, \tilde{F}\} = \left\{ \frac{1}{NM} \sum_{j=1}^M \sum_{k=1}^N |F(j,k) - \tilde{F}(j,k)|^p \right\}^{1/p} \quad \text{pour } p \geq 1$$

Si $p_1 > p_2 \geq 1$ alors :

$$L_{p_1}\{F, \tilde{F}\} \geq L_{p_2}\{F, \tilde{F}\}$$

Cette distance est une fonction croissante de p .

En particulier :

$$L_1\{F, \tilde{F}\} = \frac{1}{MN} \sum_{j=1}^M \sum_{k=1}^N |F(j,k) - \tilde{F}(j,k)| \leq \lim_{p \rightarrow \infty} (L_p\{F, \tilde{F}\}) = \max_{j,k} |F(j,k) - \tilde{F}(j,k)|$$

De même que précédemment, l'augmentation du facteur p implique une plus forte influence des différences les plus importantes.

Pour les premières valeurs de p on obtient des distances très utilisées en traitement d'image :

$$p = 1 \quad L_1\{F, \tilde{F}\} = \frac{1}{MN} \sum_{j=1}^M \sum_{k=1}^N |F(j,k) - \tilde{F}(j,k)| = AD \quad (\text{Average Difference})$$

$$p = 2 \quad L_2\{F, \tilde{F}\} = \left\{ \frac{1}{MN} \sum_{j=1}^M \sum_{k=1}^N |F(j,k) - \tilde{F}(j,k)|^2 \right\}^{1/2} = RMSE \quad (\text{Root Mean Squared Error})$$

On utilise aussi L_2^2 qui est l'erreur quadratique moyenne ou MSE (Mean Squared Error). Ces deux mesures ont les mêmes propriétés pour leurs minima et maxima, cependant la MSE est plus sensible aux grandes différences que la RMSE.

La MSE peut être présentée sous l'une des deux formes normalisées :

$$PMSE = \left\{ \frac{1}{MN} \sum_{j=1}^M \sum_{k=1}^N |F(j,k) - \tilde{F}(j,k)|^2 \right\} / \left\{ \max_{j,k} |F(j,k)| \right\}^2 \quad (\text{Peak Mean Squared Error})$$

$$NMSE = \left\{ \sum_{j=1}^M \sum_{k=1}^N |F(j,k) - \tilde{F}(j,k)|^2 \right\} / \left\{ \sum_{j=1}^M \sum_{k=1}^N (F(j,k))^2 \right\} \quad (\text{Normalized Mean Squared Error})$$

A l'aide de ces normalisations, les valeurs des mesures de distance sont moins dépendantes de l'image de référence. Le même type de normalisation peut être réalisé pour les autres valeurs de p .

B - Critères issus du traitement du signal

Si l'on considère que l'image originale $F(j,k)$ est le signal d'entrée d'un système physique et l'image dégradée $\tilde{F}(j,k)$ est le signal de sortie alors l'image d'erreur peut être considérée comme un bruit $B(j,k)$:

$$\tilde{F}(j,k) = F(j,k) + B(j,k)$$

Dès lors, il est possible d'exprimer le rapport signal à bruit ("Signal to Noise Ratio") d'un tel système :

$$SNR = 10 \text{Log}_{10} \frac{\sum_{j=1}^M \sum_{k=1}^N [\tilde{F}(j,k)]^2}{\sum_{j=1}^M \sum_{k=1}^N [B(j,k)]^2} \quad (\text{dB})$$

Une autre manière d'exprimer ce rapport est d'utiliser une mesure de distance du type MSE :

$$SNR = 10 \text{Log}_{10} \frac{1}{MN} \frac{\sum_{j=1}^M \sum_{k=1}^N [\tilde{F}(j,k)]^2}{MSE} \quad (\text{dB})$$

On peut aussi utiliser le "Peak Signal Noise Ratio" ou PSNR :

$$PSNR = 10 \text{Log}_{10} \frac{[\text{Max}(F(j,k))]^2}{MSE} \quad (\text{dB})$$

C - Les critères de Linfoot généralisés

E.H. Linfoot [Lin56, Lin58] a développé une théorie sur la détermination de la qualité d'un système optique à partir d'une modélisation du système par des fonctions de transfert et des propriétés statistiques de l'image originale. Cette théorie mesure en quelque sorte la ressemblance entre l'image originale et l'image dégradée.

Linfoot définit un coefficient de fidélité, un contenu structurel et une corrélation pour évaluer la ressemblance entre l'image à l'entrée du système et l'image de sortie.

Eskicioglu et Fisher [EsF93] appliquent les critères de Linfoot dans le domaine spatial échantillonné et définissent :

- la fidélité spatiale :

$$F = 1 - \frac{\sum_{j=1}^M \sum_{k=1}^N [F(j,k) - \tilde{F}(j,k)]^2}{\sum_{j=1}^M \sum_{k=1}^N [F(j,k)]^2}$$

- le Structural Content (SC) est défini par :

$$SC = \frac{\sum_{j=1}^M \sum_{k=1}^N [\tilde{F}(j,k)]^2}{\sum_{j=1}^M \sum_{k=1}^N [F(j,k)]^2}$$

- le Normalized Cross-Correlation (NCC) :

$$NCC = \frac{\sum_{j=1}^M \sum_{k=1}^N F(j,k)\tilde{F}(j,k)}{\sum_{j=1}^M \sum_{k=1}^N [F(j,k)]^2}$$

Ces critères restent cependant plus difficiles à appréhender physiquement que les précédents. Notons cependant qu'une image non dégradée a une fidélité de 1 et un contenu structural (SC) de 1.

D - Conclusion

La liste des critères non pondérés présentée ci-dessus n'est pas exhaustive. Certains critères graphiques, comme par exemple un histogramme de l'image d'erreur, peuvent aussi être utilisés pour déterminer la qualité. Cependant, quel que soit le critère non pondéré envisagé, l'évaluation de la qualité reste mal corrélée à celle de l'observateur. La raison est que les critères mathématiques comparent les images parfaites et dégradées pixel à pixel, alors que le système visuel moyenne les informations sur des ensembles de pixels. A titre d'illustration, la Figure 3 représente deux images identiques, l'une compressée avec un algorithme JPEG et l'autre par un algorithme utilisant la transformée en ondelettes. Bien que ces deux images présentent des dégradations très différentes l'une de l'autre, elles ont la même valeur de PSNR.



Image compressée Ondelette : PSNR = 28.66



Image compressée JPEG : PSNR = 28.65

Figure 3 : Comparaison d'images avec le même PSNR.

Les critères mathématiques spatiaux ne constituant pas une mesure significative de la qualité visuelle des images [Mar86], plusieurs critères pondérés ont été développés. Ces critères utilisent généralement une modélisation du système visuel humain.

III - Critères pondérés bivariants

Modéliser la vision humaine nécessite non seulement d'avoir une connaissance physiologique du SVH (cf. ANNEXE B -), mais aussi d'avoir les outils mathématiques pour le décrire et l'informatique pour l'implémenter. Le premier modèle du SVH a été proposé par Shade en 1956 [Sha56]. C'était un simulateur photoélectrique reproduisant la sensibilité du SVH à des cibles. Par la suite, d'autres modèles simples ont représenté le SVH comme un système linéaire caractérisé par une fonction de transfert (modèle simple canal). Ces modèles se

révélant incapables de prédire certains comportements complexes de la vision, des modèles plus complexes (multi canaux), ont été élaborés en juxtaposant plusieurs modèles simple canal.

A - Modèles simple canal

1) Le modèle de O. Shade [Sha56]

Ce modèle suppose que l'image neuronale (la représentation corticale de l'image) peut être obtenue par une translation et une transformation linéaire de l'image rétinale. Il en découle que cette transformation peut être déterminée par la réponse d'un seul neurone (c'est une sorte de réponse impulsionnelle).

Si $x[i]$ est un stimulus de N échantillons à 1 dimension, si $y[i]$ la représentation neuronale de $x[i]$ et $h[i]$ la réponse impulsionnelle d'un champ réceptif :

$$y[i] = \sum_{l=0}^{N-1} h[i-l]x[l]$$

A l'aide d'observateurs, Shade a mesuré la sensibilité du SVH à des mires sinusoïdales à contraste variable. Il a déterminé h , réponse impulsionnelle psychophysique.

Le modèle de Shade permet de prédire correctement la visibilité de mires simples (horizontales ou verticales) mais il s'avère inutilisable pour des mires mixtes (somme de sinusoïdes de différentes périodes) et des mires de très basse fréquence [Wan95].

2) Le modèle isotropique de J.L. Mannos et D.J. Sakrison [MaS74]

Ce modèle repose sur les travaux psychophysiques de Campbell et Robson [CaR68].

A partir de l'observation de mires horizontales sinusoïdales du type :

$$S(x) = L_0(1 + m \cos(2\pi f x))$$

Mannos et Sakrison proposent une modélisation de la fonction de sensibilité visuelle au contraste (CSF) :

$$CSF(f) \approx \frac{1}{m_s} = c(f_r) \cdot e^{-f_r}$$

où f_r est la fréquence radiale normalisée, $c(f_r)$ est une fonction non linéaire correspondant au processus d'inhibition latérale engendré par l'interaction neuronale au niveau de la rétine et des couches supérieures. La fonction exponentielle caractérise la résolution limitée de l'oeil, due à la taille non nulle et à l'espacement des photorécepteurs (cônes et bâtonnets).

A partir des hypothèses de Stockham [Sto72], le SVH est modélisé par une première transformation non linéaire suivi d'une transformation linéaire (Figure 4).

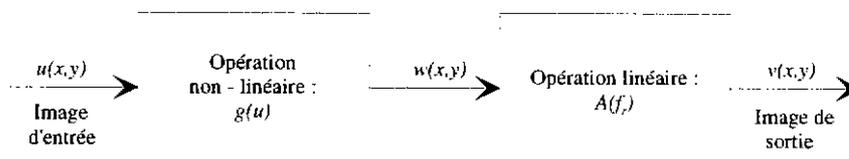


Figure 4 : modèle de Mannos et Sakrison.

Afin de simplifier la modélisation du SVH, Mannos et Sakrison proposent un modèle isotropique :

$$A(f_x, f_y) = A(f_r) \quad \text{Avec} \quad f_r = \sqrt{f_x^2 + f_y^2}$$

$$A(f_r) = \left[c + (f_r / f_0)^{k_1} \right] \exp \left[- (f_r / f_0)^{k_2} \right]$$

Après étalonnage des paramètres sur des observateurs, le modèle devient :

$$A(f_r) = 2.6 \cdot [0,0192 + 0,114 f_r] \exp[-(0,114 f_r)^{1,1}]$$

$$g(u) = u^{0,33}$$

où f_r est la fréquence radiale en cycles par degré.

La mesure de qualité image de Mannos et Sakrison fut la première à utiliser un modèle de vision pour l'évaluation des techniques de traitement d'images. Ce modèle prédit néanmoins une CSF maximale pour une fréquence spatiale de 8 cycles/deg (cf. Figure 7) alors que des expériences plus récentes ont montré (cf. Chapitre 1 :III -A -5)) que le maximum de sensibilité était atteint pour une fréquence spatiale comprise entre 4 et 5 cycles/deg [NLS89,ChR90].

3) Le modèle de C.F. Hall et E.L. Hall [HaH77]

Hall et Hall proposent un modèle tenant compte de la non linéarité ainsi que de la réponse fréquentielle du SVH. A la différence du modèle de Mannos et Sakrison, le filtre passe-bande est remplacé par un filtre passe-bas suivi par un filtre passe-haut, la non linéarité étant placée entre les deux filtres (Figure 5). Le filtre passe-bas modélise la formation de l'image sur la rétine, la non linéarité modélise la sensibilité des cellules de la rétine et le filtre passe-haut modélise la formation de l'image neuronale. C'est aussi un modèle isotropique.

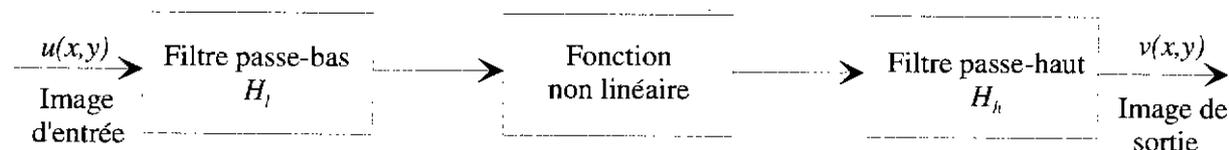


Figure 5 : modèle de Hall et Hall.

Le filtre passe-bas correspond à la réponse impulsionnelle optique de l'œil, modélisée par :

$$H_l(f_r) = \frac{2\alpha}{\alpha^2 + f_r^2}$$

Avec f_r fréquence spatiale radiale en cycles par degré et α une constante.

Le filtre passe-haut est un modèle du mécanisme d'inhibition latérale des photorécepteurs :

$$H_h(f_r) = \frac{a^2 + f_r^2}{2a_0 a + (1 - a_0)(a^2 + f_r^2)}$$

Le modèle est calibré à l'aide de tests visuels, et les constantes fixées : $\alpha=0,7$ pour un diamètre de pupille égal à 3 mm, $a=0,01$ et $a_0=0,2$. La fonction non linéaire est une fonction logarithmique.

Plus tard, Hall établit une mesure de la qualité image en pondérant une RMSE par son modèle visuel [Hal81].

4) Le modèle de J.O. Limb [Lim79]

Limb recherche une mesure objective de la qualité image qui s'approche le plus possible du jugement de l'observateur. Pour trouver cette mesure, il fait évaluer, par un ensemble d'observateurs, cinq types d'images ayant subi 16 dégradations (codage DPCM, bruits, filtrage, ...) de différentes intensités. Les notes subjectives sont comparées avec un ensemble de mesures objectives basées sur la norme L_p (cf. Chapitre 1 :II -A -2)).

Limb trouve que la norme L_p est une bonne mesure pour estimer les dégradations de bruit blanc sur des images ayant de nombreuses zones uniformes. Il modifie donc cette norme pour prendre en compte le phénomène de masquage du SVH :

$$LM_p = \left\{ \frac{1}{m} \sum_{i=1}^m \left| \frac{x_i - \bar{x}}{W_i} \right|^p \right\}^{1/p}$$

m est le nombre de pixels dans l'image et W_i une fonction de pondération pour le pixel i . Cette fonction est calculée à partir d'une fonction d'activité qui est une mesure de la variation de l'intensité au voisinage du pixel i . Plusieurs fonctions d'activité sont définies, par exemple :

$$A_{av} = \sum_{i=1}^9 |x_i - \bar{x}| \quad \text{et} \quad W_i = \exp(k A_{av})$$

Cette mesure semble plus robuste que la norme L_p pour différents types d'images et de dégradations.

Finalement, Limb propose un modèle complet de la vision humaine en incluant un filtrage de l'erreur par un filtre passe-bas et une pondération par une fonction de masquage (Figure 6).

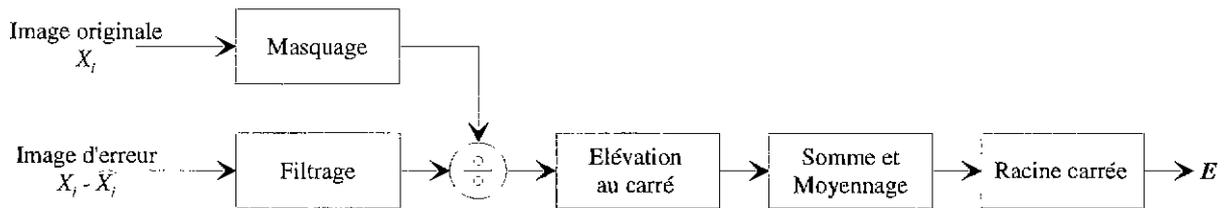


Figure 6 : modèle de Limb.

Limb a été l'un des premiers à prendre en compte l'effet de masquage dans une mesure de la qualité image et à étudier la corrélation entre des mesures objectives et subjectives pour différentes dégradations.

5) Autres modèles

A partir des travaux de Mannos et Sakrison [MaS74] ainsi que ceux de DePalma et Lowry [DeL62], Nill [Nil85] propose un nouveau modèle de CSF :

$$H(f_r) = (0.2 + 0.45 f_r) e^{-0.18 f_r}$$

Ngan [NLS89] et Chitprasert [Chr90] corrigent à nouveau ce modèle quelques années plus tard.

La Figure 7 compare les différents modèles de CSF. Celui de Mannos et Sakrison, le plus ancien d'entre eux, est maintenant généralement abandonné car il prédit un maximum de la CSF pour une fréquence spatiale de 8 cycles/deg, qui ne correspond pas à la valeur observée, proche de 4 ou 5 cycles/deg.

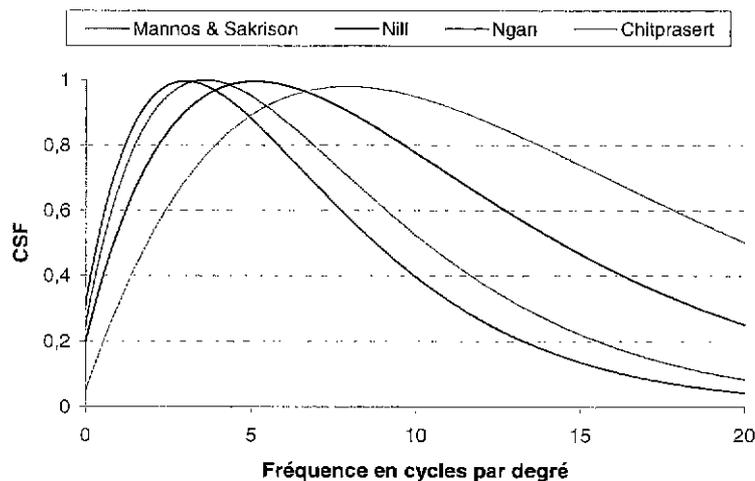


Figure 7 : Comparaison des différents modèles de CSF en fonction de la fréquence spatiale.

De plus, Nill [NiB92] propose une mesure de la qualité image basée sur une pondération de la densité spectrale de puissance par son modèle de CSF et un filtre de Wiener.

Saghri [SaC89] propose une mesure de la qualité image basée sur le modèle de CSF de Nill et une meilleure prise en compte des paramètres d'observation (étalonnage de l'écran, distance de visualisation et résolution de l'image).

B - Modèles multicanaux

Les modèles simple canal ne permettent pas de prédire la perception des formes complexes. De plus, des expériences psychophysiques [CaR68, SNR71] mettent en évidence la structure multirésolution du SVH. Pour cette raison des modèles multicanaux ont été développés.

Ces modèles sont généralement constitués de plusieurs étages :

- premier étage : filtre passe-bande (modèle de CSF) ;
- deuxième étage : découpage fréquentiel avec x canaux pour les fréquences spatiales et y canaux pour les orientations ;
- troisième étage : non linéarité (loi de Weber) ;
- quatrième étage : mesure finale, réalisée par la différence entre l'image originale et l'image dégradée après passage dans les étages précédents.

Ces modèles sont généralement assez lourds. Pour cette raison les deux seuls d'entre eux qui seront utilisés comme références au cours de notre élaboration de critères de qualité univariants sont développés dans la suite : le modèle de Karunasekera et Kingsburry et celui de Daly. Le premier est un modèle bivariant d'évaluation de la qualité visuelle des images JPEG, le second est un modèle qui donne une carte des différences visibles entre l'image dégradée et la référence parfaite.

Cependant, d'autres modèles multicanaux peuvent être trouvés dans [SaK77, Wat87b, ZeH89, CoM90, Lub93].

Une approche pondérée quelque peu différente (combinaison linéaire de critères simple canal) peut être trouvée dans [Miy88, AKM92, MKA96].

1) Le VDP ("Visible Difference Predictor") de S. Daly, [Dal92, Dal93, Dal94]

Le VDP constitue l'un des modèles les plus connus de la vision humaine. Il a été utilisé pour mesurer la qualité des méthodes de codages [AvA96, AAE97] et des images médicales [Mat96]. Il permet de mesurer la qualité d'une image dégradée par comparaison à l'image originale à l'aide d'une carte indiquant la probabilité de détection d'une différence visuelle entre les deux images. Son schéma de principe est donné Figure 8.

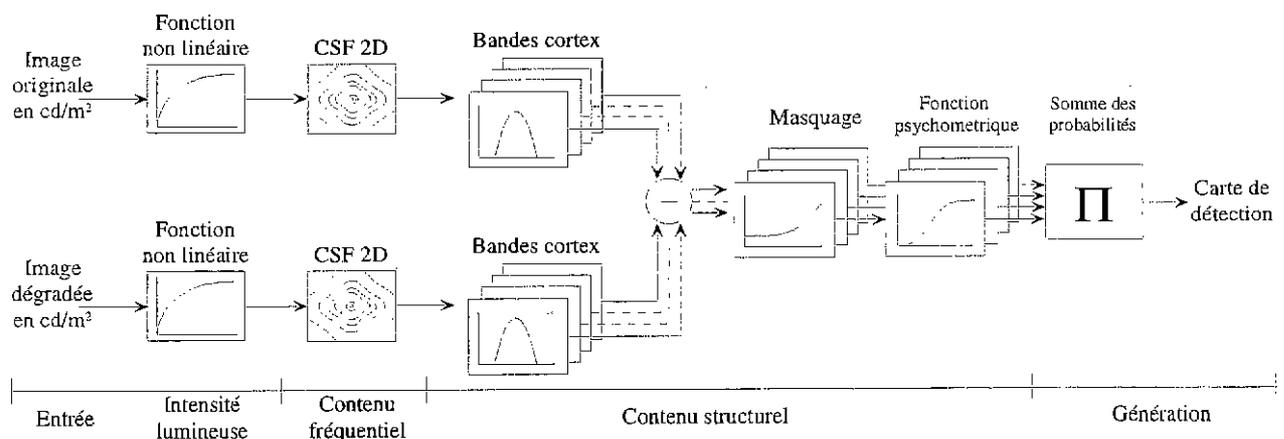


Figure 8 : le "Visible Difference Predictor" de S. Daly.

a - Entrées du modèle

Le modèle nécessite :

- les images originale et dégradée
- la distance de visualisation (pour transformer les fréquences spatiales en cycles par degré)
- les caractéristiques de l'écran (pour convertir les niveaux de gris des images en luminance cd/m^2)

b - Sensibilité à l'intensité lumineuse

La fonction non linéaire utilisée par Daly est définie pour chaque pixel par :

$$\frac{R(i, j)}{R_{max}} = \frac{L(i, j)}{L(i, j) + c L(i, j)^b}$$

Où $R(i, j)/R_{max}$ est la réponse normalisée de la rétine, $L(i, j)$ la luminance arrivant sur la rétine, $b=0,63$ et $c=12,6$ (constantes déterminées d'après ses expériences psychovisuelles).

c - Sensibilité fréquentielle (CSF)

La formulation de la CSF utilisée par Daly est donnée par le modèle de Barten [Bar90]:

$$CSF(u, v) = P \cdot \text{Min} \left\{ CSF_1 \left(\frac{\rho}{r_a \cdot r_b}, L_A, I^2 \right), CSF_1(\rho, L_A, I^2) \right\}$$

$$CSF_1(\rho, L_A, I^2) = \left(\left(3,23(\rho^2 I^2)^{-0,3} \right)^5 + 1 \right)^{-1/5} \cdot 0,9 \rho A e^{-0,9 \rho B}$$

Avec A et B fonctions de L_A (niveau de l'intensité lumineuse), $\rho = \frac{\sqrt{u^2 + v^2}}{I}$ fréquence radiale

(cycles/deg), $\theta = \text{Arctan} \frac{v}{u}$ l'orientation en degré, $I = \text{arctan} \frac{D_w}{D_v}$ taille apparente de l'image en degré visuel (D_w et D_v représentent la largeur et la hauteur de l'image, exprimées dans la même unité). A partir d'expériences psychophysiques, Daly donne : $r_a = 0,856 D_v^{0,14}$, $r_b = 0,11 \cos(4\theta) + 0,89$ et une valeur maximale de la CSF égale à 250.

d - Sensibilité au contenu structurel de l'image

Cet étage intervient après la CSF.

Dans la détection des formes, le SVH opère une sélection qui est fonction de la fréquence du stimulus, de son orientation et du phénomène de masquage.

La sélectivité du SVH est modélisée par une *transformée en cortex* [Wat87a] : l'espace fréquentiel radial est décomposé en cinq parties (1/2, 1/4, 1/8, 1/16 et 1/32) et en six angles d'orientations (0, 30, 60, 90, 120 et 150 degrés). Ces choix ont été faits après une série d'expériences psychophysiques [PhW84].

En incluant la bande de base, 31 filtres sont utilisés (Figure 9).

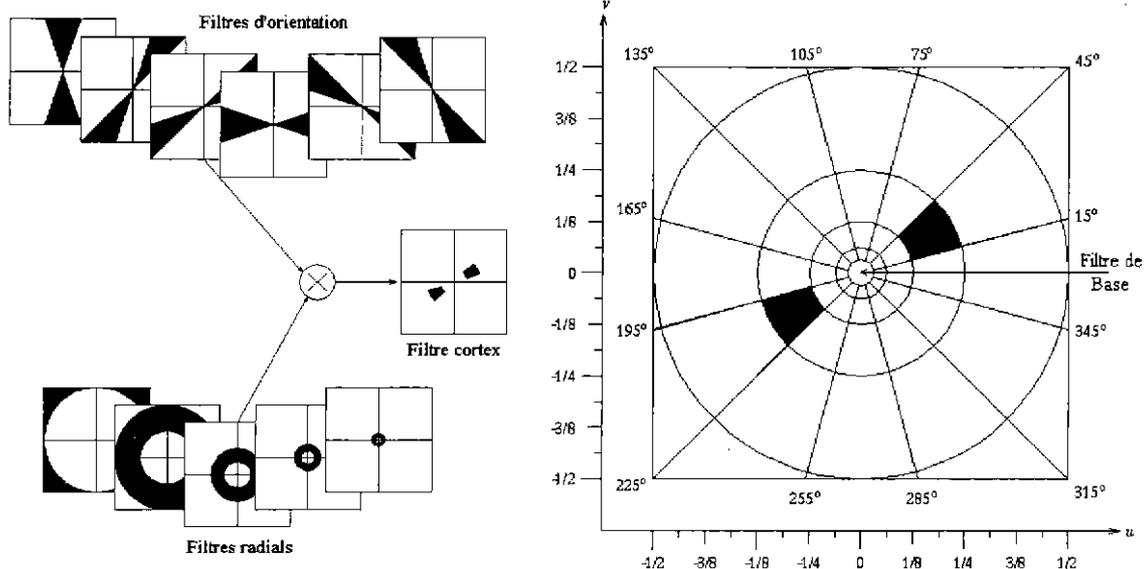


Figure 9 : construction des filtres cortex et exemple de filtre ($f = 3/16$ et $\theta = 30$ degrés).

Ces filtres étant utilisés en parallèle, on obtient 31 sous-images cortex directionnelles de l'image dégradée et autant de la référence.

A ce stade, les effets de masquage du SVH doivent être pris en compte.

Un objet (ou *cible*) sur une image peut être masqué par son voisinage (ou *fond*) si celui-ci n'est pas uniforme. Ce masquage est d'autant plus important que le contraste du voisinage est grand, la relation étant non linéaire (Figure 10).

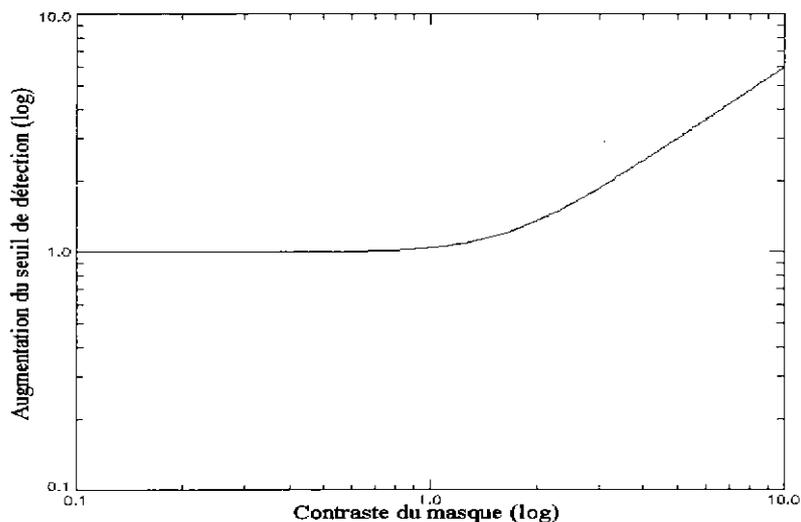


Figure 10 : effet de masquage, modèle de Legge et Foley [LeF80]

Le but étant de détecter les différences visibles entre l'image dégradée et la référence, seront considérées comme cibles les sous-images cortex d'erreurs, c'est-à-dire résultats de la différence entre les sous-images cortex dégradées et les sous-images cortex de référence.

Le masque normalisé en contraste m correspondant à l'image cortex, non dégradée, dans la bande k, l (orientation, fréquence) et ne tenant pas compte de la non linéarité du phénomène s'écrit :

$$m^{k,l}[i, j] = TF^{-1} \{ TF \{ I[i, j] \} \cdot CSF[u, v] \cdot cortex^{k,l}[u, v] \}$$

où $I(i, j)$ est l'image originale, TF : Transformée de Fourier.

Le masque normalisé en contraste prenant en compte la non linéarité (élévation du seuil) est :

$$T_e^{k,l}[i, j] = \left(1 + \left(k_1 (k_2 |m^{k,l}[i, j]|^s)^b \right)^{1/b} \right)$$

où k_1 , k_2 , s et b sont des constantes du modèle [Dal92].

Finalement, les cibles étant les différences entre les images originales et dégradées, le masque à appliquer est :

$$T_{em}^{k,l}[i, j] = \text{Min}\{T_{e1}^{k,l}[i, j], T_{e2}^{k,l}[i, j]\}$$

où les indices 1 et 2 correspondent aux images originale et dégradée.

La probabilité de détection d'une distorsion est donnée par [Nac81, RoG81]. Cette fonction indique la probabilité de détection d'une cible de contraste normalisé c en présence d'un masque de contraste normalisé α :

$$P(c) = 1 - e^{-(c/\alpha)^\beta}$$

β est un paramètre déterminé expérimentalement.

Finalement, dans le cas du VDP, la probabilité de détection du pixel (i, j) sur la sous-image cortex (k, l) est donnée par :

$$P_{k,l}[i, j] = 1 - \exp\left\{-\left(\Delta C_{k,l}[i, j]/T_{em}^{k,l}[i, j]\right)^\beta\right\} \quad \text{avec} \quad \Delta C_{k,l}[i, j] = \frac{B1_{k,l}[i, j]}{\bar{B}_K} - \frac{B2_{k,l}[i, j]}{\bar{B}_K}$$

où $B1_{k,l}$ et $B2_{k,l}$ les images cortex dans la bande (k, l) , et \bar{B}_K est la moyenne de l'image cortex dans la bande de base.

e - Génération de la carte des distorsions visibles

Après avoir généré les 31 cartes de probabilité de détection $P_{k,l}$ dans chacune des bandes cortex (k, l) , la carte de probabilité totale est donnée par :

$$P_t[i, j] = 1 - \prod_{k,l} (1 - P_{k,l}[i, j])$$

Cette carte de probabilité est améliorée en incluant le signe de la dégradation (positif pour une dégradation plus brillante, négatif pour une dégradation moins brillante).

Le seuil de détection visuelle est donné par : $|P| > 0,5$.

f - Quelques exemples

La Figure 11 présente les images dégradées, les cartes de détection des erreurs visibles et la superpositions de images dégradées et des cartes de détection.

Le bruit est plus gênant sur les zones uniformes de l'image, alors que le flou dégrade avant tout les zones structurées. La compression JPEG dégrade la texture des zones hautes fréquences et donne lieu à un phénomène structuré appelé artefacts de bloc, plus visible sur les zones uniformes.

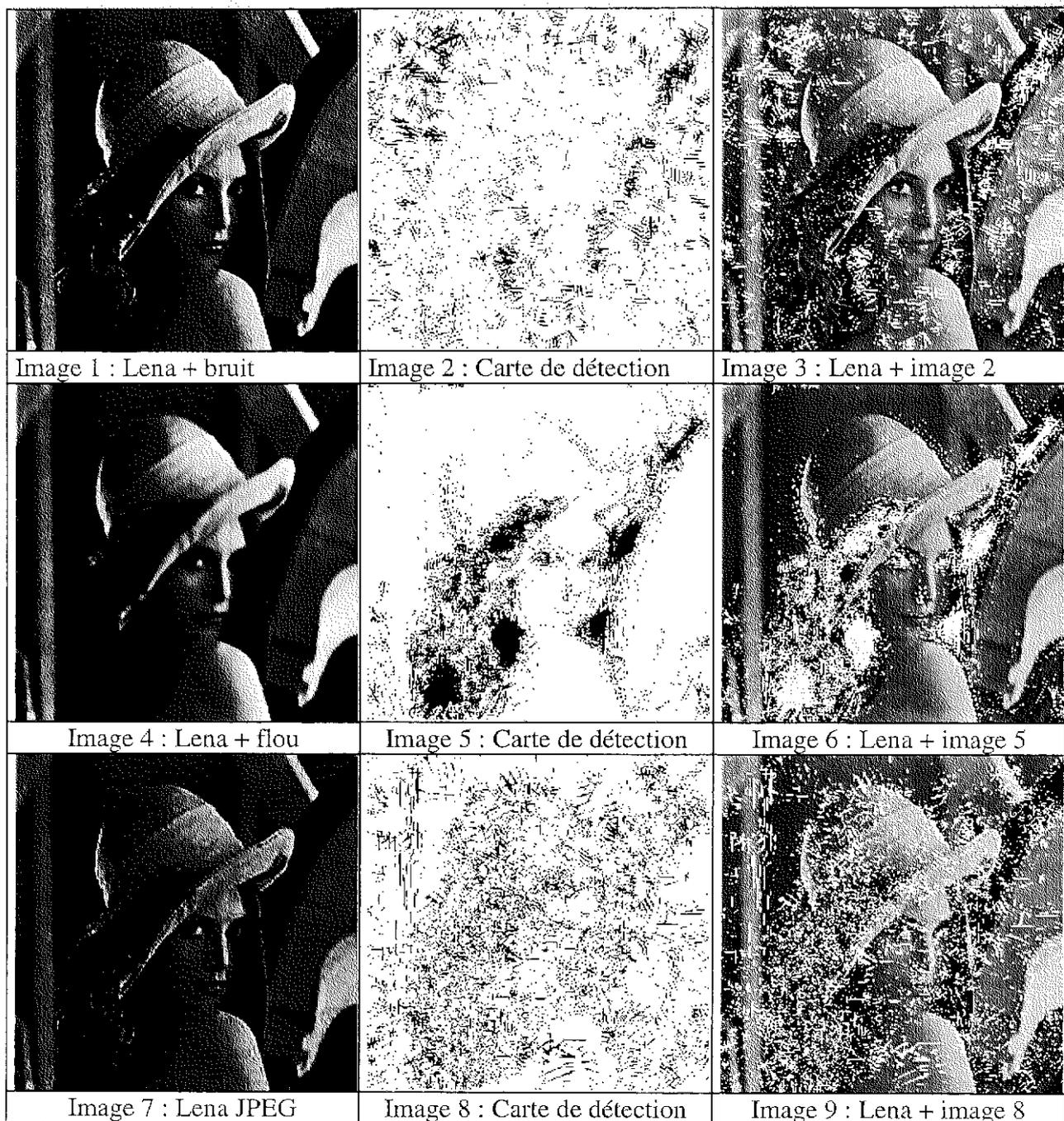


Figure 11 : Cartes de détection (inversion vidéo) des distorsions du VDP pour différentes dégradations.

g - Mesure de la qualité et limitations

La qualité de l'image peut être déterminée de plusieurs manières à partir de la carte d'erreur. Daly propose de compter le nombre de pixels dont la probabilité de détection est supérieure à 90 %. [Mat96] propose quant à lui de calculer une erreur quadratique moyenne entre l'image originale et l'image dégradée seulement pour les pixels dont la probabilité de détection est supérieure 90%.

Ce modèle fonctionne bien dans l'ensemble, même si le nombre de paramètres à régler est important. De nombreux tests psychophysiques ont été réalisés afin de le valider [Dal94]. Le VDP est néanmoins un algorithme qui demande un temps de calcul important. En effet, il nécessite 64 transformées de Fourier et approximativement 200 opérations (multiplications et

divisions) en virgule flottante. Pour une image de taille $N \times N$, le nombre d'opérations est de l'ordre de $200 \times N^2 + 64 \times 2 \times 2 \times N \times \log_2 N$. Avec une image de taille 256×256 , cela correspond à 13 millions d'opérations en virgule flottante, c'est-à-dire 4 minutes sur une station SUN 'SPARC 20'.

2) Le modèle de S. Karunasekera et N. Kingsbury [KaK93, KaK94, KaK95]

Ce modèle, un peu comme le précédent, calcule des images d'erreurs dans 5 bandes entre l'image originale et l'image dégradée, avant de les recombiner pour calculer une erreur subjective totale. La Figure 12 représente le modèle complet.

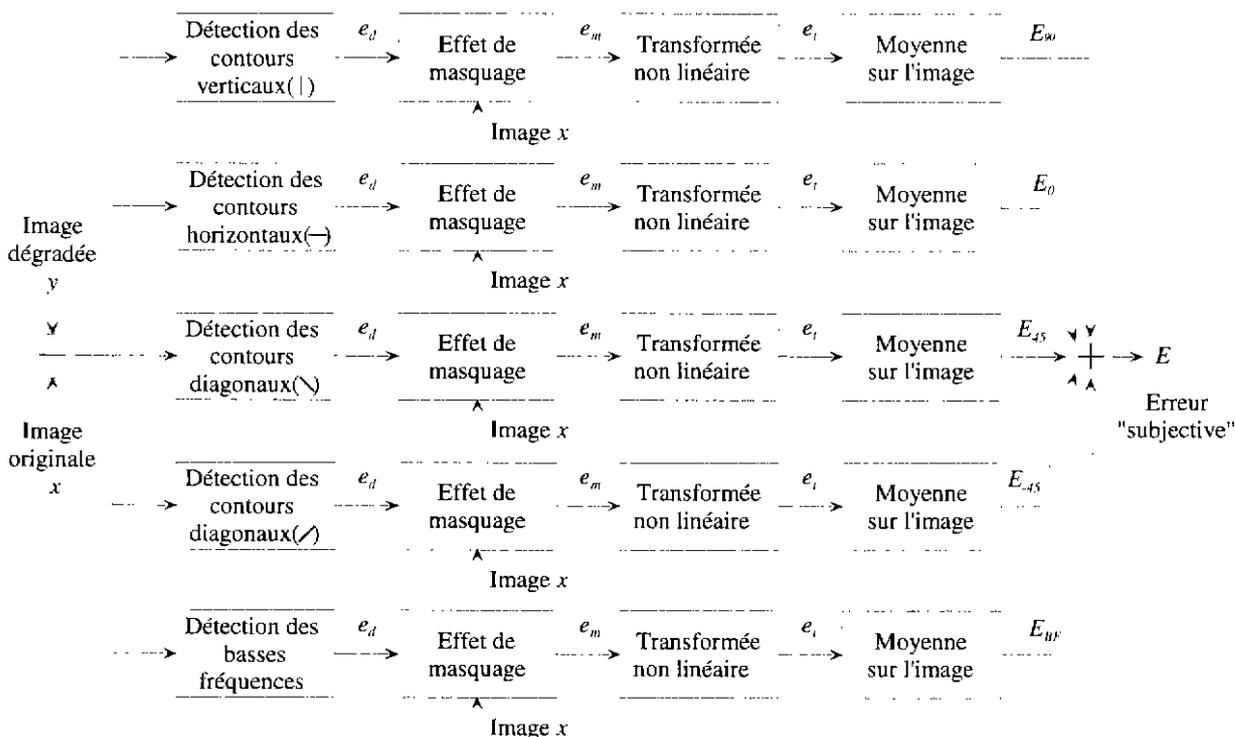


Figure 12 : modèle de Karunasekera et Kingsbury pour une mesure de la qualité image.

a - Détection des contours pour les erreurs directionnelles ($E_{90}, E_0, E_{45}, E_{-45}$)

Après des expériences visuelles, les contours sont détectés par un filtre passe-bas dans la direction du contour et un filtre passe-bande dans la direction perpendiculaire, ce dernier étant plus en accord avec les modèles de CSF qu'un filtre passe-haut.

Pour chaque direction de détection, le filtre utilisé est le même à une rotation près.

b - Modélisation de l'effet de masquage pour les erreurs directionnelles ($E_{90}, E_0, E_{45}, E_{-45}$)

Deux types de masquages sont considérés : le premier dû à l'activité (contenu structurel) de l'image dans une direction particulière et le second dû à la luminance moyenne de l'image. Les mesures de masquage sont réalisées à partir de l'image originale.

La Figure 13 représente le calcul de l'activité de masquage dans une direction θ . L'image originale (x) est filtrée de manière directionnelle (filtres spécifiques $H_D(\theta)$) pour détecter les contours dans la direction θ ($x_{d\theta}$). Afin de tenir compte de l'étalement spatial de l'effet de masquage, la valeur absolue de $x_{d\theta}$ est ensuite filtrée par un filtre passe-bas dans la direction perpendiculaire du contour ($H_D(\theta+90^\circ)$). On obtient ainsi la fonction d'activité directionnelle A_θ .

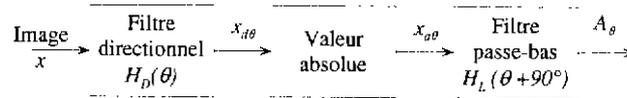


Figure 13 : mesure de l'activité directionnelle.

La fonction de masquage de l'image d'erreur par l'activité directionnelle, M_θ , est calculée d'après le modèle de Legge et Foley [LeF80] :

$$M_\theta = 1 + \left(\frac{A_\theta}{a_0} \right)^{r_a}$$

Les constantes a_0 et r_a ainsi que les caractéristiques des filtres sont déterminées par des expériences visuelles.

Le masquage de l'image d'erreur par la fonction de masquage est :

$$e_{a\theta} = \frac{e_{d\theta}}{M_\theta}$$

où $e_{d\theta}$ est l'image d'erreur sans masquage et $e_{a\theta}$ l'image d'erreur avec masquage de l'activité.

Le masquage $e_{m\theta}$ dû à la luminance moyenne de l'image est calculé à partir de la fonction de masquage de luminance B , obtenue par filtrage passe-bas de l'image originale :

$$e_{m\theta} = \frac{e_{a\theta}}{\left(1 + \left(\frac{B}{b_0} \right)^{r_b} \right)}$$

où $e_{a\theta}$ l'image d'erreur après le masquage de l'activité, b_0 et r_b sont des constantes déterminées expérimentalement.

La Figure 14 synthétise la modélisation du phénomène de masquage.

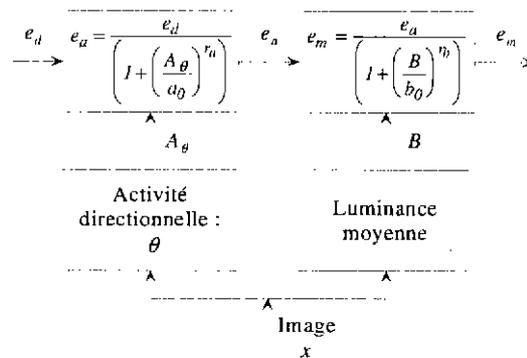


Figure 14 : modélisation de l'effet de masquage.

c - Calcul de l'erreur basse fréquence (E_{LF})

Les dégradations basses fréquences sont détectées à l'aide d'un filtre passe-bas Gaussien à symétrie circulaire qui permet de compléter l'espace de filtrage des contours des erreurs directionnelles. Deux masquages identiques aux précédents sont calculés :

d - Transformée non linéaire

Elle est modélisée par :

$$e_{r\theta} = (e_{m\theta})^{p_b}$$

où $e_{m\theta}$ est l'image d'erreur après masquage et p_0 est une constante déterminée expérimentalement dont la valeur est égale à 2.

e - Calcul de l'erreur finale:

Les erreurs directionnelles (E_{90} , E_0 , E_{45} , E_{-45}) et basses fréquences (E_{LF}) sont obtenues par une moyenne sur l'image d'erreur $e_{i\theta}$ et l'erreur finale est donnée par la somme des erreurs :

$$E_{kk} = \overline{E_{90}} + \overline{E_0} + \overline{E_{45}} + \overline{E_{-45}} + \overline{E_{LF}}$$

f - Conclusion

Le modèle de Karunasekera et Kingsbury est moins robuste que le VDP pour un grand nombre de dégradations. En effet, il repose sur une détection de contours dans 4 directions privilégiées au lieu de 30 pour le VDP. Cependant, pour des défauts directionnels marqués (artefacts de compression par exemple), sa fiabilité est plus grande.

Le principal avantage de ce modèle est qu'il ne nécessite aucune transformée de Fourier et peu d'opérations de multiplication. Pour une image 256x256 pixels, le temps de calcul est de l'ordre de 20 s sur une station SUN 'SPARC 20'.

C - Conclusion

Depuis le VDP, de nombreuses améliorations ont été apportées à la modélisation du SVH. L'utilisation de nouvelles transformées pour la décomposition multi canaux : filtres gaussiens [VDB96], en quadrature de phase [TeH94, WLB96], transformée en ondelette [LaJ97, Bra99], évolution du découpage en canaux fréquentiels [CSB00], permet des temps de calcul plus faibles et une meilleure reconstruction de l'image.

Des expériences psychophysiques [Hee92, FoB94, FoI94] ont montré que la modélisation de l'effet de masquage en considérant les canaux visuels indépendants ne permet pas de prédire correctement la réponse du SVH. [Hee92, TeH94, SiH97] proposent de pondérer le masquage dans un canal visuel, par le masquage dans les autres canaux.

Néanmoins, ces modèles bivariants, aussi complexes soient-ils, se révèlent totalement inutiles dans les nombreux cas où l'on ne dispose pas d'image de référence : recette en vol, évaluation de la qualité d'une image satellitaire de provenance inconnue ou peu fiable... Dans la plupart de ces situations, l'évaluation de la qualité des images est réalisée visuellement, par le personnel d'un service qualité. Un outil d'évaluation univariant, c'est-à-dire ne nécessitant pas l'image de référence, de la qualité des images serait une aide précieuse lorsque le nombre d'images à traiter est grand.

IV - Critères univariants

Nous n'avons trouvé, au cours de notre recherche, aucun critère général et univariant de qualité visuelle des images. En revanche, certains critères particuliers ont été développés pour des applications précises.

Par exemple, différents critères, la plupart statistiques, permettent d'évaluer le rapport signal à bruit des images [BoC92, KeH91, LeH89, MJR90]. Pour tous ces critères, la précision et la pertinence de l'estimation restent faibles. Nous n'avons pas trouvé de critère de qualité visuelle des images bruitées.

D'autres critères permettent d'évaluer la gêne visuelle occasionnée par les artefacts de compression [ABP96, CGC97].

Les exemples précédents ne sont en aucun cas une énumération exhaustive de tous les critères univariants que l'on peut trouver, néanmoins ils restent très peu nombreux et développés au cas par cas.

V - Conclusion

Les critères bivariants d'évaluation de la qualité des images existent en nombre. Des améliorations y sont apportées très régulièrement et leur fiabilité commence à devenir importante. L'évolution des moyens informatiques permet de modéliser des propriétés visuelles de plus en plus complexes, découpant l'espace perceptuel en canaux fréquentiels, directionnels...

En revanche il n'existe pas de critère univariant et universel pour évaluer la qualité d'une image. Des critères univariants spécifiques, assez peu nombreux, sont développés au cas par cas pour répondre à un besoin précis.

La méthode univariante qui est présentée dans la suite n'est pas un moyen universel d'évaluer la qualité d'une image. Son originalité est de pouvoir être adaptée d'un contexte à une autre avec le même concept. Elle peut être étalonnée, selon le contexte, soit sur des observateurs, soit sur des modèles bivariants. Les modèles bivariants que nous avons utilisés ne sont pas les plus performants, mais l'objectif est avant tout de valider le principe de cette méthode. Nous avons utilisé les modèles qui avaient déjà été utilisés à l'ONERA au cours d'une thèse en qualité image [JLO98] et qui s'étaient avérés performants.

Chapitre 2 : Une nouvelle approche

I - Introduction

Les méthodes bivariantes d'évaluation de la qualité des images, mentionnées dans le chapitre précédent, se révèlent inefficaces dans un nombre important de situations. Dans le cas particulier des images du satellite SPOT, par exemple, des observateurs (contrôle qualité) inspectent méticuleusement les images de manière à détecter et corriger d'éventuels défauts (cf. ANNEXE E -). Ces images sont de très grande taille et la tâche est fastidieuse. Il n'existe évidemment pas d'image de référence permettant de rechercher ces défauts automatiquement. En revanche, les défauts susceptibles d'être présents dans l'image sont connus et disposer d'un outil capable d'en détecter la présence et, le cas échéant, d'en évaluer l'importance serait une aide précieuse.

Selon l'application considérée, on peut souhaiter une détection du défaut calquée sur le comportement visuel, quantifiée sur une échelle représentant la gêne occasionnée, ou une détection plus mathématique quantifiée sur une échelle absolue. En effet, une dégradation invisible dans l'image peut parfois en gêner l'utilisation et, inversement, une dégradation très visible peut ne pas perturber certaines applications. A titre d'exemple, un bruit d'égalisation présent sur une zone très structurée de l'image peut être invisible mais fausser d'éventuelles mesures radiométriques.

Nous proposons ici une méthode permettant d'évaluer la qualité d'une image sans utiliser d'image de référence. Il s'agit dans un premier temps d'apprendre à détecter une dégradation sur des exemples connus pour ensuite utiliser cet apprentissage sur des images inconnues. Cela consiste à étalonner un modèle sur des mesures connues pour, ensuite, l'utiliser en interpolation. En raison de la complexité des phénomènes à modéliser nous avons choisi d'utiliser un réseau de neurones artificiels (RN). Ce modèle, non linéaire, a le double avantage d'être un excellent interpolateur tout en disposant de méthodes très simples d'étalonnage de ses paramètres.

En pratique, il s'agit dans un premier temps de caractériser la dégradation sur l'image pour en extraire une signature mathématique. Ensuite le RN apprend à associer, grâce à des exemples connus, la signature du défaut et la qualité de l'image. Finalement il est utilisé de façon autonome sur des images inconnues pour, à partir de la signature des défauts qu'elles contiennent, en évaluer la qualité.

II - Le modèle neuronal

A - Introduction

1) Le cerveau humain - La mémoire

Le cerveau humain est composé de milliards de neurones interconnectés pour former un réseau de communication complexe. En effet, chaque neurone est connecté à des milliers d'autres et cette connexion est utilisée pour transmettre de l'information. Cette information peut varier en intensité d'une connexion à une autre, permettant ainsi à un neurone d'avoir une influence plus ou moins grande sur ses voisins : les liaisons neuronales sont pondérées.

Bien que le processus de mémorisation ne soit pas parfaitement expliqué, on peut en partie l'expliquer par la fixation de la valeur des poids des connexions entre les neurones d'une partie bien précise du cerveau, qui varie selon le contexte.

2) Le réseau de neurones artificiels (RN)

Un réseau de neurones artificiels est un modèle non linéaire qui permet d'associer à un jeu de N paramètres dits « paramètres d'entrée », P valeurs qui en dépendent. Ce modèle est construit avec une architecture semblable à celle du cerveau humain. Il comporte un grand nombre de cellules individuelles, les neurones, fonctions élémentaires, connectées les unes aux autres. Ces « connections neuronales », pondérées, constituent les paramètres du modèle. Le réseau de neurones artificiels est donc une combinaison linéaire de fonctions non linéaires élémentaires.

Un des avantages du réseau de neurones sur d'autres modèles est la méthode utilisée pour le calibrer : des exemples (entrée/sortie) lui sont présentés et un algorithme mathématique adapte progressivement les connections neuronales de manière à permettre au réseau de retrouver seul la sortie attendue. Ces exemples constituent la « base d'apprentissage » du RN. Ce type d'apprentissage permet de modéliser des phénomènes complexes sans modèle physique explicite. En particulier, dans le cas de la vision humaine, la constitution d'une base d'apprentissage réalisée avec des tests visuels permettra de reproduire un comportement proche de celui du système visuel humain, sans modèle explicite de vision.

B - Principe de fonctionnement

1) Le neurone artificiel

Le neurone artificiel, brique élémentaire du RN, est une fonction non linéaire de $\mathcal{R}^n \rightarrow \mathcal{R}$. Cette fonction associe dans un premier temps au vecteur d'entrée (e), une valeur qui est la somme pondérée des entrées e_i . La sortie du neurone est donnée par application à cette somme d'une fonction non linéaire, dite fonction d'activation. La seule condition imposée par l'algorithme de calcul, plus tard utilisé pour évaluer le poids des connections neuronales, est que cette fonction soit différentiable. En pratique, il est aussi préférable qu'elle soit bornée. La Figure 15 représente un neurone artificiel.

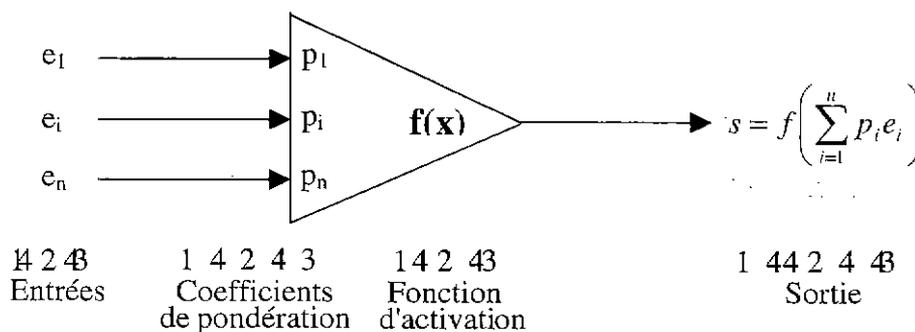


Figure 15 : Le neurone artificiel.

Une fonction d'activation fréquemment utilisée est la fonction sigmoïde :

$$f(x) = \frac{1}{1 + e^{-kx}}$$

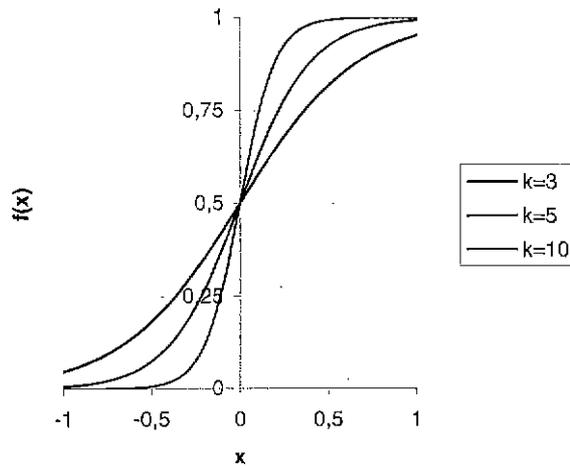


Figure 16 : Quelques exemples de sigmoïdes

Une propriété de la sigmoïde est de vérifier :

$$f'(x) = k \cdot f(x)[1 - f(x)] \tag{1}$$

2) Le réseau de neurones

a - Architecture

Les neurones sont interconnectés pour former un « réseau de neurones ». Selon l'architecture du réseau, l'algorithme d'apprentissage et les performances peuvent être différentes. Une des architectures les plus communément utilisées, en raison de la simplicité de sa mise en œuvre, est une architecture de réseau multicouche dont l'algorithme d'apprentissage est une rétropropagation de gradient (Figure 17). C'est cet algorithme d'apprentissage qui est détaillé plus loin. Un tel RN est constitué d'une « couche d'entrée », composée de neurones à une entrée recevant les paramètres d'entrée du modèle, d'une ou plusieurs « couches cachées », et d'une couche de sortie, composée d'autant de neurones que de sorties souhaitées pour le modèle. Au sein du RN, les entrées de chaque neurone sont les sorties de tous les neurones de la couche précédente.

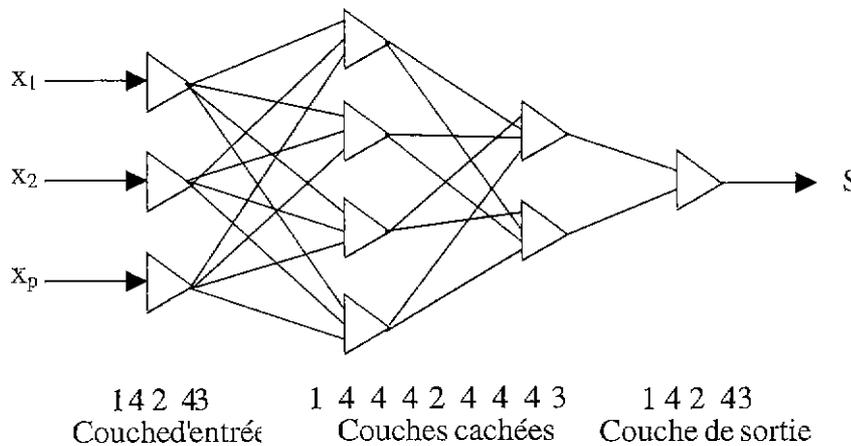


Figure 17 : Architecture du RN à rétropropagation de gradient.

Concevoir l'architecture d'un RN, c'est donc choisir le nombre de couches cachées ainsi que le nombre de neurones sur chacune de ces couches. Il n'existe pas de règle pour effectuer ce choix, qui reste empirique [SiD88, LoC86]. Néanmoins, pour un grand nombre d'applications, essentiellement en classification, un RN à une couche cachée s'avère suffisant et il existe des méthodes de réduction, destinées à réduire le nombre de neurones dans un réseau. Le paragraphe suivant présente rapidement l'algorithme d'apprentissage ainsi que les différentes méthodes de réduction.

b - Apprentissage et validation

Le RN, comme tout autre modèle physique, nécessite d'être étalonné. La première phase, dite phase d'apprentissage, consiste à lui présenter des exemples connus (entrée/sortie) qui vont permettre d'ajuster les poids des connexions neuronales. Malgré leur nombre très important, cet ajustement est réalisé de manière très simple par un algorithme de rétropropagation du gradient (cf. paragraphe suivant). Après la phase d'apprentissage, il est utilisé en mode autonome et fonctionne comme un interpolateur.

i -Rétropropagation de gradient

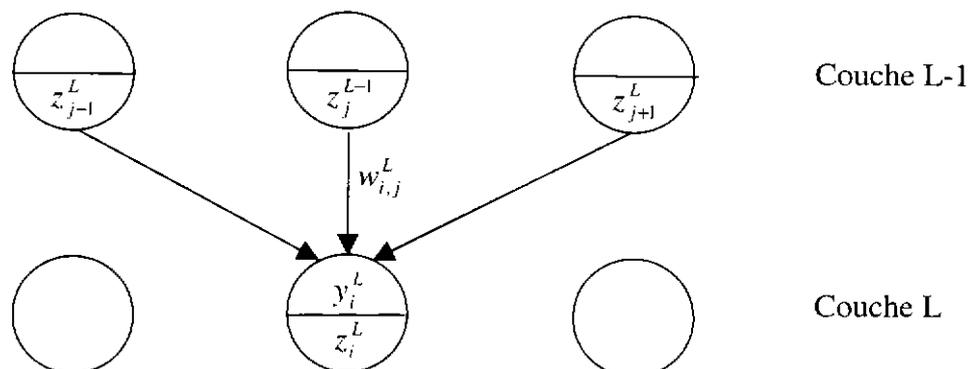
A chaque itération de l'apprentissage, un exemple (entrée/sortie) est présenté au RN. Celui-ci calcule une estimation de la sortie compte tenu de sa configuration, et réajuste les poids des connexions neuronales de manière à minimiser l'erreur (sortie calculée - sortie théorique). L'algorithme réajuste dans un premier temps le poids du neurone de sortie et une relation de récurrence permet de propager cette correction aux neurones de la couche précédente et ainsi de suite jusqu'à la couche d'entrée.

Pour cette raison, cet algorithme est appelé « gradient stochastique ». Il repose sur l'hypothèse que minimiser l'erreur à chaque itération revient statistiquement à minimiser l'espérance de l'erreur sur l'ensemble des exemples de la base.

Dans le cas d'un unique exemple d'apprentissage, le principe est le suivant :

Pour un RN comportant q couches numérotées de 1 à q , un neurone de sortie et dont la fonction d'activation est f , on note :

- n_L le nombre de neurones de la couche L ;
- $w_{i,j}^L$ le poids affecté à la connexion entre le $j^{\text{ème}}$ neurone de la couche L-1 et le $i^{\text{ème}}$ neurone de la couche L ;
- f la fonction d'activation du RN ;
- z_j^L la sortie du $j^{\text{ème}}$ neurone de la couche L ;
- y_i^L l'entrée totale du $i^{\text{ème}}$ neurone de la couche L : $y_i^L = \sum_j w_{i,j}^L z_j^{L-1}$;
- t la sortie théorique attendue.



On a donc :

$$z_j^L = f(y_j^L) \quad (2)$$

L'erreur à minimiser est :

$$E = (z_1^q - t)^2 \quad (3)$$

La méthode de descente du gradient repose sur une approximation linéaire de la fonction d'erreur :

$$E(\vec{w} + \Delta\vec{w}) \approx E(\vec{w}) + \Delta\vec{w}^T E'(\vec{w}) \quad \text{où} \quad \Delta\vec{w} = -\eta \cdot \text{grad}(E(\vec{w})) \quad (4)$$

$\eta > 0$ est appelé le taux d'apprentissage.

Pour $L < q$ on a :

$$\Delta w_{i,j}^L = -\eta \frac{\partial E}{\partial w_{i,j}^L} = -\eta \frac{\partial E}{\partial y_j^L} \frac{\partial y_j^L}{\partial w_{i,j}^L} = -\eta \cdot z_i^{L-1} \frac{\partial E}{\partial y_j^L} \quad (5)$$

Or :

$$\frac{\partial E}{\partial y_j^L} = \frac{\partial E}{\partial z_j^L} \frac{\partial z_j^L}{\partial y_j^L} = f'(y_j^L) \frac{\partial E}{\partial z_j^L} \quad (6)$$

Comme :

$$\frac{\partial E}{\partial z_j^L} = \frac{\partial E}{\partial y_i^{L+1}} \frac{\partial y_i^{L+1}}{\partial z_j^L} = \frac{\partial E}{\partial y_i^{L+1}} \frac{\partial}{\partial z_j^L} \sum_k w_{i,k}^{L+1} z_k^L = \frac{\partial E}{\partial y_i^{L+1}} w_{i,j}^{L+1} z_i^L \quad \forall i \in [1, n_{L+1}] \quad (7)$$

D'après (5) utilisé pour la couche L+1 on a :

$$\frac{\partial E}{\partial y_i^{L+1}} = \frac{\Delta w_{i,j}^{L+1}}{z_i^L} \quad (8)$$

Donc finalement avec (5), (6) et (8) :

$$\Delta w_{i,j}^L = -\eta \cdot z_i^{L-1} f'(y_j^L) w_{i,j}^{L+1} \Delta w_{i,j}^{L+1} \quad (9)$$

Pour $L=q$:

$$\Delta w_{1,j}^q = -\eta \frac{\partial E}{\partial w_{1,j}^q} = -\eta \frac{\partial E}{\partial z_1^q} \frac{\partial z_1^q}{\partial y_1^q} \frac{\partial y_1^q}{\partial w_{1,j}^q} = -2\eta (z_1^q - t) f'(y_1^q) z_1^{q-1} \quad (10)$$

En résumé, la relation de récurrence qui permet de rectifier les poids des connexions neuronales est, compte tenu de (1) :

$$\boxed{\begin{aligned} \Delta w_{1,j}^q &= -2\eta k \cdot (z_1^q - t)(1 - z_1^q) z_1^q z_1^{q-1} \\ \Delta w_{i,j}^L &= -\eta k \cdot z_i^{L-1} z_j^L (1 - z_j^L) w_{i,j}^{L+1} \Delta w_{i,j}^{L+1} \end{aligned}} \quad (11)$$

La démonstration de l'algorithme de rétropropagation du gradient, généralisée au cas où l'erreur à minimiser est calculée sur plusieurs exemples, peut-être trouvée dans [RhW85].

ii -La phase d'apprentissage

Si l'on se souvient que les paramètres « flottants » du RN sont les poids affectés aux connexions neuronales, on peut en calculer le nombre. Pour un RN comportant N couches, chacune constituée de n_L neurones, le nombre de connexions neuronales est :

$$\sum_{L=1}^{N-1} n_L \cdot n_{L+1}$$

Le nombre d'exemples à présenter au RN pour réaliser l'apprentissage dépend du nombre de connexions neuronales. Il doit cependant être suffisamment important pour permettre ensuite

une interpolation convenable sur des exemples inconnus. On considère généralement que le nombre d'exemples pour la phase d'apprentissage varie en $(\sum n_i)^2$. Un choix méticuleux des exemples d'apprentissage sur le domaine de stabilité désiré peut cependant permettre de réduire significativement ce nombre. En particulier, lorsqu'on dispose de peu d'exemples, on peut intervertir régulièrement en cours d'apprentissage les exemples de la base d'apprentissage qui ont été les mieux appris avec ceux de la base de test sur lesquels le RN généralise le moins bien. Cette méthode permet d'augmenter la capacité de généralisation du RN.

L'ensemble des exemples (entrée/sortie) présentés au RN pendant la phase d'apprentissage constitue la « base d'apprentissage ». L'apprentissage est réalisé et réitéré sur les éléments de la base tant que l'erreur commise entre la sortie prédite par le RN et la sortie attendue diminue. Cependant, étant donné que le nombre de neurones du RN et par conséquent le nombre de paramètres, est fixé empiriquement, on peut facilement tomber en phase de surapprentissage si le RN contient trop de neurones par rapport au nombre d'exemples. A ce moment là, il apprend le bruit contenu dans la base d'apprentissage et perd sa capacité de généralisation. Pour éviter cela, le RN est régulièrement testé, pendant la phase d'apprentissage, sur une base indépendante de la première, la base de test. En pratique, tant que l'erreur sur chacune des bases diminue, la capacité de généralisation du RN augmente. Quand l'erreur moyenne sur la base de test cesse de diminuer et se met à augmenter alors que l'erreur sur la base d'apprentissage diminue encore, l'apprentissage est arrêté. Au delà on entre en phase de surapprentissage.

Notons par ailleurs que le RN aura un comportement d'autant plus fiable que les images à évaluer seront proches de celles contenues dans la base d'apprentissage.

iii -Les différentes méthodes de réduction du RN (« pruning »)

Etant donné que seul le nombre d'entrées et celui de sorties du RN sont imposés par le problème physique à modéliser, le nombre de couches cachées et celui de neurones par couches cachées est choisi empiriquement. Il n'existe pas de règle mathématique de décision. Pour cette raison, certains algorithmes de réduction (*pruning* en anglais) ont été développés de manière à réduire en cours d'apprentissage le nombre de neurones utilisés.

L'intérêt de réduire un RN peut-être :

- gagner en temps de calcul et en utilisation de la mémoire ;
- améliorer la capacité de généralisation ;
- prouver l'importance physique de certains paramètres d'entrée ;
- réduire la taille de la base d'apprentissage...

La plupart des méthodes utilisent des algorithmes qui observent la sensibilité des connexions neuronales afin de supprimer celles qui ont les poids les plus faibles.

Aucune méthode de réduction n'a été utilisée dans le cadre de ce travail, l'objectif étant de montrer l'intérêt d'utilisation d'un RN sans pour autant chercher à optimiser les modèles. Pour cette raison elles ne sont pas décrites plus en détails. Les algorithmes peuvent être trouvés dans [Bie94, Sch94]. Cependant une étude de sensibilité a posteriori permettra de déterminer, parmi les composantes d'entrées du RN, celles qui ont une contribution physique importante à l'étude de la qualité des images.

[SiD88] arrivent aux conclusions suivantes quant à la réduction des RN :

- la réduction d'un RN augmente sa résistance au bruit ;
- une réduction trop importante introduit des distorsions ;
- l'apprentissage et la réduction gagnent à être réalisés sur des données bruitées.

III - Mise en œuvre

L'objectif est d'évaluer la qualité d'une image. Comme déjà expliqué, la qualité peut être une simple considération esthétique, ou bien peut dépendre de l'utilisation à laquelle est destinée l'image. Notre objectif ici est de montrer que l'on peut utiliser une approche neuronale dans n'importe lequel de ces contextes. Pour ce faire, nous allons montrer que l'on peut étalonner la méthode neuronale univariante sur plusieurs notations de qualité bivariantes ou visuelles. La méthode générale, applicable à tout type de dégradation et d'évaluation de la qualité, est décrite ci-dessous. Elle comporte trois étapes :

- une caractérisation de la dégradation, c'est-à-dire une extraction de sa signature ;
- un apprentissage avec un RN pour mettre en relation la signature et la qualité ;
- une validation sur des exemples inconnus.

La difficulté de mise en œuvre est liée à la multiplicité des dégradations qui peuvent exister sur une image. De plus une combinaison successive de deux dégradations, par exemple, n'aboutit pas forcément au même résultat en terme de qualité image. La Figure 18 montre que la combinaison flou+bruit n'est pas commutative. Pour ces exemples le flou est obtenu grâce à un modèle de défocalisation d'instrument (cf. Chapitre 3 :II -B -1)) réglé sur la même valeur pour les différentes images, le bruit ajouté est gaussien et le SNR avec l'image parfaite est de 10.

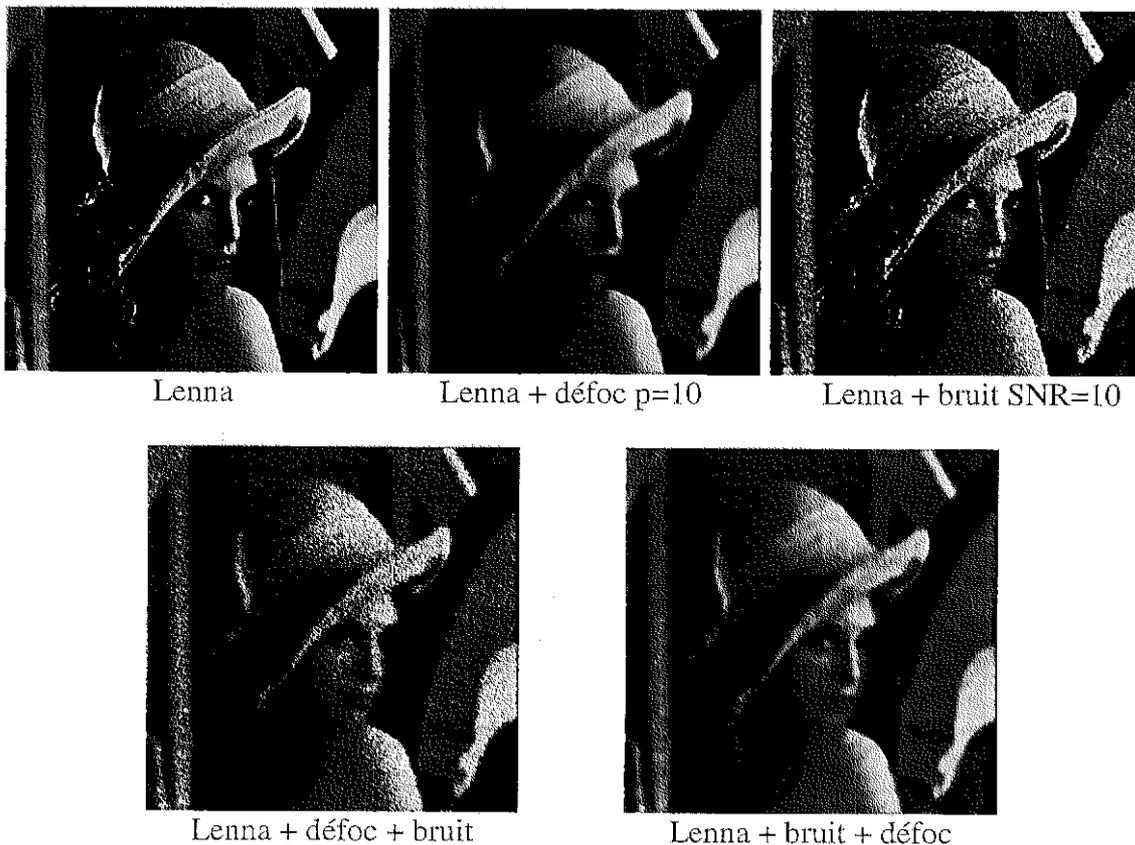


Figure 18 : Mise en évidence de la non permutabilité de deux dégradations.

De la même façon, on peut imaginer qu'à une même image dégradée correspondent plusieurs « familles » de dégradations possibles.

Pour cette raison, nous avons choisi dans un premier temps de nous limiter à des dégradations simples, c'est-à-dire non mélangées les unes avec les autres : flou, artefacts de compression,

bruit... Ces dégradations seront classées en deux catégories : celles qui sont réparties uniformément sur toute l'image (flou, bruit...) et celles qui sont très localisées, voire ponctuelles.

A - Etape fondamentale : caractérisation de la dégradation

La littérature montre peu de tentatives de caractérisation de la qualité d'une image grâce à un RN, et celles qui existent restent peu convaincantes [ABP96] (cf. Chapitre 3 :I -). Il semble que ce soit plus souvent le fruit d'une mauvaise utilisation du RN que d'une inadaptation du modèle. Ces tentatives utilisent très souvent comme paramètres d'entrée du RN, les niveaux de gris de tous les pixels sur un voisinage de taille donné, sans extraire la plupart du temps de signature de la dégradation à mettre en évidence.

Pour cette raison, **l'étape la plus importante de notre démarche consiste à extraire la signature de la dégradation**. C'est un jeu de paramètres, calculé sur l'image dégradée, qui doit être suffisant pour évaluer la qualité de l'image et qui constituera l'entrée du modèle neuronal. Selon que la note de qualité désirée caractérise la qualité visuelle ou simplement une grandeur caractéristique de la dégradation considérée, ce jeu de paramètres, que l'on a choisi d'appeler le « vecteur caractéristique » de la dégradation, peut être plus ou moins complexe. En effet, donner une note de qualité visuelle, par exemple, implique de prendre en compte certaines propriétés du système visuel : par exemple un même rapport signal sur bruit sur les zones uniformes et sur les zones structurées d'une image provoquera une gêne plus importante sur les zones uniformes. A ce moment là, le vecteur caractéristique doit contenir des informations non seulement sur le défaut, mais aussi sur le type de zone ou le type d'image. Un défaut ponctuel peut être masqué par son voisinage, si ce dernier est très structuré, et donc être invisible.

D'une manière générale, plus le vecteur caractéristique contient d'informations, plus l'évaluation de la qualité est précise. Cependant ne perdons pas de vue que la taille du vecteur caractéristique détermine la taille du RN et donc de la base d'apprentissage nécessaire... Il s'agit de trouver le meilleur compromis précision/taille du RN.

Dans la pratique, nous disposons de tous les outils mathématiques, issus du traitement du signal et des images, de la connaissance ou non des dégradations, pour extraire le vecteur caractéristique. L'image nécessite souvent un prétraitement pour amplifier les caractéristiques de la dégradation.

B - Apprentissage – validation

1) Apprentissage

La phase d'apprentissage du RN, comme tout autre modèle, va conditionner son comportement après coup. Un soin très important doit être en particulier accordé au choix des exemples qui vont constituer la base d'apprentissage. Le RN est un très bon interpolateur, mais son comportement en extrapolation n'est pas aussi bien maîtrisé que celui d'un modèle fabriqué d'après des considérations physiques. En conséquence, les images de la base devront être représentatives de tous les types d'images sur lesquels on désire un comportement stable et fiable du RN. Pour un apprentissage général, il faut des images uniformes, structurées... de tous types. Plusieurs intensités de dégradations sont aussi nécessaires.

Deux cas de figure peuvent se présenter. Il se peut que l'on sache reproduire la dégradation : par exemple une compression JPEG. Dans ce cas, les base d'apprentissage et base de test

peuvent être fabriquées de toutes pièces, respectant un éventuel contexte qui peut être imposé par une application précise (image satellitaires, résolution...). Ils se peut aussi que l'on ne sache pas modéliser la dégradation. Dans ce cas un apprentissage peut être réalisé « sur le tas » à partir des images dégradées dont on dispose (cf. Chapitre 3 :II -C -).

Pour le premier cas de figure, nous avons testé deux types d'apprentissages pour chaque dégradation : un apprentissage général, sur une base d'images très variées et un apprentissage spécifique sur une base d'images SPOT de même résolution (mode P, 10m de résolution). Des tests croisés (apprentissage général/test spécifique et apprentissage spécifique/test général) permettent d'évaluer la stabilité des modèles.

La base d'images générales que nous avons utilisée (cf. ANNEXE A -) est constituée de 157 images 512x512 pixels non dégradées de toutes sortes. Pour chacune des dégradations étudiées, chaque image de cette base est dégradée avec dix intensités différentes couvrant la plage de qualité que l'on désire apprendre à estimer. Pour l'apprentissage, nous disposons donc de 1570 exemples dont la qualité doit être préalablement évaluée par le modèle bivariant dont on veut reproduire le comportement (éventuellement un observateur).

La base d'images spécifiques est quant à elle constituée de 157 morceaux 512x512 pixels d'images SPOT acquises en mode P, c'est à dire dans une unique bande spectrale à la résolution de 10m. Ces images sont dégradées et utilisées de la même manière que celle de la base générale.

Il est important de préciser que ces images, qui servent d'exemples pour la phase d'apprentissage, ne sont pas toutes parfaites. Elles ont cependant été triées visuellement lors de la constitution des bases.

2) Validation

Une fois les apprentissages réalisés, ils sont validés sur des images inconnues, dégradées. Les bases de validation générale et spécifique utilisées dans le cadre de cette étude sont toujours les mêmes. Les images parfaites de ces bases sont montrées Figure 19 et Figure 20. Pour la validation, ces images sont dégradées selon différentes intensités et évaluées à la fois par le modèle univariant et par le modèle qui a servi à étalonner l'apprentissage : modèle bivariant ou évaluation visuelle.

Au cours des études qui suivent, nous ne représenterons l'influence des différentes grandeurs physiques que pour les images Test000, Test005 et Test014. L'image Test000 est une image de champs, uniforme, l'image Test014 est quand à elle une image de ville, structurée dans son ensemble. L'image Test005, Lena, très utilisée en traitement d'images, est une image mixte contenant à la fois des régions uniformes et des régions structurées, des contrastes importants...

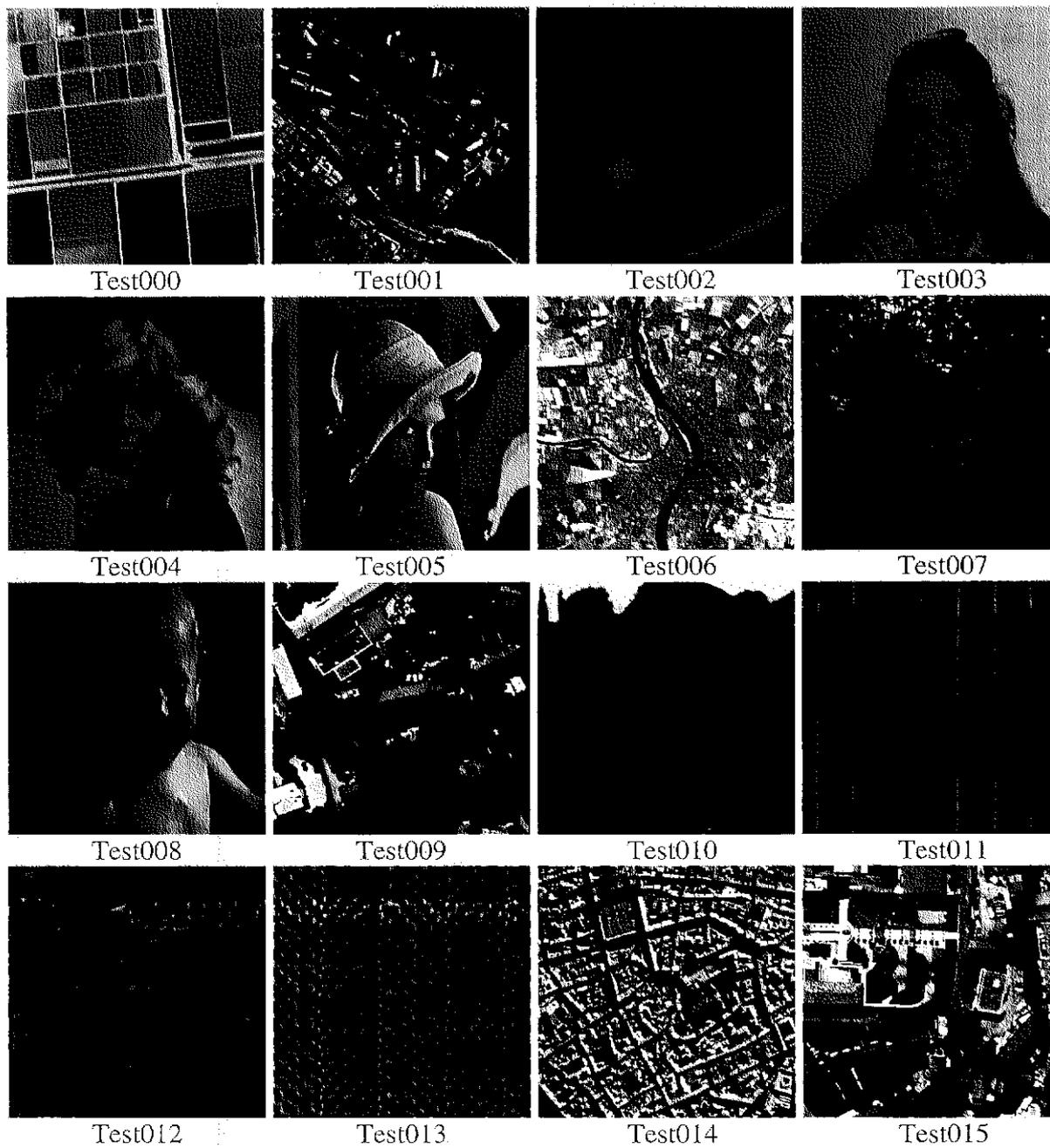
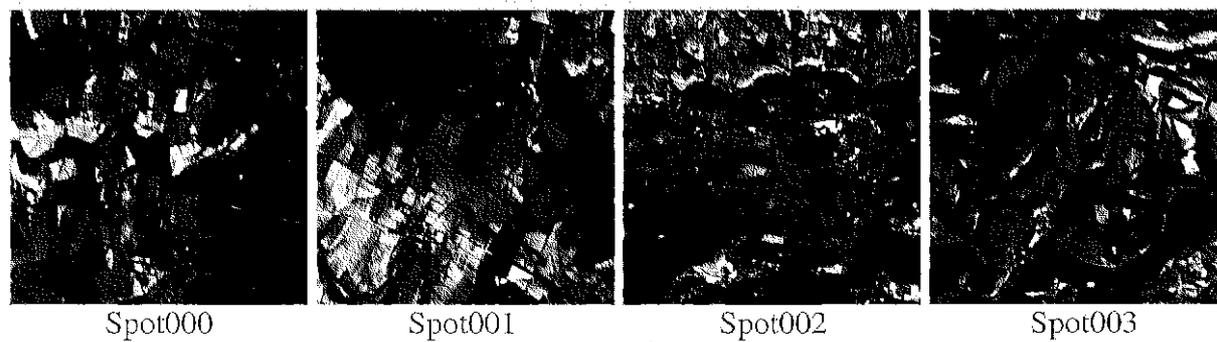


Figure 19 : Base de validation générale.



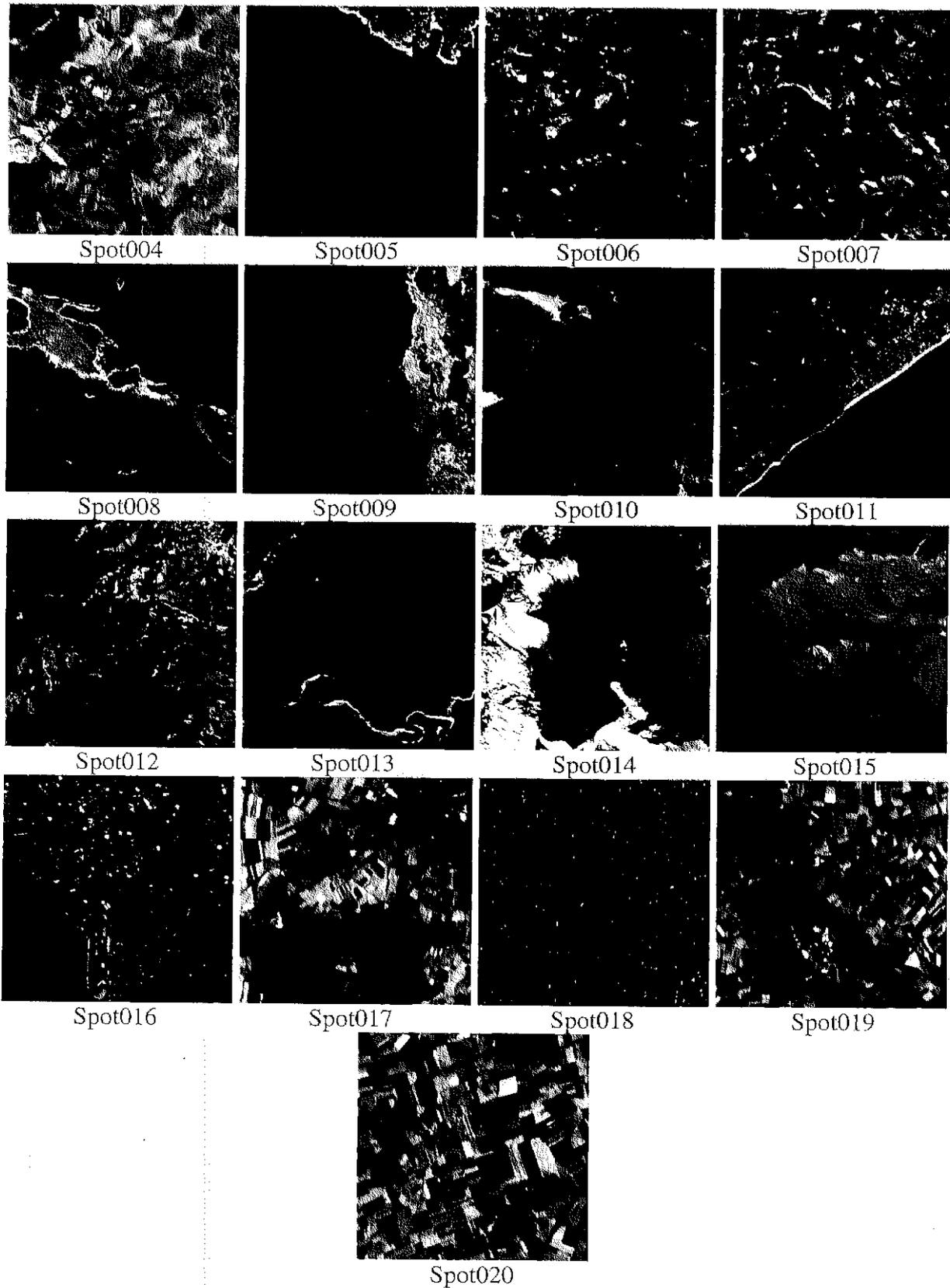


Figure 20 : Base de validation spécifique.

C - Etude de sensibilité

L'avantage du modèle neuronal est de permettre une sélection des données physiquement importantes pour décrire un problème. Lors de la constitution du vecteur caractéristique, nous pouvons nous autoriser, compte tenu du nombre d'images de la base d'apprentissage, une trentaine de composantes. Certaines s'imposent d'elles même en fonction de la dégradation à mettre en évidence, mais d'autres, à la signification physique moins évidente (un moment d'ordre élevé en particulier), peuvent être intégrées dans le vecteur caractéristique. La phase d'apprentissage, qui attribue des poids aux connexions neuronales, favorisera les composantes qui ont effectivement une influence pour estimer la grandeur attendue. Nous utiliserons souvent cette propriété du modèle neuronal.

L'importance de chacune des composantes du vecteur d'entrée du RN, difficilement accessible directement, est estimée par une étude de sensibilité une fois l'apprentissage réalisé. Cette étude doit permettre de supprimer les composantes inutiles ou redondantes.

Nous avons défini la sensibilité du RN de la manière suivante : pour chaque type d'image, le vecteur d'entrée correspondant à l'image parfaite est calculé. La sensibilité vis-à-vis de la $j^{\text{ème}}$ composante est calculée en la faisant varier sur $[-1,1]$ (qui est l'intervalle de définition des entrées), toutes les autres étant constantes, et en évaluant la déviation maximale introduite sur la sortie du RN par rapport à la valeur correspondant à l'image parfaite.

Si $(x_0, \dots, x_j, \dots, x_p)$ sont les composantes du vecteur caractéristique d'une image non comprimée et $f(x_0, \dots, x_j, \dots, x_p)$ est la sortie du RN associée à cette image, la sensibilité $S(j)$ à la $j^{\text{ème}}$ composante est définie de la manière suivante :

$$S(j) = \text{Max}_{x_j \in [-1,1]} [f(x_0, \dots, x_j', \dots, x_p) - f(x_0, \dots, x_j, \dots, x_p)]$$

La sensibilité moyenne à la composante x_j sur l'ensemble des images est la moyenne sur les images des $S(j)$. La sensibilité vis-à-vis d'une composante représente la variation maximale que peut introduire cette composante, seule, dans l'estimation de la qualité.

Après l'étude de sensibilité, des apprentissages sont réalisés en supprimant les composantes les moins importantes en apparence. Très souvent, une conclusion sur l'importance des grandeurs utilisées à ce moment là est alors possible. Cette technique est connue sous le nom de « data mining » ou « sélection de données ».

Parfois, cependant, l'étude de sensibilité peut donner des résultats différents pour deux apprentissages réalisés sur un même problème, avec un même RN. Ce phénomène peut être expliqué par les redondances contenues dans le vecteur caractéristique, de dimension N. Le RN cherche à minimiser une erreur, dans un espace à N dimensions, entre la sortie attendue et celle calculée à partir du vecteur caractéristique. Par le jeu des redondances, le RN peut converger vers un minimum local, favorisant l'une ou l'autre des composantes ou un mélange des deux. Cette convergence dépend de l'état initial du RN qui est fixé aléatoirement.

Pour tenter de pallier ce problème d'interprétation, nous avons toujours réalisé, avant de conclure, plusieurs apprentissages pour plusieurs initialisations.

IV - Conclusion

La méthode présentée dans ce chapitre est une méthode générale, applicable quel que soit le type de défaut présent sur l'image et quelle que soit la façon d'en estimer la qualité : qualité visuelle, mathématique...

Cependant, il reste difficile de caractériser simplement les dégradations et le nombre d'exemples requis pour l'apprentissage est élevé.

Nous appliquerons dans un premier temps la méthode pour évaluer la qualité d'une image dégradée par une compression JPEG. Le défaut observé dans ce cas est structuré et le vecteur caractéristique relativement simple à extraire. Nous l'appliquerons aussi à des images floues, et nous montrerons qu'elle peut être utilisée pour retrouver certaines caractéristiques du système dégradant. Nous l'utiliserons ensuite pour modéliser la détection visuelle de défauts ponctuels sur une image, avant de proposer un moyen pour évaluer la qualité d'images dégradées par un mélange de défauts.

Par ailleurs, une étude de sensibilité systématique permet de tirer des conclusions sur l'importance physique des composantes du vecteur caractéristique pour décrire la qualité d'une image, relativement à une application donnée.

Chapitre 3 :
Application à une image dégradée
uniformément par un défaut unique

I - Evaluation de la qualité des images JPEG

A - Introduction

Dans ce chapitre, nous nous proposons d'évaluer la qualité d'une image comprimée selon la norme JPEG (cf. ANNEXE D -). Cette compression, en raison du protocole selon lequel elle est effectuée sur l'image, provoque des dégradations dont la plus gênante visuellement est périodique. Elle peut, pour cette raison, sembler facile à caractériser. Pourtant, la plupart des méthodes d'évaluation de la qualité des images JPEG sont des méthodes bivariantes, qui, de plus, ont des domaines de stabilité ou de validité très différents, selon le taux de compression ou la nature des images [MEC98]. Peu de méthodes univariantes ont été développées à cette fin. [ABP96] utilise, dans le cadre de l'évaluation de la qualité d'une image comprimée selon la norme JPEG, des coefficients de continuité entre les blocs à partir desquels un réseau de neurones estime la qualité de l'image. Cependant les résultats sont de l'ordre de 25 % de précision par rapport à la qualité théorique attendue, alors que ceux obtenus avec un modèle bivariant peuvent atteindre 5-10 % de précision [Oli98].

Dans cette partie, nous allons montrer que la méthode neuronale univariante peut être utilisée pour évaluer la qualité d'une image JPEG. Dans un premier temps, elle sera utilisée pour retrouver le facteur de qualité JPEG fixé lors de la phase de compression. Ce facteur est, en quelque sorte, une caractérisation de la quantité d'information perdue pendant la phase de quantification de la compression.

Dans un deuxième temps, nous montrerons qu'il est aussi possible de prédire la gêne visuelle occasionnée par la compression. Comme nous ne disposons pas de suffisamment d'observateurs pour mettre en place un protocole d'évaluation de la qualité visuelle des images JPEG, nous réaliserons un apprentissage de manière à reproduire, sans image de référence, le comportement du modèle visuel bivariant de S. Karunasekera et N. Kingsbury [KaK95].

B - Dégradations introduites par le codage

En raison de la nature même du codage [Wal91, EuC96], la perte de qualité de l'image s'effectue de manière cohérente sur les blocs de 8x8 pixels, indépendamment d'un bloc à l'autre (cf. ANNEXE B -). Le seul lien susceptible d'exister entre deux blocs dépend de la statistique de répartition des niveaux de gris dans l'image, et non du codage. La dégradation introduite par le codage, très visible pour de fortes compressions, est connue sous le nom *d'artefacts de blocs*. Au sein d'un même bloc, les pertes dépendent du contenu fréquentiel. Les hautes fréquences apparaissent assez vite dégradées quand le taux de compression augmente alors que les artefacts de blocs apparaissent plus tard, essentiellement sur les zones uniformes, comme le montre la Figure 21.

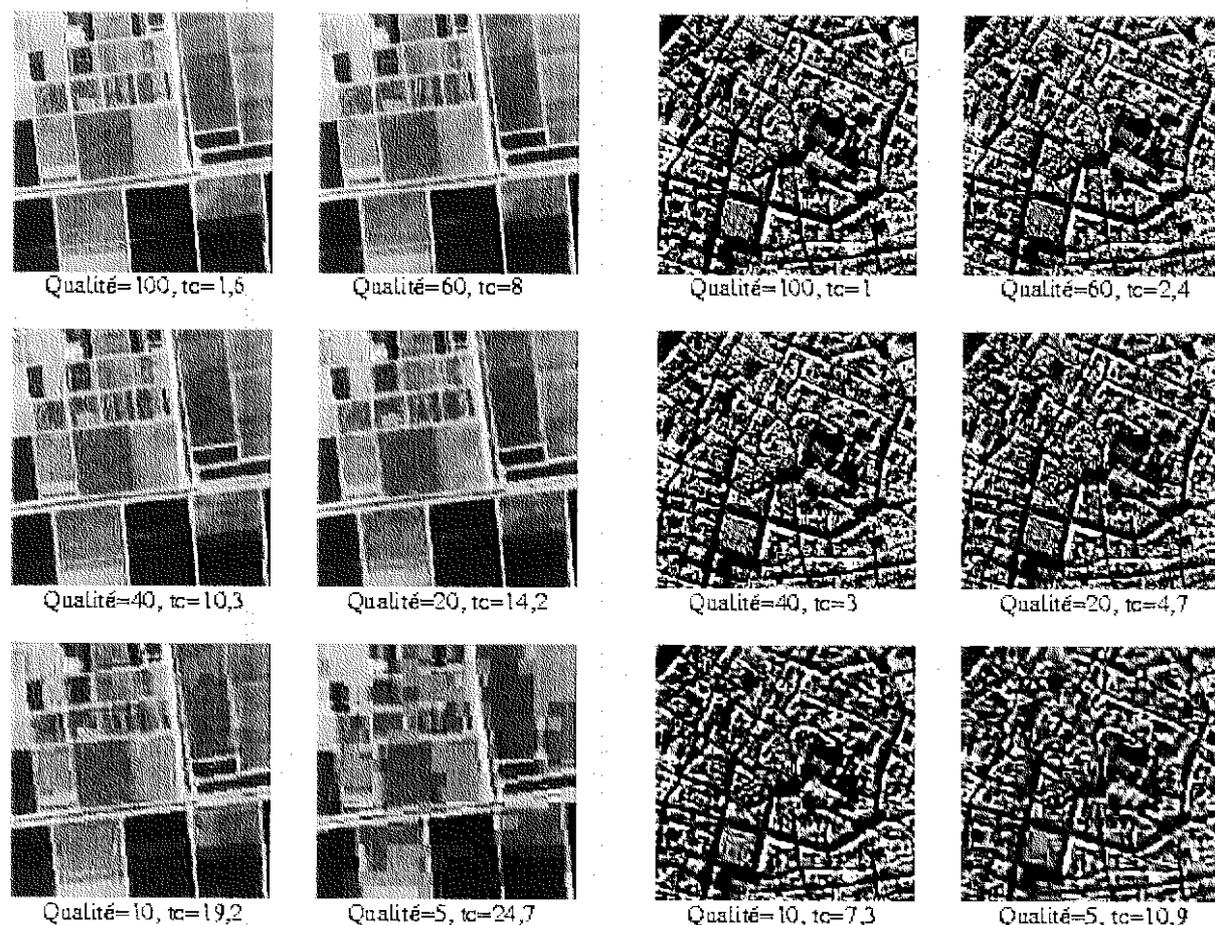


Figure 21 : Relation entre la qualité JPEG et le taux de compression pour une image uniforme de type champs et pour une image structurée de type ville.

Plusieurs outils existent déjà pour quantifier de manière bivalente la qualité d'une image JPEG [Oli98]. A partir d'une mesure de distance entre l'image non comprimée et l'image dégradée, ces outils permettent d'évaluer la qualité de l'image comprimée. Notre objectif est d'arriver à évaluer cette qualité, qui peut être soit l'indice de qualité JPEG, soit tout autre indice visuel ou purement mathématique, directement à partir de l'image dégradée.

L'apprentissage sera effectué d'une part sur l'indice de qualité JPEG choisi lors de la phase de compression, indice purement mathématique, d'autre part sur un modèle représentatif d'une notation visuelle, donc subjective.

C - Extraction du vecteur caractéristique

Rappelons une fois de plus que le choix des composantes du vecteur caractéristique n'est pas exclusivement fondé sur des considérations physiques (cf. Chapitre 2 :III -C -). Des grandeurs ayant une signification évidente sont bien sûr introduites dans le vecteur caractéristique, mais l'avantage de la méthode d'apprentissage avec un RN est qu'elle permet de sélectionner elle-même les données importantes. Pour cette raison, au cours de toutes les étapes du prétraitement de l'image, certaines grandeurs statistiques seront ajoutées au vecteur caractéristique.

1) Prétraitement : caractérisation des blocs

L'objectif de cette phase est d'arriver à quantifier l'importance du phénomène de bloc, indépendamment du type d'image. Cette information, capitale lors de la phase d'apprentissage, est pourtant difficilement observable sur l'image comprimée non traitée.

Les différentes étapes du prétraitement de l'image que nous avons développé sont représentées sur la Figure 22.

Pour une image Im de N lignes et N colonnes constituée des pixels $Im(i,j)$, on calcule tout d'abord le laplacien de l'image. Le noyau choisi permet de ne retenir que les composantes horizontale et verticale de l'image comprimée et donc de mettre en évidence les artefacts de blocs :

$$\Delta Im = Im * \begin{pmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{pmatrix}$$

Nous ne considérerons par la suite que la valeur absolue du laplacien de l'image, car seuls les contrastes aux frontières des blocs ont une importance dans l'évaluation de la qualité.

Les artefacts de blocs ont une structure dont la taille est un multiple de 8 pixels et peut varier d'un bloc à l'autre. Pour cette raison, il est difficile de les observer sur la transformée de Fourier du laplacien de l'image. Pour amplifier encore le phénomène, nous avons cumulé d'une part les lignes et d'autre part les colonnes de l'image pour former 2 vecteurs V_{horiz} et V_{vertic} :

$$V_{horiz}(k) = \sum_{i=1}^N |\Delta Im(i, k)|$$

$$V_{vertic}(k) = \sum_{j=1}^N |\Delta Im(k, j)|$$

Ces vecteurs caractérisent l'amplitude du phénomène dans les directions horizontale et verticale. Nous avons formé le vecteur $V(k)$ qui caractérise l'importance des artefacts de blocs sur l'image, indépendamment de son orientation :

$$V(k) = V_{horiz}(k) + V_{vertic}(k)$$

Ce vecteur est constitué d'impulsions périodiques, de période 8 pixels, et on retiendra comme valeur caractéristique de l'importance du phénomène de bloc, la valeur FF_8 de la FFT de $V(k)$ qui correspond à une périodicité de 8 pixels.

Notons que ce prétraitement permet de réduire d'un facteur $2N \cdot \log(N)$ la complexité de la FFT effectuée, sur le vecteur $V(k)$ plutôt que sur l'image. La Figure 23 montre l'évolution de ce coefficient en fonction de l'indice de qualité JPEG pour des images de types différents. On constate que, pour un même indice de qualité, ce coefficient est toujours plus grand pour une image très structurée (Test014) que pour une image uniforme (Test000), alors que visuellement l'image structurée paraît moins dégradée, comme le montre la Figure 21.

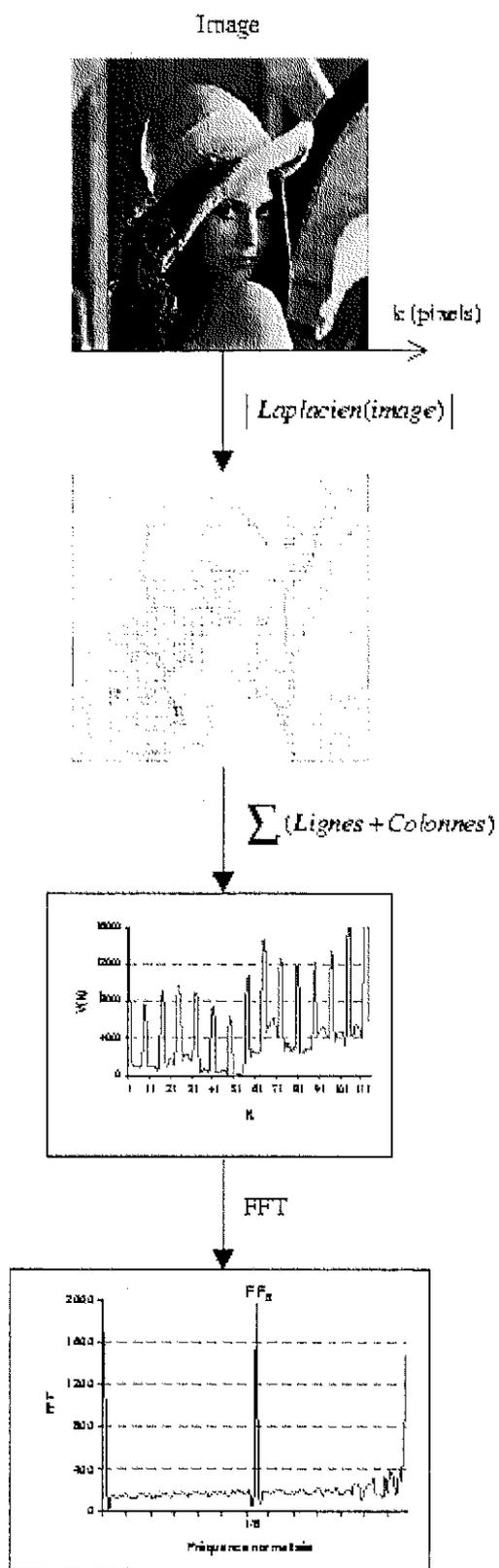


Figure 22 : Prétraitement et détection des artefacts de blocs. Exemple sur l'image Test005 (Lena) pour un facteur de qualité JPEG égal à 20 (taux de compression de 22)

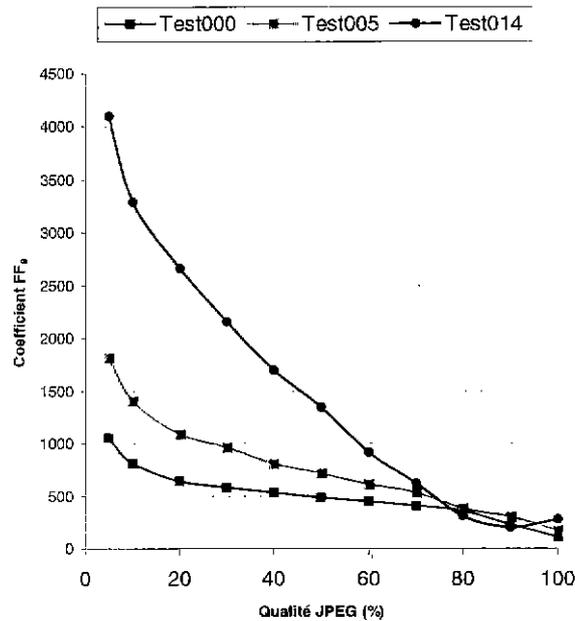


Figure 23 : Variation du coefficient FF_8 en fonction de la qualité JPEG selon le type d'image.

Ce coefficient, seul, n'est donc pas suffisant pour déterminer la qualité de l'image. En effet, selon le type d'image, il décrit l'une ou l'autre des courbes du faisceau de la Figure 23 représentant sa variation en fonction de la qualité de l'image. Pour pouvoir déterminer la qualité, quelle qu'elle soit, il devra être associé à des informations concernant le type de l'image, telles que son variogramme (cf. ANNEXE C -A3 -) ou ses moments. Ces informations permettront de déterminer la courbe sur laquelle on se situe (Figure 23), pour accéder ensuite à l'indice de qualité de manière bijective.

Parmi les grandeurs susceptibles de fournir une information complémentaire, nous avons choisi de retenir la valeur $FF_{7,9}$ de la FFT de $V(k)$, moyenne de FF_7 et FF_9 , ainsi que FF_{16} , de manière à disposer d'un élément de comparaison pour FF_8 . En effet, il se peut qu'une image contienne une structure périodique de période proche de 8 pixels. Ce sera le cas si $FF_{7,9}$ a sensiblement le même ordre de grandeur que FF_8 . Dans le cas contraire, la probabilité que la valeur de FF_8 soit uniquement due au phénomène de bloc est grande.

D'autre part et pour la même raison, lors de chaque étape du prétraitement, certaines grandeurs statistiques seront conservées sans raison physique évidente. La phase d'apprentissage du RN permettra de sélectionner les informations pertinentes.

2) Informations sur le type d'image

a - Le variogramme

Le variogramme est décrit en annexe (cf. ANNEXE C -A3 -).

La Figure 24 montre les variogrammes des images de type "ville" (Test014) et "champs" (Test000) pour différentes qualités de compression JPEG. Deux images différentes ont des variogrammes différents et la qualité de la compression n'influe que peu sur l'allure générale du variogramme. C'est donc un bon descripteur du type d'image. La compression a tendance à diminuer les valeurs du variogramme car, en coupant les hautes fréquences, elle augmente la corrélation entre des points voisins d'un même bloc. Pour cette même raison, la compression affecte plus le variogramme d'une image structurée (Test014) que celui d'une image très uniforme.

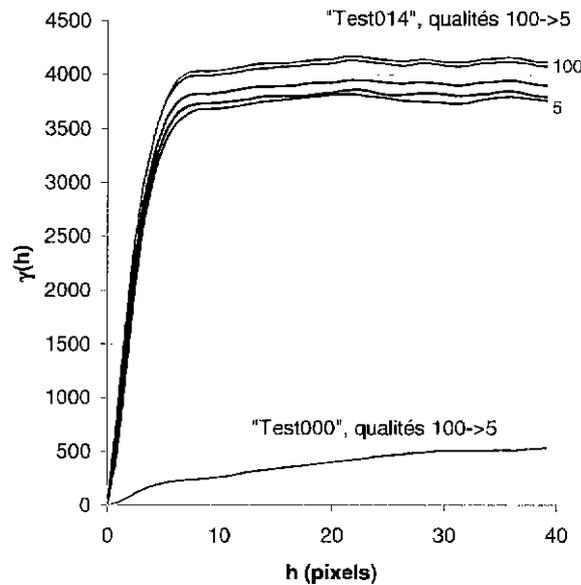


Figure 24 : Variogrammes d'une image de type "ville" (Test014) et "champs" (Test000) pour les qualités JPEG 100, 60, 40, 20, 10, 5.

Nous avons choisi une modélisation à quatre paramètres, du type :

$$\gamma(h) = A(1 - e^{-Bh}) + C(1 - e^{-Dh}) = E_1 + E_2$$

Cette modélisation est aussi décrite en annexe.

b - Autres statistiques

Toute grandeur statistique pouvant fournir une information sur le contenu de l'image est intéressante, à condition qu'elle ne soit pas modifiée par la compression. La Figure 25 montre que c'est le cas en ce qui concerne la moyenne et la variance. La légère fluctuation de la variance pour l'image Test014 avait déjà été observée sur l'asymptote du variogramme. On retiendra aussi les moments d'ordre 3 et 4 de l'image comme paramètres susceptibles de fournir une information complémentaire pour déterminer la qualité. L'étude de sensibilité (cf. Chapitre 3 :I-D -3)) donnera des informations sur la pertinence de ce choix.

Toutes ces grandeurs statistiques ne sont affectées par la compression que pour des facteurs de qualité JPEG faibles, c'est-à-dire de très forts taux de compression, elles donnent en général une information sur l'image parfaite.

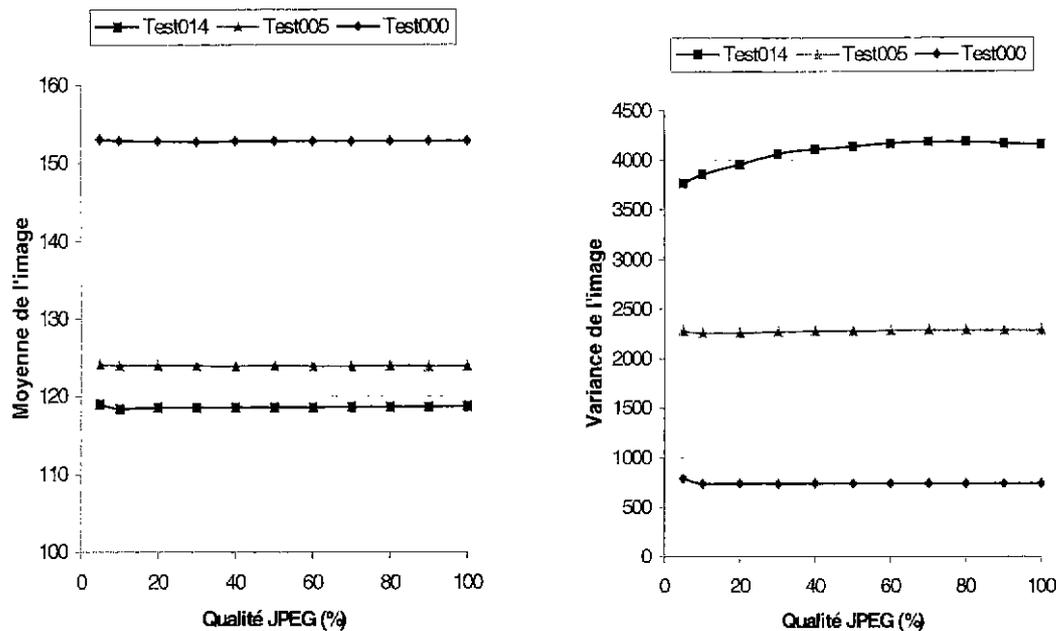


Figure 25 : Variations de la moyenne et de la variance d'une image en fonction de la qualité JPEG, selon le type d'image.

3) Le vecteur caractéristique

C'est l'ensemble des paramètres qui vont caractériser la dégradation et le type d'image pour servir d'entrée au RN.

Il est défini de manière à tenir compte des propriétés essentielles des artefacts de bloc ainsi que de celles de l'image. L'utilisation d'un RN permet lors de l'apprentissage de mesurer l'importance de l'une ou l'autre de ces composantes. Certaines d'entre elles, à la signification physique moins évidente, ont été retenues parce que l'apprentissage a montré qu'elles étaient significatives dans l'appréciation de la qualité. L'étude de sensibilité présentée au Chapitre 3 : I -D -3) permettra de déterminer quelles sont les composantes réellement importantes dans l'évaluation de la qualité et donc d'éliminer d'éventuelles redondances.

Le vecteur caractéristique est constitué des composantes suivantes :

- 0) Moyenne de l'image
- 1) Variance de l'image
- 2) Moment d'ordre 3 de l'image
- 3) Moment d'ordre 4 de l'image
- 4) Coefficient A du variogramme
- 5) Coefficient B du variogramme
- 6) Coefficient C du variogramme
- 7) Coefficient D du variogramme
- 8) Moyenne de la valeur absolue du Laplacien de l'image
- 9) Variance de la valeur absolue du Laplacien de l'image
- 10) Moment d'ordre 3 de la valeur absolue du Laplacien de l'image
- 11) Moment d'ordre 4 de la valeur absolue du Laplacien de l'image
- 12) Moyenne de V_{vertic}
- 13) Variance de V_{vertic}
- 14) Moment d'ordre 3 de V_{vertic}
- 15) Moment d'ordre 4 de V_{vertic}

- 16) Variance de V_{horiz}
- 17) Moment d'ordre 3 de V_{horiz}
- 18) Moment d'ordre 4 de V_{horiz}
- 19) Valeur de FF_8 sur $V(k)$
- 20) Valeur de FF_{16} sur $V(k)$
- 21) Valeur de $FF_{7,9}$ sur $V(k)$ (pour comparaison avec FF_8)
- 22) Valeur de FF_8 sur l'image
- 23) Valeur de $FF_{7,9}$ sur l'image (pour comparaison avec FF_8)
- 24) Moyenne de $V(k)$
- 25) Variance de $V(k)$
- 26) Moment d'ordre 3 de $V(k)$
- 27) Moment d'ordre 4 de $V(k)$

La moyenne de V_{horiz} est rigoureusement la même que celle de V_{vertic} , pour cette raison seule la deuxième a été retenue. En revanche, aucune relation évidente ne semble lier les autres composantes.

D - Résultats

Nous allons montrer que le vecteur caractéristique, tel qu'il est constitué, permet d'apprendre différentes notations de qualité des images JPEG.

Dans un premier temps, l'apprentissage a été réalisé de manière à retrouver l'indice de qualité JPEG fixé lors de la compression de l'image (cf. ANNEXE D -). Cet indice, choisi avant la compression, détermine l'importance de la quantification qui est réalisée sur l'image.

Dans un deuxième temps, l'apprentissage a été réalisé de manière à reproduire le comportement d'un modèle bivariant représentatif d'une évaluation subjective de la qualité : le modèle de S. Karunasekera et N. Kingsbury (modèle K&K) [KaK95].

Pour chacun de ces deux modèles, l'apprentissage a été effectué successivement sur les bases variées et spécifiques (cf. Chapitre 2 :III -B -). Les images sont comprimées avec des facteurs de qualité JPEG variant de 10 à 100. Pour le premier apprentissage, il s'agit de retrouver cette qualité à partir de l'image dégradée. Pour le deuxième apprentissage, les images comprimées sont notées par le modèle K&K et c'est sur cette notation que l'apprentissage est réalisé.

1) Apprentissage du facteur de qualité JPEG

L'intérêt de cet apprentissage est de montrer que l'on peut estimer un tel paramètre, fixé par l'utilisateur lors de la phase de compression et qui, d'une certaine manière, traduit la quantité d'information perdue pendant cette phase.

L'apprentissage est dans un premier temps réalisé sur la base d'images variées, puis testé à la fois sur des images variées et sur des images spécifiques. Il est ensuite réalisé sur la base d'images spécifiques pour vérifier que la précision des résultats en est améliorée.

a - Apprentissage sur la base d'images variées

Pour réaliser cet apprentissage, l'indice de qualité JPEG est normalisé sur l'intervalle $[0,10]$, toutes les entrées sont normalisées sur $[-1,1]$. La Figure 26 présente les résultats obtenus après apprentissage de l'indice de qualité sur la base d'images variées.

L'erreur quadratique moyenne (EQM) entre la qualité théorique de l'image et celle prédite par le RN est de 5,5 %. L'erreur maximale de 14 % est obtenue pour l'image Test000, qui est une image de champs, très uniforme. On constate aussi pour cette même image que l'erreur est plus importante pour des qualités moyennes que pour les extrêmes. Aucune explication n'est proposée pour expliquer ce phénomène. Le RN est un très bon interpolateur, son

comportement en extrapolation n'est pas toujours bien connu. Cette image est vraisemblablement mal située par rapport à la base d'apprentissage.

Pour mesurer la similitude entre la notation du RN et la note théorique, le coefficient de corrélation de Pearson a été calculé.

Si x_i et y_i sont respectivement les notes de qualité théorique et celles données par le modèle, ce coefficient vaut :

$$R = \frac{\sum_i x_i y_i - \frac{1}{N} \sum_i x_i \sum_i y_i}{\sqrt{\left[\sum_i x_i^2 - \frac{1}{N} \left(\sum_i x_i \right)^2 \right] \cdot \left[\sum_i y_i^2 - \frac{1}{N} \left(\sum_i y_i \right)^2 \right]}}$$

Comme le montre le Tableau 1, il existe toujours une très forte corrélation entre les indices de qualité JPEG théoriques et les indices calculés par le RN, preuve que la notation du RN est très semblable à la notation théorique : pour une même image, le RN arrive toujours à classer les images en fonction de leur qualité.

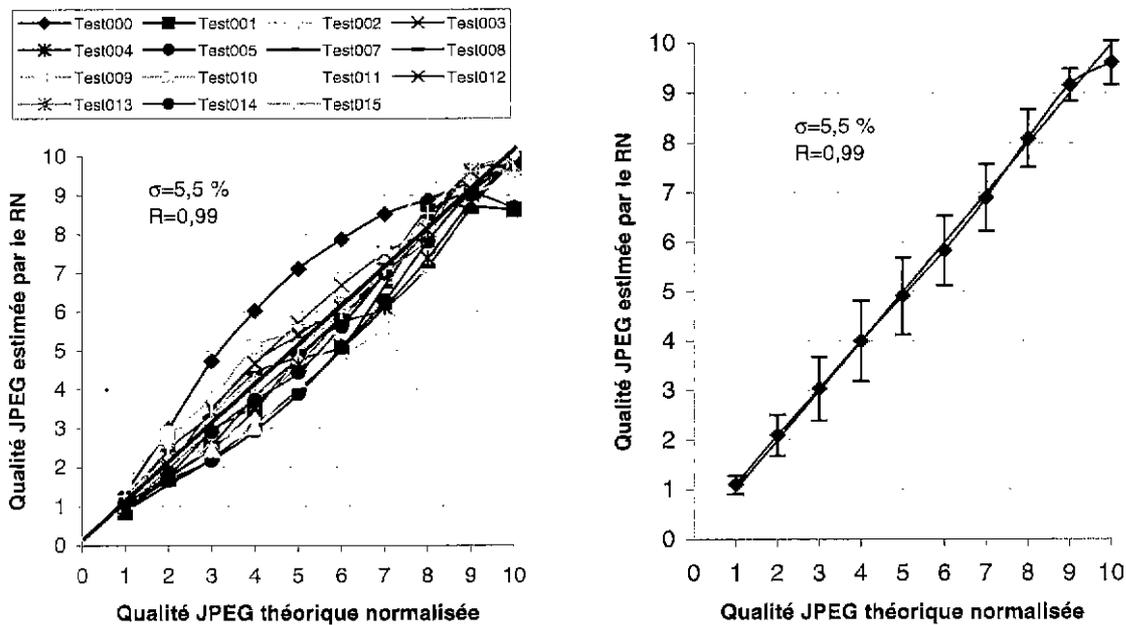


Figure 26 : Apprentissage de l'indice de qualité sur la base variée, test sur la base variée. Le graphique de gauche montre les résultats obtenus image par image pour des indices de qualité normalisés variant de 1 à 10, le graphique de droite présente les résultats moyens sur toutes les images pour ces mêmes indices de qualité. Les barres d'erreurs sont égales aux écarts quadratiques, de part et d'autre de la moyenne.

La vérification sur la base de validation spécifique montre une très grande stabilité du RN. L'erreur moyenne du RN par rapport à la qualité théorique reste stable autour de 4%, quel que soit le type d'image et quel que soit l'indice de qualité JPEG.

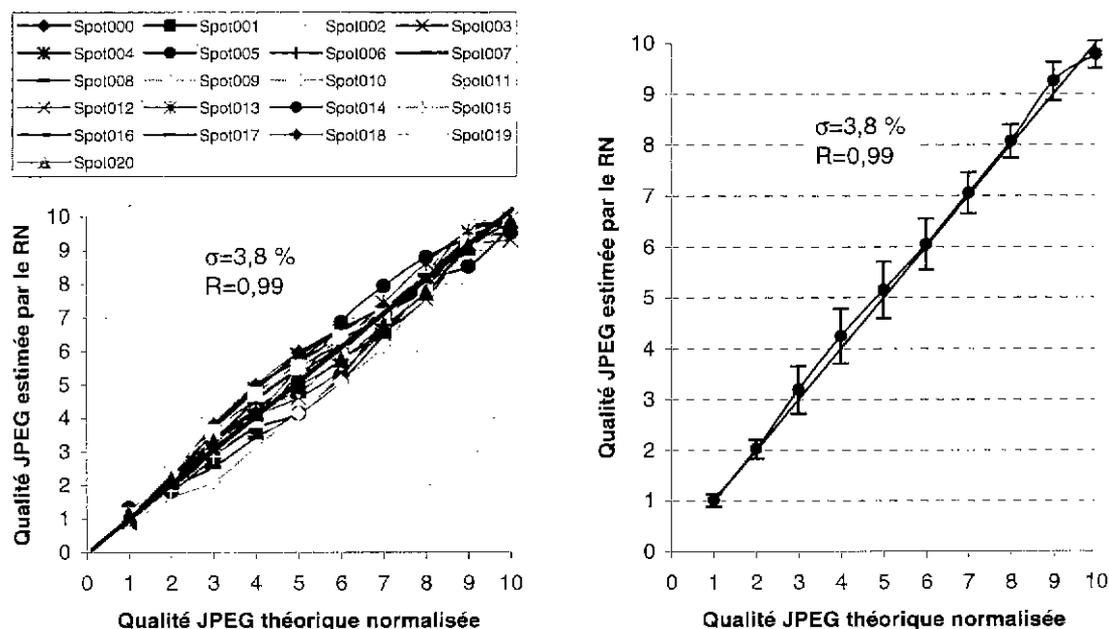


Figure 27 : Apprentissage de l'indice de qualité sur la base variée, test sur la base spécifique. Le graphique de gauche montre les résultats obtenus image par image pour des indices de qualité normalisés variant de 1 à 10, le graphique de droite présente les résultats moyens sur toutes les images pour ces mêmes indices de qualité. Les barres d'erreurs sont égales aux écarts quadratiques, de part et d'autre de la moyenne.

| Images SPOT | | |
|-------------|---------|-------|
| Image | EQM (%) | R |
| Spot000 | 3,8 | 0,995 |
| Spot001 | 1,9 | 0,998 |
| Spot002 | 2,5 | 0,997 |
| Spot003 | 3,2 | 0,997 |
| Spot004 | 4,9 | 0,991 |
| Spot005 | 4,3 | 0,994 |
| Spot006 | 3,5 | 0,996 |
| Spot007 | 2,8 | 0,997 |
| Spot008 | 5,0 | 0,991 |
| Spot009 | 6,4 | 0,991 |
| Spot010 | 3,0 | 0,997 |
| Spot011 | 3,9 | 0,995 |
| Spot012 | 3,7 | 0,996 |
| Spot013 | 3,5 | 0,997 |
| Spot014 | 5,2 | 0,989 |
| Spot015 | 1,5 | 1,000 |
| Spot016 | 4,9 | 0,993 |
| Spot017 | 4,4 | 0,996 |
| Spot018 | 5,8 | 0,992 |
| Spot019 | 4,0 | 0,996 |
| Spot020 | 1,9 | 0,999 |
| Moyenne | 3,8 | 0,995 |

| Images variées | | |
|----------------|---------|-------|
| Image | EQM (%) | R |
| Test000 | 13,9 | 0,963 |
| Test001 | 5,5 | 0,986 |
| Test002 | 7,6 | 0,985 |
| Test003 | 4,4 | 0,996 |
| Test004 | 4,9 | 0,995 |
| Test005 | 6,8 | 0,990 |
| Test006 | 4,1 | 0,980 |
| Test007 | 6,6 | 0,989 |
| Test008 | 4,8 | 0,991 |
| Test009 | 2,3 | 0,997 |
| Test010 | 6,0 | 0,995 |
| Test011 | 6,5 | 0,980 |
| Test012 | 3,4 | 0,997 |
| Test013 | 2,8 | 0,996 |
| Test014 | 5,1 | 0,987 |
| Test015 | 3,8 | 0,995 |
| Moyenne | 5,5 | 0,989 |

Tableau 1 : Statistiques concernant les résultats obtenus sur les bases de test dans le cadre de l'apprentissage de l'indice de qualité JPEG sur la base variée: coefficient de corrélation entre la notation théorique et celle du modèle univariant, erreur quadratique moyenne.

b - Apprentissage spécifique

L'apprentissage est maintenant réalisé sur la base spécifique d'images SPOT. La Figure 28 montre les résultats obtenus sur la base spécifique. On constate que le résultat image par image est amélioré, en moyenne, pour tous les indices de qualité. Néanmoins, lorsqu'une image est à la limite du contexte, c'est le cas pour l'image Spot005 qui est quasiment uniforme, le RN commet une erreur importante. Ce n'était pas le cas pour l'apprentissage réalisé sur la base variée. On a donc gagné en précision moyenne mais perdu en stabilité. On le constate nettement sur la Figure 29 : lorsque cet apprentissage spécifique est testé sur des images variées, l'évaluation de l'indice de qualité devient fantaisiste.

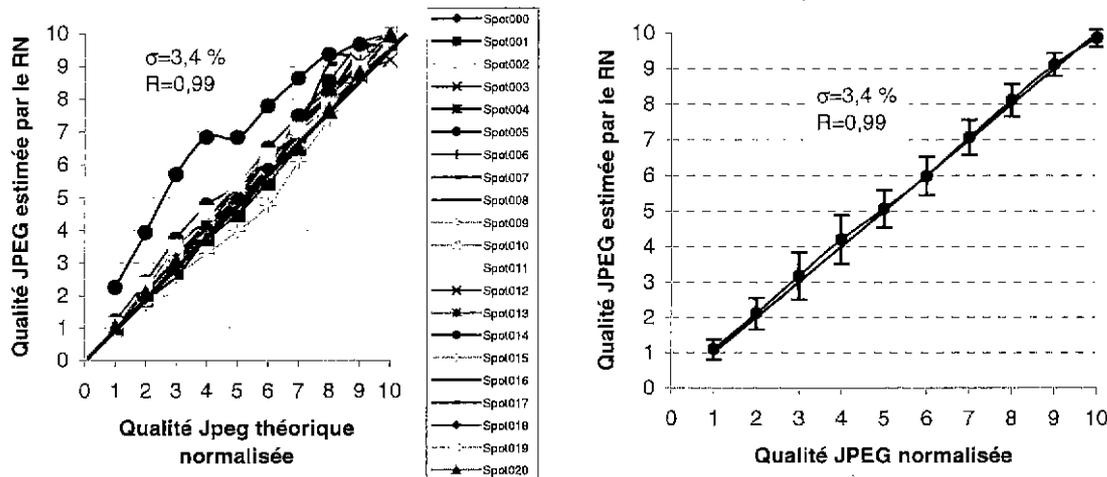


Figure 28 : Apprentissage de l'indice de qualité sur la base spécifique, test sur la base spécifique. Le graphique de gauche montre les résultats obtenus image par image pour des indices de qualité normalisés variant de 1 à 10, le graphique de droite présente les résultats moyens sur toutes les images pour ces mêmes indices de qualité. Les barres d'erreurs sont égales aux écarts quadratiques, de part et d'autre de la moyenne.

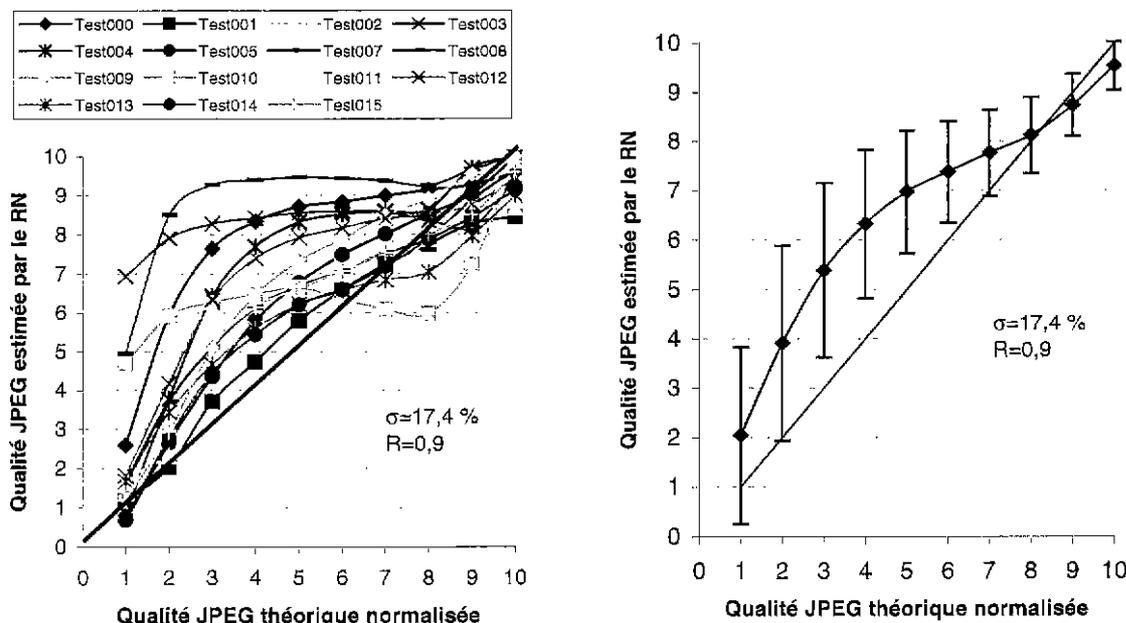


Figure 29 : Apprentissage de l'indice de qualité sur la base spécifique, test sur la base variée. Le graphique de gauche montre les résultats obtenus image par image pour des indices de qualité normalisés variant de 1 à 10, le graphique de droite présente les résultats moyens sur toutes les images pour ces mêmes indices de qualité. Les barres d'erreurs sont égales aux écarts quadratiques, de part et d'autre de la moyenne.

Le Tableau 2 détaille image par image, sur chacune des bases, les résultats précédents.

| Images SPOT | | | Images variées | | |
|-------------|---------|-------|----------------|---------|-------|
| Image | EQM (%) | R | Image | EQM (%) | R |
| Spot000 | 1,4 | 0,999 | Test000 | 29,3 | 0,822 |
| Spot001 | 2,6 | 0,998 | Test001 | 7,2 | 0,977 |
| Spot002 | 1,6 | 0,999 | Test002 | 11,5 | 0,963 |
| Spot003 | 0,8 | 1,000 | Test003 | 37,5 | 0,742 |
| Spot004 | 1,1 | 1,000 | Test004 | 22,2 | 0,882 |
| Spot005 | 17,9 | 0,961 | Test005 | 11,1 | 0,967 |
| Spot006 | 1,9 | 0,999 | Test006 | 9,3 | 0,936 |
| Spot007 | 2,0 | 0,998 | Test007 | 12,9 | 0,964 |
| Spot008 | 1,8 | 0,999 | Test008 | 40,7 | 0,676 |
| Spot009 | 2,2 | 0,998 | Test009 | 15,4 | 0,957 |
| Spot010 | 1,3 | 0,999 | Test010 | 23,8 | 0,728 |
| Spot011 | 3,5 | 0,997 | Test011 | 5,1 | 0,993 |
| Spot012 | 3,7 | 0,999 | Test012 | 20,9 | 0,887 |
| Spot013 | 1,6 | 1,000 | Test013 | 11,3 | 0,963 |
| Spot014 | 5,3 | 0,997 | Test014 | 8,8 | 0,967 |
| Spot015 | 6,7 | 0,991 | Test015 | 12,0 | 0,953 |
| Spot016 | 4,6 | 0,991 | Moyenne | 17,4 | 0,899 |
| Spot017 | 2,5 | 0,999 | | | |
| Spot018 | 1,9 | 0,998 | | | |
| Spot019 | 3,1 | 0,998 | | | |
| Spot020 | 3,2 | 0,997 | | | |
| Moyenne | 3,4 | 0,996 | | | |

Tableau 2 : Statistiques concernant les résultats obtenus sur les bases de test dans le cadre de l'apprentissage de l'indice de qualité JPEG sur la base spécifique: coefficient de corrélation entre la notation théorique et celle du modèle univariant, erreur quadratique moyenne.

2) Apprentissage d'une notation de qualité visuelle : modèle de Karunasekera et Kingsbury [KaK93, KaK94, KaK95].

La notation de l'image envisagée maintenant est beaucoup plus représentative d'une notation subjective que l'indice de qualité JPEG. Nous nous proposons de retrouver la note de qualité visuelle d'une image JPEG prédite par le modèle bivariant de S. Karunasekera et N. Kingsbury (cf. Chapitre 1 :III -B -2)). Dans ce modèle, la note de qualité est fonction du temps que l'observateur met pour différencier une image comprimée de l'image parfaite. Cette note n'a plus rien à voir avec l'indice de qualité utilisé précédemment. Notre objectif est de montrer que le modèle peut reproduire de façon univariante une telle notation.

Comme précédemment, nous avons réalisé un apprentissage sur la base d'images variées, puis sur la base d'images spécifiques.

a - Apprentissage sur la base d'images variées

La Figure 30 et le Tableau 3 présentent les résultats obtenus sur des images variées après apprentissage sur la base d'images variées. L'erreur moyenne sur la base variée est inférieure à 6 %, avec un pic de 10 % pour l'image Test014 qui est une ville. Sur la base spécifique les résultats sont très stables et l'erreur moyenne inférieure à 4 %.

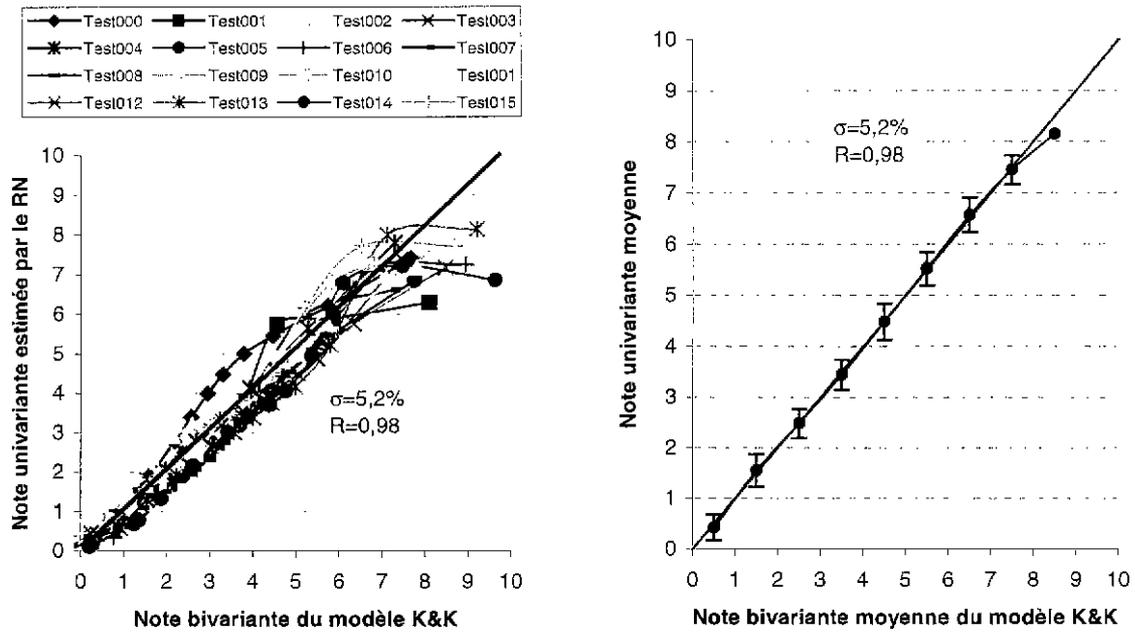


Figure 30 : Apprentissage de la notation de qualité visuelle sur la base variée, test sur la base variée. Le graphique de gauche montre les résultats obtenus image par image pour des notes de qualité variant de 1 à 10, le graphique de droite présente les résultats moyens sur toutes les images pour ces mêmes indices de qualité. Les barres d'erreurs sont égales aux écarts quadratiques, de part et d'autre de la moyenne.

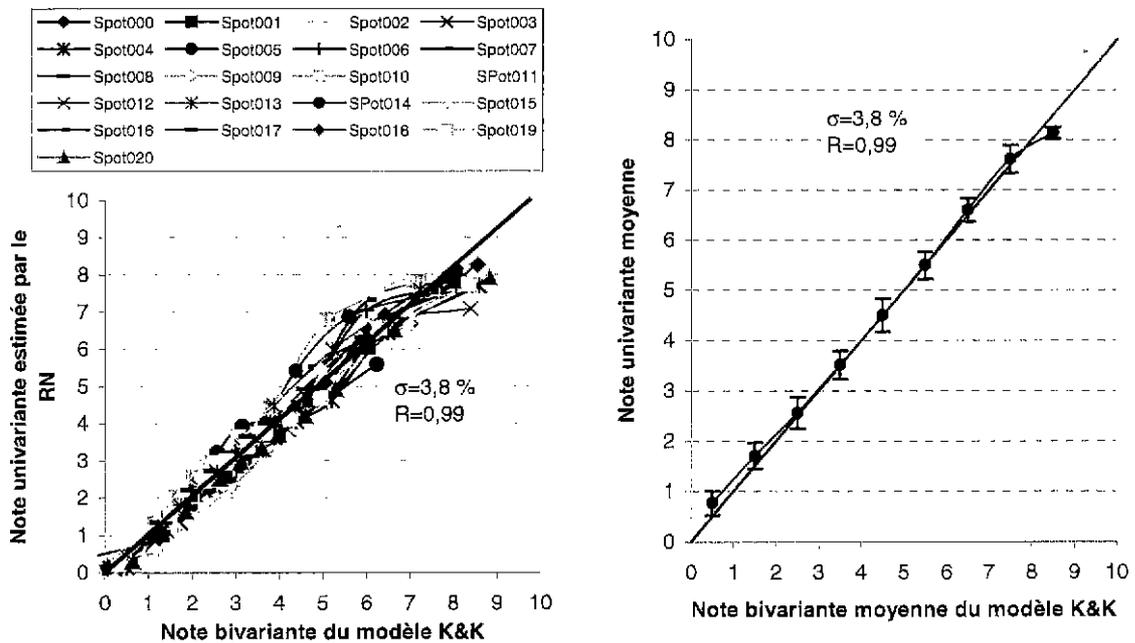


Figure 31 : Apprentissage de la notation de qualité visuelle sur la base variée, test sur la base spécifique. Le graphique de gauche montre les résultats obtenus image par image pour des notes de qualité variant de 1 à 10, le graphique de droite présente les résultats moyens sur toutes les images pour ces mêmes indices de qualité. Les barres d'erreurs sont égales aux écarts quadratiques, de part et d'autre de la moyenne.

Le Tableau 3 présente les résultats obtenus image par image.

| Images SPOT | | | Images variées | | |
|-------------|---------|-------|----------------|---------|-------|
| Image | EQM (%) | R | Image | EQM (%) | R |
| Spot000 | 2,3 | 0,997 | Test000 | 7,9 | 0,971 |
| Spot001 | 1,6 | 0,998 | Test001 | 8,0 | 0,934 |
| Spot002 | 4,5 | 0,989 | Test002 | 2,6 | 0,996 |
| Spot003 | 1,8 | 0,997 | Test003 | 2,8 | 0,994 |
| Spot004 | 4,1 | 0,995 | Test004 | 2,7 | 0,997 |
| Spot005 | 4,7 | 0,967 | Test005 | 5,0 | 0,997 |
| Spot006 | 3,7 | 0,987 | Test006 | 5,9 | 0,975 |
| Spot007 | 4,3 | 0,985 | Test007 | 4,9 | 0,978 |
| Spot008 | 3,3 | 0,992 | Test008 | 2,0 | 0,995 |
| Spot009 | 4,5 | 0,984 | Test009 | 3,5 | 0,989 |
| Spot010 | 7,0 | 0,989 | Test010 | 2,9 | 0,991 |
| Spot011 | 5,9 | 0,967 | Test011 | 7,7 | 0,958 |
| Spot012 | 4,8 | 0,986 | Test012 | 5,1 | 0,998 |
| Spot013 | 4,0 | 0,994 | Test013 | 6,6 | 0,977 |
| Spot014 | 5,5 | 0,979 | Test014 | 10,1 | 0,935 |
| Spot015 | 2,7 | 0,996 | Test015 | 5,9 | 0,968 |
| Spot016 | 3,2 | 0,992 | Moyenne | 5,2 | 0,978 |
| Spot017 | 2,9 | 0,998 | | | |
| Spot018 | 2,4 | 0,995 | | | |
| Spot019 | 2,6 | 0,995 | | | |
| Spot020 | 3,9 | 0,997 | | | |
| Moyenne | 3,8 | 0,989 | | | |

Tableau 3 : Statistiques concernant les résultats obtenus sur la base de test dans le cadre de l'apprentissage, sur la base variée, de l'indice de qualité visuelle : coefficient de corrélation entre la notation théorique et celle du modèle univariant, erreur quadratique moyenne.

b - Apprentissage sur la base spécifique

L'apprentissage sur la base spécifique permet, comme précédemment, d'améliorer les résultats obtenus pour des images du même type (Figure 32). Néanmoins, on ne constate pas d'erreur importante pour l'image Spot005 comme c'était le cas pour l'apprentissage de l'indice de qualité JPEG. On ne constate pas non plus d'estimation fantaisiste de qualité sur les images variées (Figure 33) même si l'erreur moyenne augmente légèrement.

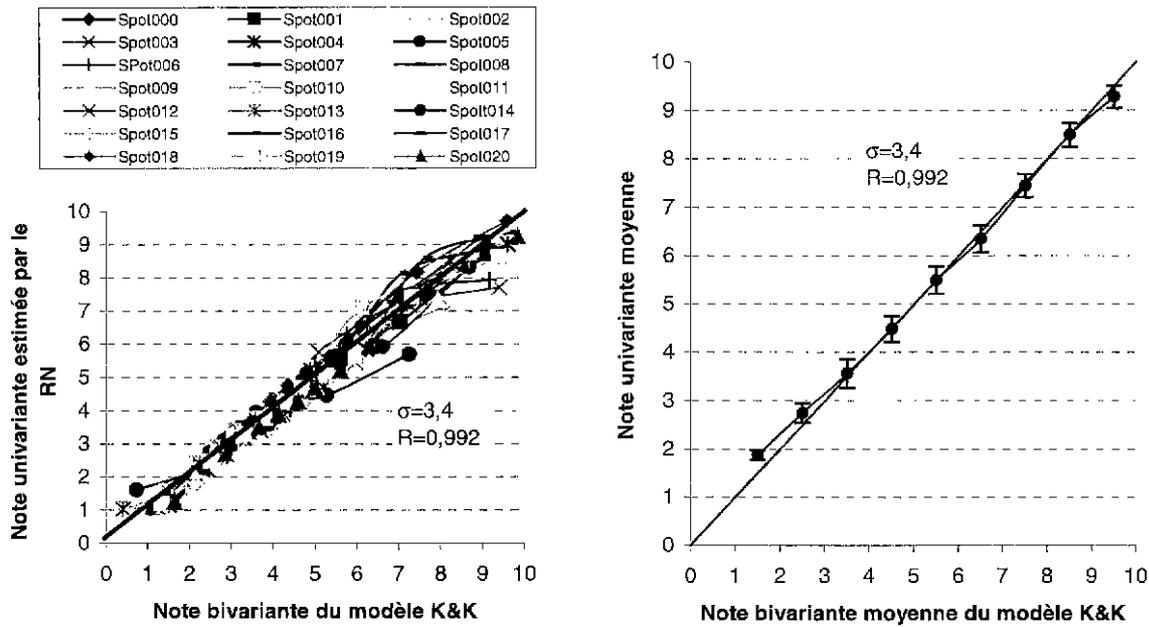


Figure 32 : Apprentissage de la notation de qualité visuelle sur la base spécifique, test sur la base spécifique. Le graphique de gauche montre les résultats obtenus image par image pour des notes de qualité variant de 1 à 10, le graphique de droite présente les résultats moyens sur toutes les images pour ces mêmes indices de qualité. Les barres d'erreurs sont égales aux écarts quadratiques, de part et d'autre de la moyenne.

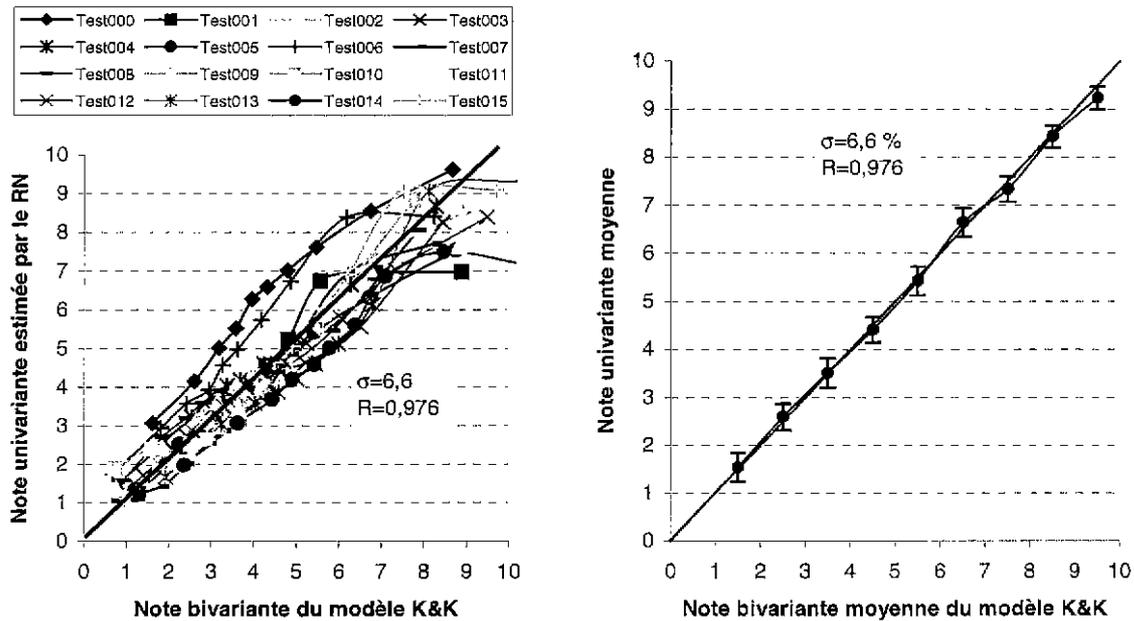


Figure 33 : Apprentissage de la notation de qualité visuelle sur la base spécifique, test sur la base variée. Le graphique de gauche montre les résultats obtenus image par image pour des notes de qualité variant de 1 à 10, le graphique de droite présente les résultats moyens sur toutes les images pour ces mêmes indices de qualité. Les barres d'erreurs sont égales aux écarts quadratiques, de part et d'autre de la moyenne.

| Images SPOT | | |
|-------------|---------|-------|
| Image | EQM (%) | R |
| Spot000 | 2,2 | 0,998 |
| Spot001 | 1,9 | 0,998 |
| Spot002 | 3,4 | 0,992 |
| Spot003 | 3,3 | 0,990 |
| Spot004 | 2,4 | 0,997 |
| Spot005 | 6,5 | 0,974 |
| Spot006 | 4,8 | 0,978 |
| Spot007 | 3,7 | 0,988 |
| Spot008 | 2,2 | 0,995 |
| Spot009 | 4,1 | 0,993 |
| Spot010 | 4,4 | 0,993 |
| Spot011 | 4,4 | 0,978 |
| Spot012 | 5,9 | 0,988 |
| Spot013 | 3,2 | 0,997 |
| Spot014 | 2,7 | 0,994 |
| Spot015 | 1,6 | 0,998 |
| Spot016 | 3,6 | 0,989 |
| Spot017 | 1,8 | 0,999 |
| Spot018 | 3,9 | 0,996 |
| Spot019 | 2,7 | 0,996 |
| Spot020 | 3,2 | 0,998 |
| Moyenne | 3,4 | 0,992 |

| Images variées | | |
|----------------|---------|-------|
| Image | EQM (%) | R |
| Test000 | 18,9 | 0,978 |
| Test001 | 7,3 | 0,938 |
| Test002 | 2,2 | 0,996 |
| Test003 | 4,3 | 0,983 |
| Test004 | 5,5 | 0,998 |
| Test005 | 4,2 | 0,993 |
| Test006 | 5,6 | 0,972 |
| Test007 | 4,4 | 0,982 |
| Test008 | 5,7 | 0,986 |
| Test009 | 2,1 | 0,996 |
| Test010 | 6,2 | 0,957 |
| Test011 | 8,9 | 0,948 |
| Test012 | 4,0 | 0,997 |
| Test013 | 7,0 | 0,974 |
| Test014 | 12,9 | 0,939 |
| Test015 | 6,1 | 0,975 |
| Moyenne | 6,6 | 0,976 |

Tableau 4 : Statistiques concernant les résultats obtenus sur la base de test dans le cadre de l'apprentissage, sur la base spécifique, de l'indice de qualité visuelle : coefficient de corrélation entre la notation théorique et celle du modèle univariant, erreur quadratique moyenne.

Cette différence de comportement est, en fait, facilement explicable. Lorsque l'indice de qualité JPEG varie, la qualité visuelle de l'image varie plus ou moins selon le type d'image considérée. En particulier, pour un même indice de qualité JPEG, la qualité visuelle d'une image relativement uniforme est moins bonne que celle d'une image très structurée et pour deux indices de qualité proches, l'appréciation visuelle peut être quasiment identique (Figure 34). Cette dernière mesure de qualité, moins sensible que la première, est plus facilement approchée.

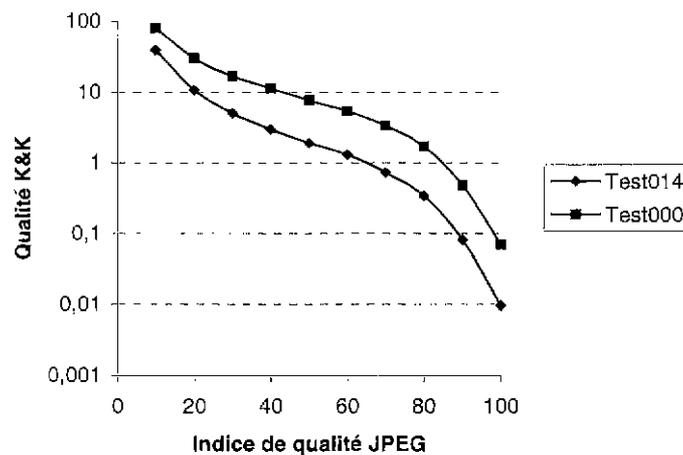


Figure 34 : Variation de la notation de qualité visuelle en fonction de l'indice de qualité JPEG.

3) Etude de sensibilité - Réduction du vecteur caractéristique

L'étude de sensibilité est réalisée de la manière décrite au Chapitre 2 :III -C -. Rappelons que la sensibilité vis-à-vis d'une composante représente la déviation maximale que peut introduire cette composante, seule, dans l'estimation de la qualité.

a - Sensibilité de l'apprentissage de l'indice de qualité JPEG

La Figure 35 montre les sensibilités vis-à-vis des différentes composantes pour 3 images de types différents et en moyenne sur la base de test.

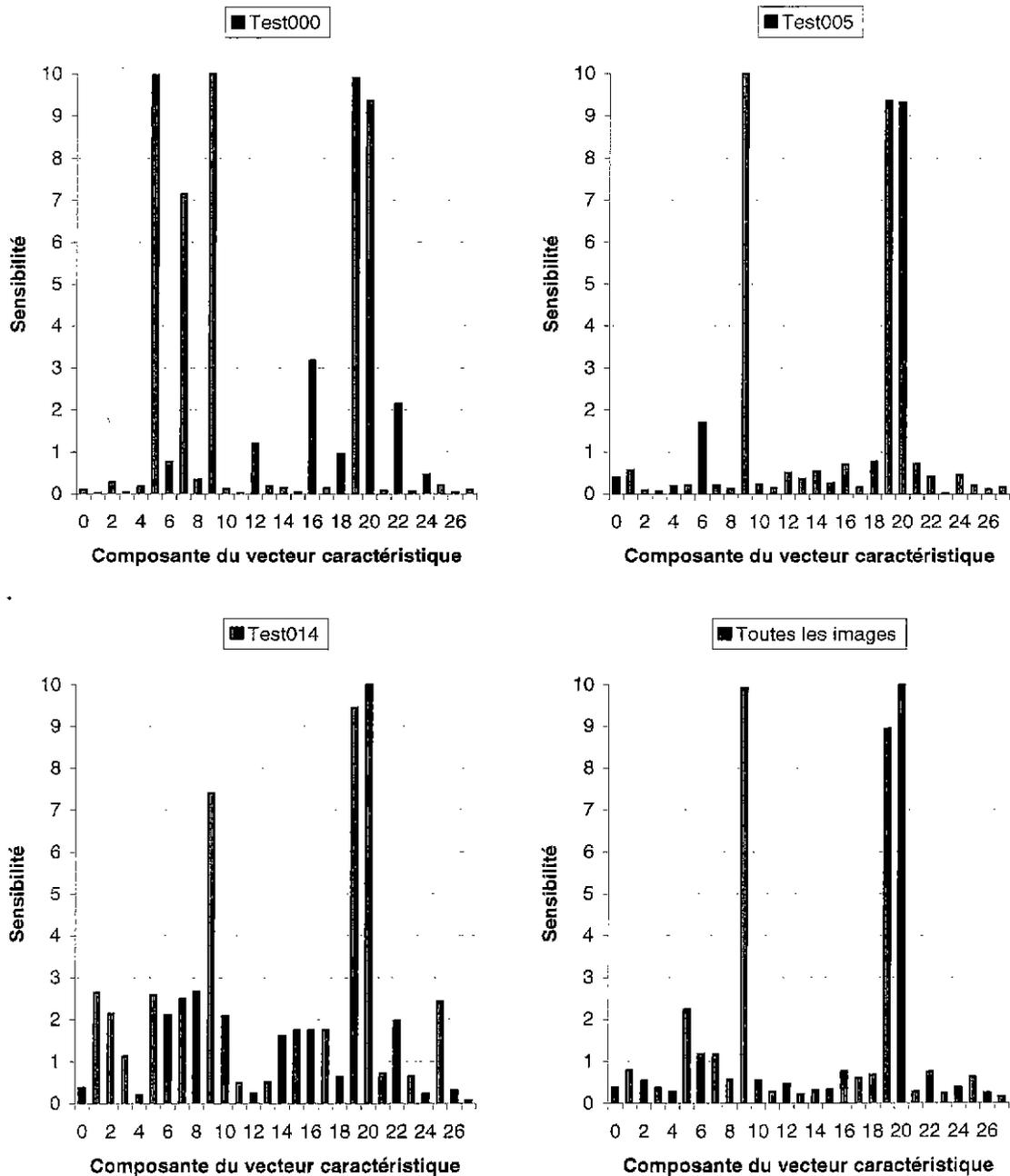


Figure 35 : Sensibilité vis-à-vis des composantes dans le cadre de l'apprentissage de l'indice de qualité JPEG.

On constate que les différentes entrées prennent des importances différentes selon le type d'image. Pour évaluer la qualité de la plupart des images, seuls 3 ou 4 paramètres semblent importants. Le coefficient FF_8 est utilisé en priorité pour estimer la qualité de l'image, pondérée par des informations plus spécifiques concernant le type d'image.

Pour apprendre la qualité d'une image uniforme (Test000), une grande importance est accordée aux coefficients B et C du variogramme qui permettent d'identifier la texture de l'image. En revanche, pour une image très structurée (Test014), beaucoup de paramètres semblent influencer la décision. Quelques entrées semblent ne jamais avoir d'influence, c'est le cas des moments d'ordre 3 et 4 de $V(k)$.

Réduction du vecteur caractéristique :

Des apprentissages ont été réalisés en supprimant les entrées d'importance moindre pour en vérifier la non importance.

Si dans un premier temps, on ne retient que les composantes d'importances supérieures à 20 sur une base 100, l'apprentissage est très mauvais, l'erreur quadratique moyenne sur l'ensemble des images est d'environ 40-45 % selon les configurations de RN choisies. Si l'on ne considère maintenant que les entrées d'importances relatives supérieures à 10/100, cette erreur reste très élevée, autour de 30-35 %.

Dans un deuxième temps, nous avons émis l'hypothèse que l'orientation des images n'avait pas d'influence sur la note de qualité pour supprimer du vecteur caractéristique les composantes correspondant à V_{horiz} et V_{vertic} , devenues redondantes avec $V(k)$. L'erreur quadratique moyenne sur l'ensemble des images se dégrade pour atteindre 9,2 % dans cette configuration et l'étude de sensibilité ne montre pas de hausse de sensibilité vis-à-vis de $V(k)$. Le fait que V_{horiz} et V_{vertic} aient une importance, même faible, peut être expliqué au regard de la base d'apprentissage qui contient un grand nombre d'images de scènes extérieures, de façades ou de visages, dans lesquelles l'une des directions horizontale ou verticale est privilégiée. En effet, les artefacts de blocs ont statistiquement la même composante sur V_{horiz} et V_{vertic} . Les différences entre ces deux vecteurs sont dues aux composantes verticales et horizontales de l'image autres que les blocs. Ces composantes sont susceptibles de perturber l'évaluation de l'importance du phénomène de bloc. Elles apparaissent entre les pics sur V_{horiz} et V_{vertic} et peuvent permettre au RN de faire la différence entre la contribution des blocs et celle de l'image sur le coefficient FF_8 .

Finalement, nous avons tenté de conserver V_{horiz} et V_{vertic} pour supprimer $V(k)$, qui semble redondant. L'erreur quadratique moyenne sur l'ensemble des images est de 6,5 % dans cette configuration et la sensibilité vis-à-vis des composantes relatives à V_{horiz} et V_{vertic} n'augmente pas de manière notable. Il semble que $V(k)$, représentatif de l'importance globale des artefacts de blocs sur l'image, apporte une information complémentaire, mais d'importance moindre, pour déterminer la qualité de l'image.

b - Sensibilité de l'apprentissage de la qualité visuelle

Une étude de sensibilité identique à la précédente a été réalisée. La Figure 36 en présente les résultats. Dans ce cas, chaque entrée intervient à un moment ou à un autre. Comme précédemment, la donnée la plus importante est le coefficient FF_8 pour tous les types d'image.

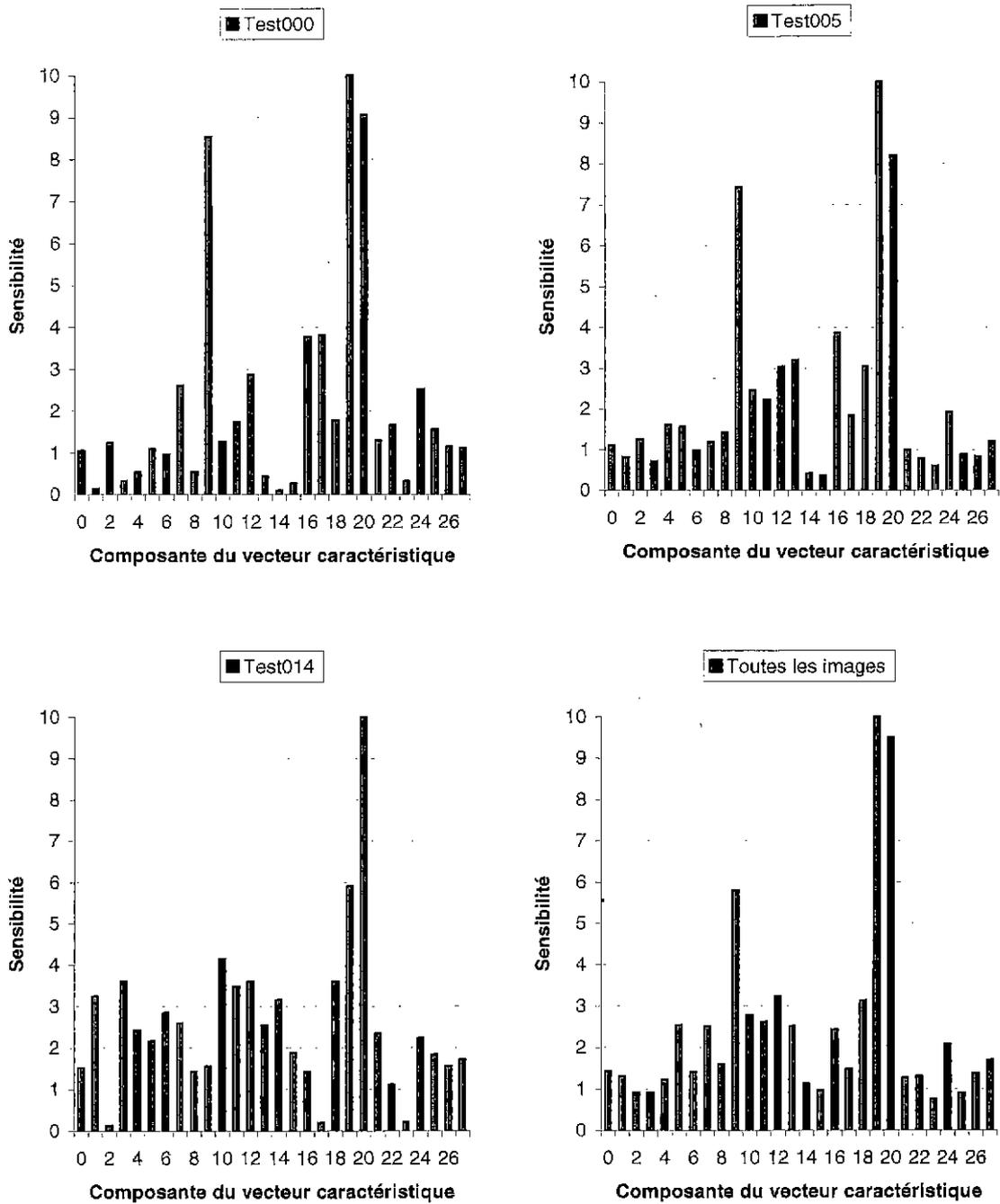


Figure 36 : Sensibilité vis-à-vis des composantes dans le cadre de l'apprentissage de la notation de qualité visuelle.

Des apprentissages ont été réalisés en ne conservant que les composantes qui semblent avoir une grande importance physique. Le Tableau 5 montre que seul un faible nombre d'entrées a une réelle importance sur l'apprentissage, mais chacune apporte sa contribution pour améliorer la précision. Contrairement à l'apprentissage de l'indice de qualité JPEG, la donnée du coefficient FF_8 associée à quelques informations sur le type d'image suffisent pour prédire la qualité visuelle avec la précision très acceptable de 7%. L'ajout des autres composantes permet d'améliorer la précision petit à petit.

| Entrées conservées pour l'apprentissage | Coefficient de corrélation R de Pearson | EQM moyenne (%) sur la base de test |
|---|---|-------------------------------------|
| Sensibilité > 30/100 | 0,974 | 7,2 |
| Sensibilité > 20/100 | 0,975 | 6,8 |
| Sensibilité > 10/100 | 0,976 | 5,7 |

Tableau 5 : Influence de la suppression de certaines composantes du vecteur caractéristique sur la qualité de l'apprentissage.

La différence de comportement entre les deux modèles semble être liée à leur nature. L'œil ne peut pas différencier tous les indices de qualité JPEG parce que sa sensibilité n'est pas assez importante. Rappelons (cf. ANNEXE D -) que le facteur de qualité n'est pas normalisé d'un outil de compression à un autre. Il est juste une fonction monotone du coefficient de quantification, calibrée le plus souvent sur des observations, mais différemment et plus ou moins sérieusement d'un outil à un autre.

La note donnée par le modèle visuel pour plusieurs indices de qualité différents mais proches peut être sensiblement identique. De ce fait elle est plus simple à approcher et nécessite moins de paramètres.

E - Conclusion

La méthode univariante permet de prévoir, avec une précision de 5 à 6 %, la qualité d'une image comprimée selon la norme JPEG qui aurait été donnée par un modèle bivariant quel qu'il soit, après apprentissage sur ce modèle. L'étude de sensibilité réalisée permet de montrer que toutes les composantes qui ont été retenues pour caractériser la dégradation ne sont pas forcément importantes pour apprendre un comportement de notation donné. En revanche, toutes sont utiles pour reproduire le comportement de l'un ou l'autre des deux modèles présentés. On peut raisonnablement supposer que tout autre information permettant de décrire les propriétés de l'image ou de la dégradation serait susceptible d'améliorer la précision de la notation univariante.

II - Etude de la qualité des images floues – Application à l'évaluation du défaut de mise au point du satellite SPOT 1

A - Introduction

La méthode neuronale univariante est appliquée ici à des images rendues floues par un défaut de mise au point de l'instrument de prise de vue. L'objectif de cette étude est de montrer qu'elle peut permettre d'évaluer certains paramètres caractéristiques du système d'acquisition de l'image à l'origine de cette dégradation. Pour un instrument satellitaire, par exemple, il peut s'agir de quantifier le défaut de mise au point en unité de longueur.

Un instrument de prise de vue peut ne pas être parfaitement focalisé lors de sa mise en route : Ce fut le cas pour l'instrument HRV2 de SPOT1 ou les deux instruments de SPOT 3. Pour le cas particulier des satellites SPOT, qui disposent de deux instruments identiques HRV1 et HRV2, une étude comparative des deux instruments peut permettre d'évaluer l'importance du défaut de mise au point de l'un d'entre eux [LeM95]. Dans le cas, le plus fréquent, d'un instrument unique, cette évaluation devient très difficile.

L'étude est dans un premier temps réalisée sur des images dégradées par un modèle de flou. Cette étape préliminaire doit servir non seulement à valider la méthode, mais aussi à trouver le vecteur caractéristique (cf. Chapitre 2 :III -A -) qui permettra de déterminer les propriétés physiques du défaut de mise au point. Ce modèle est un modèle réaliste simulant un flou géométrique dont le paramètre caractéristique, le pas de mise au point, est celui auquel on a accès sur la plupart des instruments.

Dans un deuxième temps, la méthode est appliquée pour évaluer le défaut de mise au point de l'instrument HRV2 du satellite SPOT 1. Ce défaut avait déjà été constaté et évalué de façon bivariante, par comparaison des FTM des deux instruments, lors d'une campagne menée en 1994.

Dans la suite, nous emploierons de manière équivalente les termes « refocalisation » pour « mise au point » et « défocalisation » pour « défaut de mise au point ».

B - Une première approche : utilisation d'un modèle de flou

1) Le modèle de flou géométrique

Pour préparer l'étude de la défocalisation réelle du satellite SPOT 1, nous avons utilisé un modèle de flou géométrique, caractérisé par l'écart au plan de bonne focalisation, mesuré en nombre de pas de mise au point, p , avec lequel l'image est acquise.

Précisons néanmoins que les images sur lesquelles est appliqué ce modèle sont des images du satellite SPOT, non perturbées par un défaut de mise au point, mais déjà bruitées, filtrées et échantillonnées par l'instrument de prise de vue. Leur spectre peut éventuellement déjà être replié. Dans le cas réel les bruits et l'échantillonnage viennent après le défaut de mise au point.

Le modèle utilisé est un filtre à réponse impulsionnelle finie (RIF) et phase nulle, conçu par un échantillonnage en fréquence de la FTM. L'échantillonnage est le même que celui de l'image.

Ce modèle repose sur l'approximation géométrique qu'un décalage du plan image par rapport au plan de bonne focalisation se traduit par une convolution de l'image avec un disque uniforme qui est la tâche (étalement du point) ainsi générée. On suppose aussi que le phénomène de diffraction reste négligeable vis-à-vis de la défocalisation.

La position du plan image par rapport à la position de bonne focalisation est mesurée en nombre de pas, valeur entière signée selon le sens du déplacement. Un pas correspond à un déplacement, toujours de même distance, du plan image.

L'étalement du point peut être exprimée en unité métrique (diamètre réel δ) ou angulaire (diamètre apparent vu du centre optique i). Ces notations sont définies sur la Figure 37. Pour les instruments HRV du satellite SPOT, les caractéristiques sont les suivantes :

- focale : $F=1,082$ m
- diamètre de la pupille : $D=33$ cm
- nombre d'ouverture : $N=3,3$
- pas : $\Delta=10$ μm

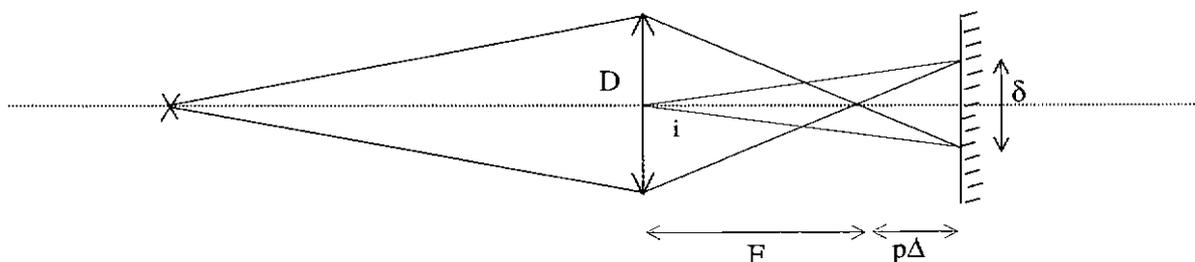


Figure 37 : Flou géométrique.

On a de manière évidente :

$$\delta = \frac{D \cdot p\Delta}{F} = \frac{p\Delta}{N} \quad \text{avec} \quad N = \frac{F}{D}$$

Et :

$$i \approx \frac{\delta}{(F + p\Delta)} \approx \frac{p\Delta}{FN} \quad \text{car} \quad p\Delta \ll F$$

Le disque uniforme d_p par lequel on convolue l'image est donc défini en fonction de i et de p par :

$$d_p(r_a) = \begin{cases} 1 & \text{si } r_a \leq \frac{i}{2} \\ \pi \left(\frac{i}{2}\right)^2 & \\ 0 & \text{si } r_a > \frac{i}{2} \end{cases}$$

où r_a désigne le rayon apparent vu du centre optique.

La FTM du modèle de défocalisation est donc, en posant $f_r = \sqrt{u^2 + v^2}$:

$$D_p(f_r) = \frac{2 \cdot J_1\left(2\pi f_r \frac{i}{2}\right)}{2\pi f_r \frac{i}{2}} = \frac{2 \cdot J_1(\alpha p f_r)}{\alpha p f_r} \quad \text{avec} \quad \alpha = \frac{\pi\Delta}{NF}$$

où $J_1(x)$ est la fonction de Bessel J d'ordre 1.

Pour SPOT1 on obtient $\alpha=8,86 \cdot 10^{-6}$ rad. La fréquence d'échantillonnage est $f_e = 83,2 \text{ kc/rad}$.

Si f_r est normalisée, on trouve $\alpha=0,732$.

La Figure 38 montre la fonction de transfert du modèle de défocalisation pour plusieurs valeurs du nombre de pas de défocalisation.

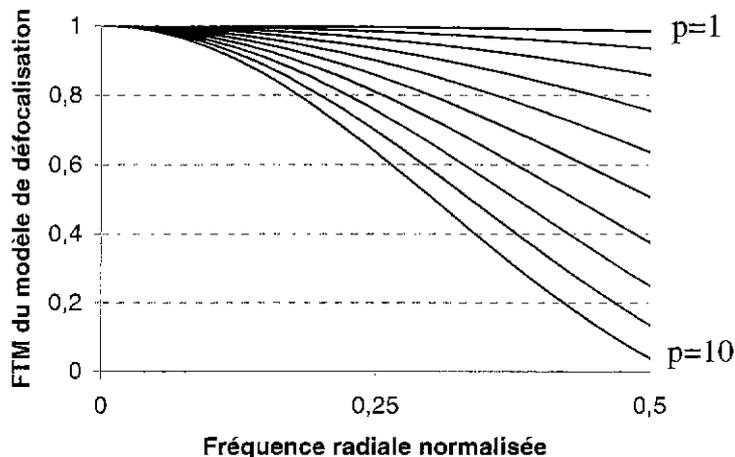


Figure 38 : Fonctions de transfert du modèle de défocalisation pour un nombre de pas variant de 1 à 10.

La Figure 39 montre la simulation d'une image SPOT défocalisée avec ce modèle pour des nombre de pas de défocalisation variant de 0 à 10.

2) Recherche des paramètres caractéristiques

La coupure des hautes fréquences pour rendre une image floue provoque un étalement de la tâche image sur les pixels voisins. Le flou est d'autant plus prononcé que le diamètre de cet étalement est important. Il est logique de rechercher les caractéristiques de cet étalement dans les propriétés fréquentielles et statistiques de l'image.

Nous ne représenterons les variations des paramètres caractéristiques que pour les images Test000, Test005 et Test014, qui sont respectivement une image uniforme, mixte et structurée.

i - Transformée de Fourier de l'image

Les propriétés fréquentielles de l'image ont été étudiées sur la transformée de Fourier (FFT, Fast Fourier Transform). Etant donné que le modèle de flou est à phase nulle, nous ne nous intéresserons pas à la phase de la FFT mais à son module uniquement.

On calcule dans un premier temps la FFT bidimensionnelle de l'image floue :

$$IM_{floue}(u, v) = FFT[Im_{floue}(x, y)]$$

Cette transformée de Fourier est le produit de la FFT de l'image originale par la FFT du filtre générateur de flou :

$$IM_{floue}(u, v) = IM_{originale}(u, v) \cdot Filtre(u, v)$$

La FFT filtre est une fonction radiale et la FFT de l'image originale est aussi, statistiquement, une fonction radiale. La FFT de l'image floue est donc aussi une fonction radiale. Il paraît dans un premier temps logique de simplifier le problème en ne nous intéressant qu'à la fonction à une dimension $IM(f_r)$ où $f_r = \sqrt{u^2 + v^2}$ et pour $f_r \in [0; f_e/2]$.

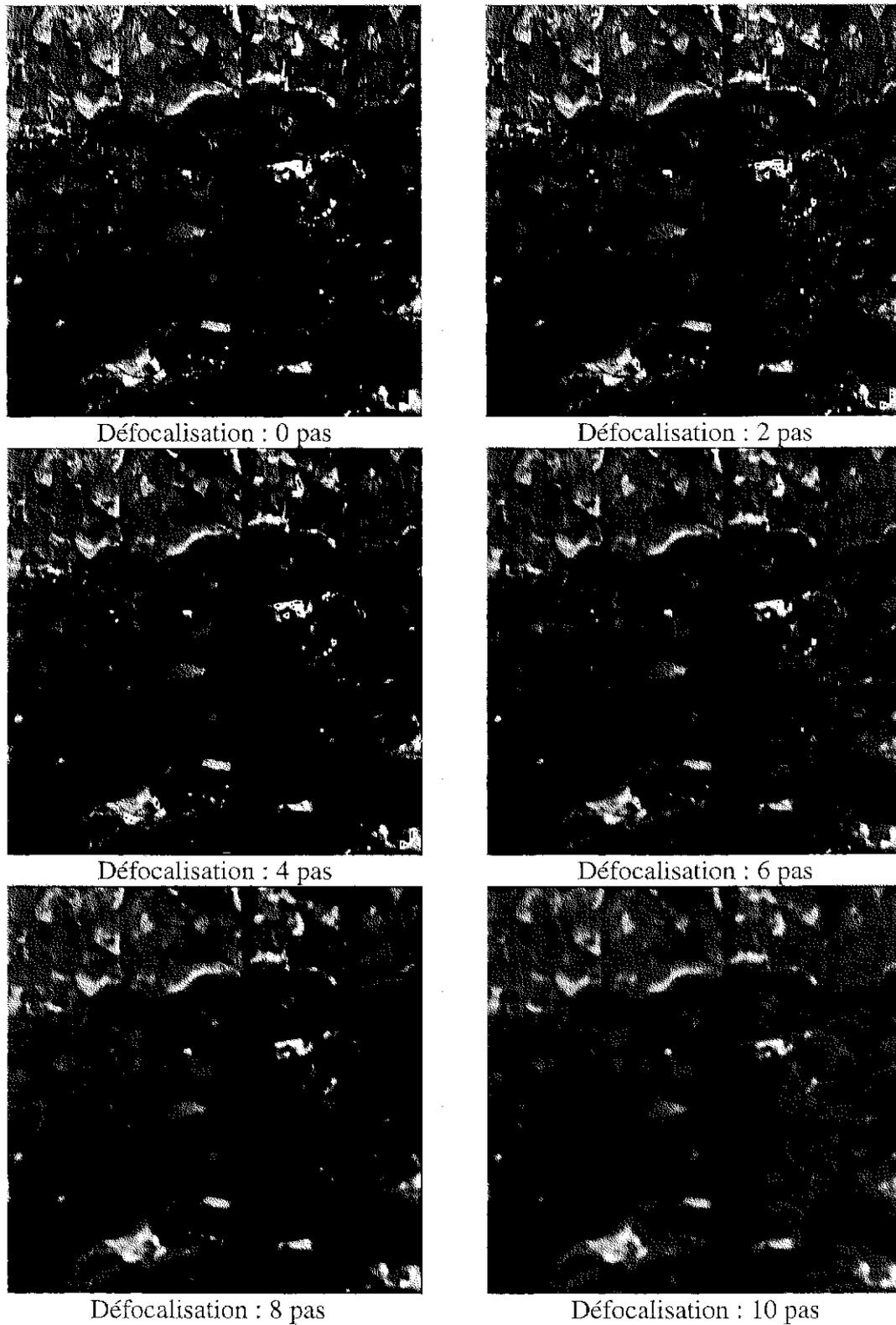


Figure 39 : Simulation de la défocalisation d'une image SPOT 1 à l'aide du modèle de défocalisation

Curieusement, nous n'avons pas réussi à extraire de $IM(f_r)$ une information fréquentielle aboutissant à une convergence du RN. Nous avons tenté d'utiliser les valeurs moyennes de $IM(f_r)$ sur plusieurs intervalles de fréquence, sans résultat. Il semble que la variation de $IM(f_r)$ en fonction du nombre de pas de défocalisation ne soit pas suffisamment importante pour permettre au RN de différencier ces derniers les uns des autres.

Nous avons imaginé une transformation sur la FFT de l'image pour définir un vecteur ayant une variation beaucoup plus importante que $IM(f_r)$.

La besseloïde, strictement monotone dans l'intervalle des défocalisations considérée, est transformée en un vecteur $V(f)$ par sommation des lignes et des colonnes :

$$V(f) = \sum_v |IM(f, v)| + \sum_u |IM(u, f)|$$

Cette transformation, mathématiquement correcte, n'a pas de réel fondement physique. Néanmoins, la relation qui associe $V(f)$ au nombre de pas de défocalisation p est bijective. A ce titre, nous pouvons utiliser $V(f)$ pour essayer de retrouver p .

La Figure 40 montre, dans le cas d'une image initialement à spectre plat, que la variation de $V(f)$ en fonction du nombre de pas de défocalisation est beaucoup plus prononcée que celle de $IM(f_r)$.

$V(f)$ n'est plus vraiment une information fréquentielle directement liée à l'image.

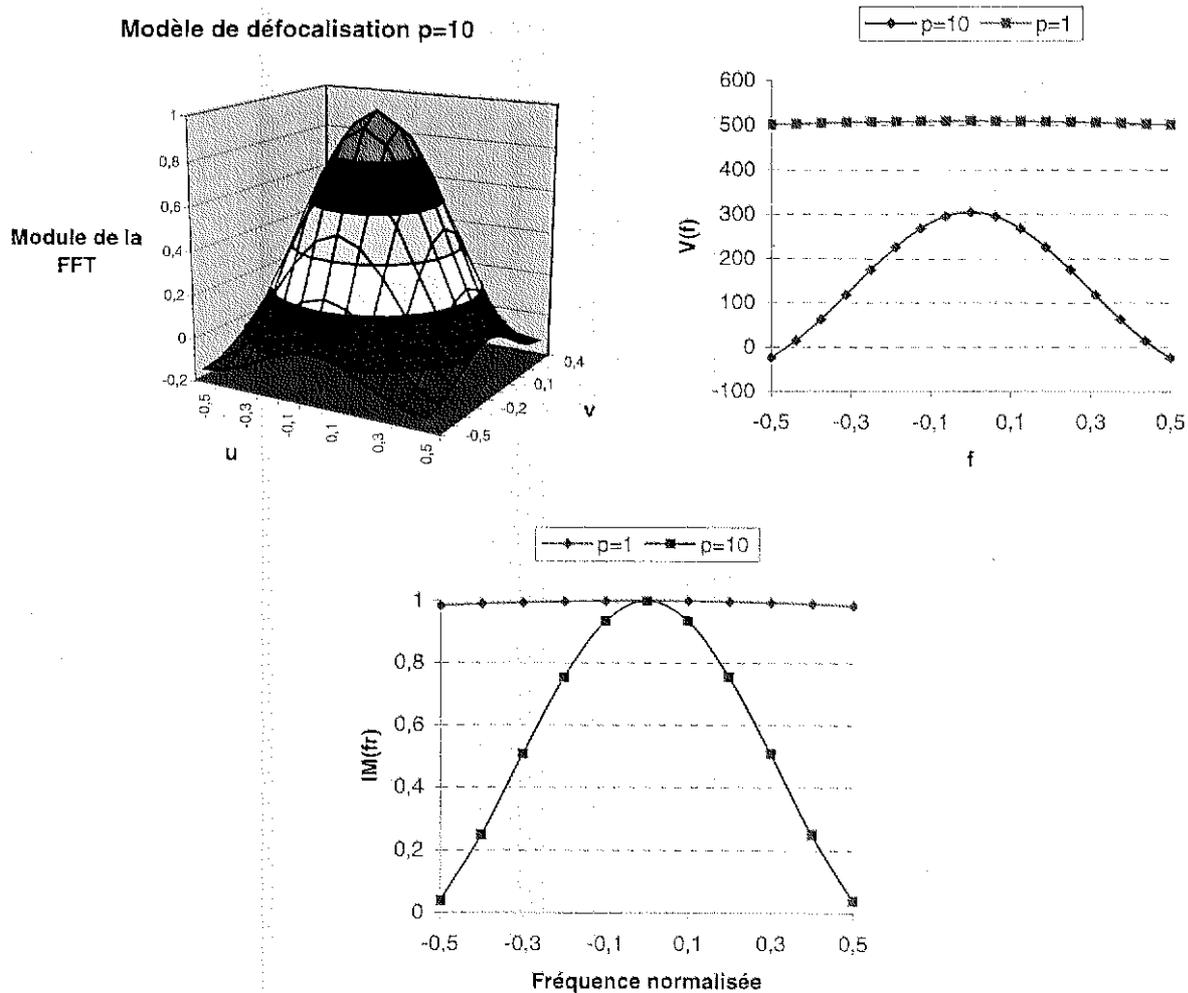


Figure 40 : Mise en évidence de la variation importante de $V(f)$ devant celle de $IM(fr)$ quand la défocalisation augmente.

La Figure 41 montre, pour des images de différents types, l'influence de la défocalisation sur $V(f)$. Notons que $V(f)$ dépend aussi du contenu fréquentiel de l'image initiale. Pour une image structurée comme Test014 et pour une même défocalisation, la valeur de $V(f)$ à une fréquence donnée est plus importante que pour une image relativement uniforme comme Test000. De plus, pour une image non dégradée, la plage de variation de $V(f)$ est moins importante que pour une image dégradée sur laquelle les hautes fréquences ont été coupées.

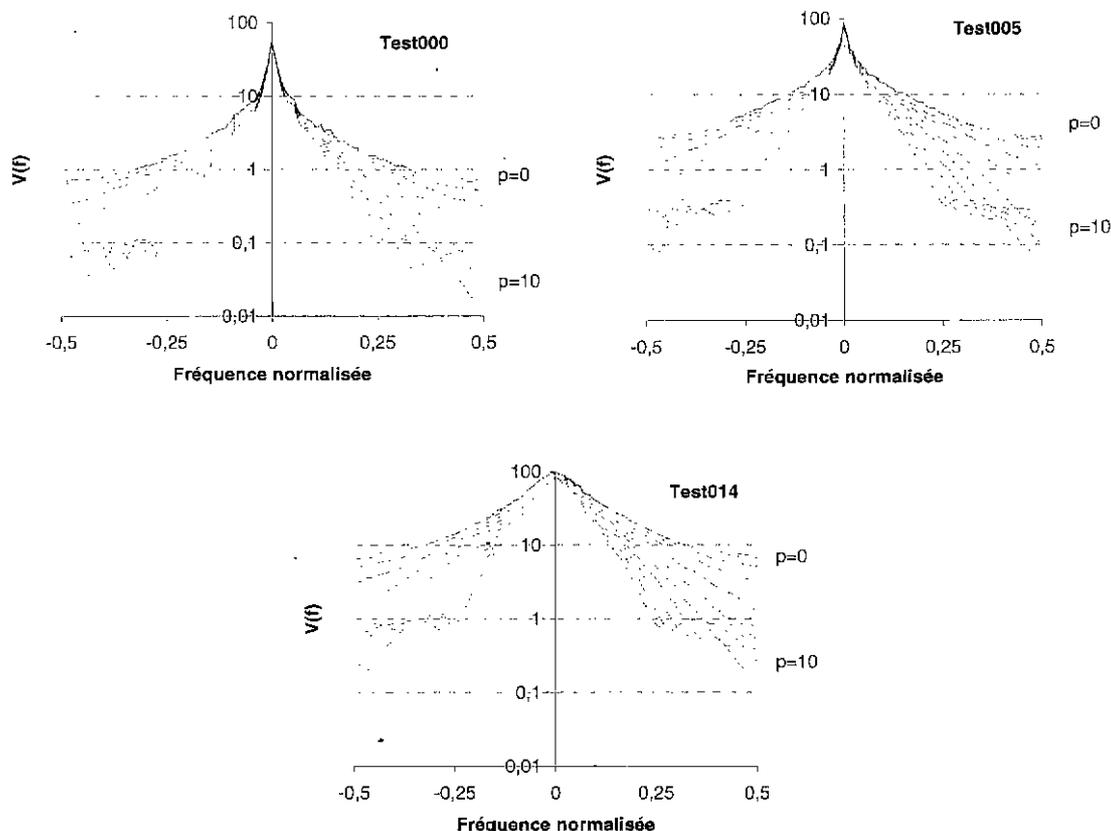


Figure 41 : Influence de la défocalisation sur la FFT pour différents types d'images.

Les valeurs de $V(f)$ pour des fréquences particulières n'ont pas véritablement de signification. Elles peuvent être très fortement bruitées par les particularités de l'image. En revanche, les valeurs moyennes de $V(f)$ sur plusieurs intervalles de fréquences sont susceptibles de donner une information à la fois sur la défocalisation et sur le type d'image.

Soient ff_1 la valeur moyenne de $V(f)$ sur l'intervalle $[0,3 fe/2 ; 0,4 fe/2]$, ff_2 la valeur moyenne de $V(f)$ sur l'intervalle $[0,4 fe/2 ; 0,5 fe/2]$, ..., ff_6 la valeur moyenne de $V(f)$ sur l'intervalle $[0,8 fe/2 ; 0,9 fe/2]$.

L'intervalle $[0 ; 0,3 fe/2[$ contient une information basse fréquence qui évolue peu avec la défocalisation. En revanche, l'intervalle $[0,9 fe/2 ; fe/2[$ contient une information haute fréquence, la plus susceptible d'être bruitée. Pour cette raison, nous n'avons pas, empiriquement, considéré ces intervalles.

La Figure 42 montre les variations des ff_i pour les trois images Test000, Test005 et Test014, pour différents pas de focalisation de l'instrument.

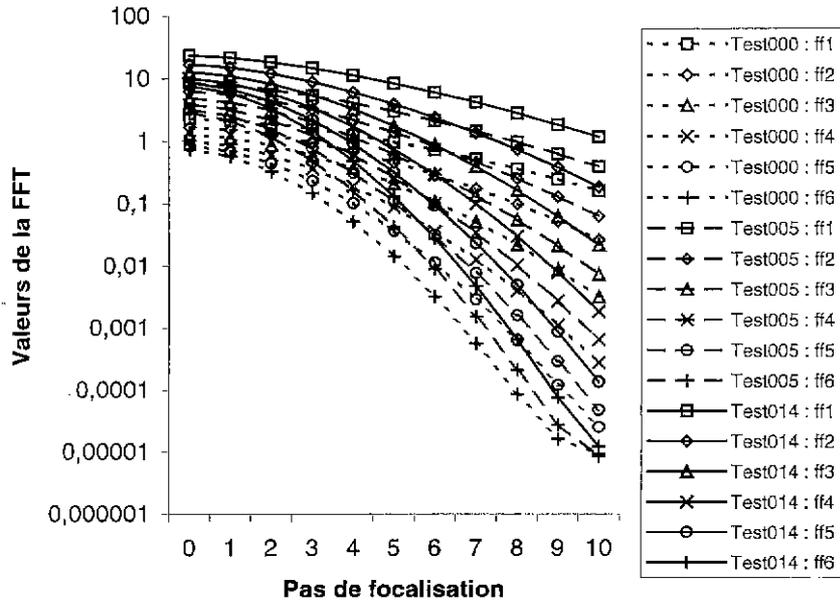


Figure 42 : Variation des ff_i pour différentes images et différents pas de focalisation.

On observe effectivement que les ff_i diminuent quand la défocalisation augmente. Cette diminution est d'autant plus forte que les fréquences considérées sont grandes. D'autre part, cette diminution dépend aussi de l'image considérée.

Nous retiendrons les valeurs ff_1, \dots, ff_6 pour le vecteur caractéristique de la dégradation.

ii -Variogramme de l'image

Le variogramme de l'image (cf. ANNEXE C -A3 -) est une statistique de l'image susceptible de compléter la FFT. Il représente l'inverse de la corrélation moyenne entre deux pixels, donné en fonction de la distance qui les sépare. Etant donné que la corrélation entre des pixels proches augmente quand la défocalisation augmente, on s'attend à ce que la forme du variogramme s'aplatisse dans ces conditions.

Nous avons montré au Chapitre 3 :II -B -1) que l'étalement du point pour une défocalisation de p pas était :

$$\delta = \frac{p\Delta}{N}$$

La taille d'un pixel sur le satellite SPOT est $w=13 \mu m$. On peut exprimer δ en fonction de p et w :

$$\delta = \frac{p\Delta}{Nw} = 0,22p \text{ en pixels}$$

Pour une défocalisation de 10 pas, le diamètre de la fonction d'étalement du point est inférieure à 3 pixels. Pour cette raison nous n'utiliserons que les valeurs du variogramme pour des distances de 1,2,3 pixels. La Figure 43 montre l'évolution de ces valeurs en fonction de la défocalisation.

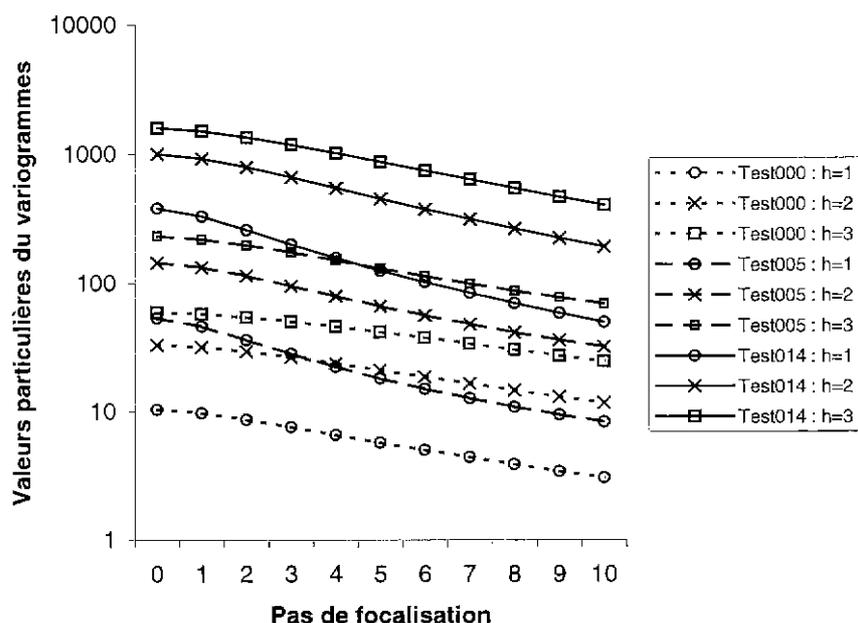


Figure 43 : Evolution du variogramme en fonction de la défocalisation de l'image. h représente la distance entre les pixels pour laquelle le variogramme est calculé

Comme attendu, les trois premières valeurs du variogramme diminuent quand la défocalisation augmente. Cette diminution est d'autant plus marquée que l'image est structurée (Test014).

Nous retiendrons ces trois valeurs pour le vecteur caractéristique de l'image, ainsi que la valeur du variogramme pour une distance de 20 pixels. Cette distance semble suffisante pour caractériser la variation à grande distance des niveaux de gris des pixels sur l'image.

iii - Moments de l'image

L'annexe (cf. ANNEXE C -A2 -) présente rapidement la signification physique des quatre premiers moments d'une image. La Figure 44 montre l'évolution de ces quatre moments en fonction de la défocalisation.

La moyenne d'une image reste constante quand la défocalisation augmente. Cela s'explique facilement si l'on se souvient que la défocalisation est une moyenne locale des pixels sur un voisinage plus ou moins grand. La moyenne apporte donc une information sur les propriétés de l'image non dégradée.

Pour la raison évoquée ci-dessus, la variance d'une image décroît quand la défocalisation augmente, elle décroît d'autant plus que l'image initiale est structurée. La variance apporte donc une information sur la défocalisation.

Il est plus délicat d'expliquer les variations des moments d'ordre trois et quatre de l'image. Le moment d'ordre 3 de l'image, ou skewness, caractérise la symétrie de l'histogramme des niveaux de gris autour de la valeur moyenne. Une distribution symétrique autour de sa moyenne a un skewness égal à zéro. Le skewness d'une image tend vers 0 quand la défocalisation augmente car tous les niveaux de gris se rapprochent du niveau de gris moyen. Cette grandeur apporte une information complémentaire sur la défocalisation.

Le moment d'ordre quatre, ou kurtosis, d'une image, représente la particularité de son histogramme à être « resserré », comparativement à une gaussienne. Un kurtosis supérieur à trois est représentatif d'une distribution plus piquée qu'une distribution gaussienne. D'une manière générale, l'histogramme des niveaux de gris de l'image a tendance à se piquer autour

de la moyenne quand la défocalisation augmente, pour cette raison le kurtosis de l'image augmente.

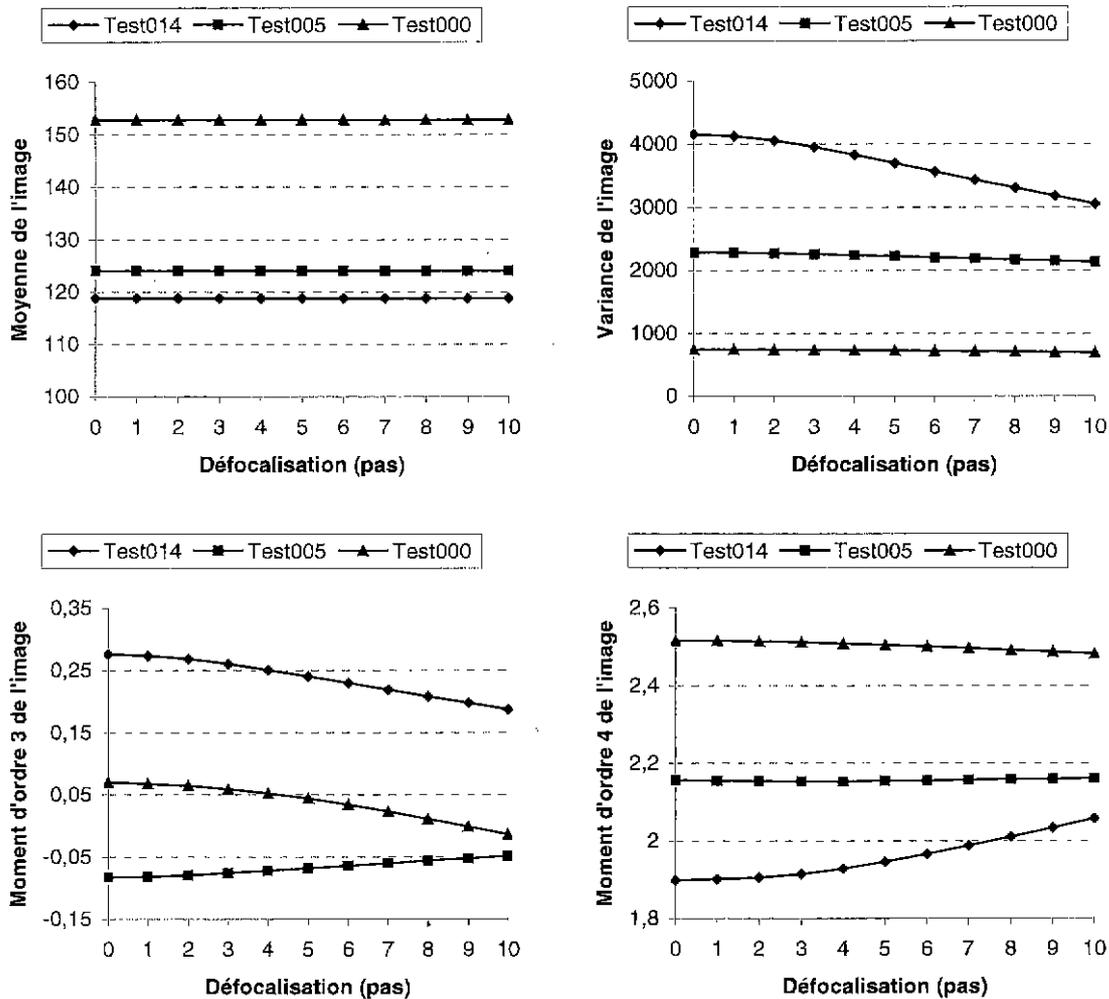


Figure 44 : Evolution des moments de l'image en fonction de la défocalisation.

Les quatre premiers moments de l'images sont retenus pour le vecteur caractéristique de la dégradation.

3) Le vecteur caractéristique de la défocalisation

Le vecteur caractéristique est finalement constitué des paramètres suivants :

0. Valeur du variogramme pour $h=1$ pixel
1. Valeur du variogramme pour $h=2$ pixels
2. Valeur du variogramme pour $h=3$ pixels
3. Valeur du variogramme pour $h=20$ pixels
4. Moyenne de l'image
5. Variance de l'image
6. Moment d'ordre 3 de l'image
7. Moment d'ordre 4 de l'image
8. FFT moyenne sur $[0,3 fe/2 ; 0,4 fe/2[$
9. FFT moyenne sur $[0,4 fe/2 ; 0,5 fe/2[$
10. FFT moyenne sur $[0,5 fe/2 ; 0,6 fe/2[$
11. FFT moyenne sur $[0,6 fe/2 ; 0,7 fe/2[$

12. FFT moyenne sur $[0,7 \text{ fe}/2 ; 0,8 \text{ fe}/2[$
13. FFT moyenne sur $[0,8 \text{ fe}/2 ; 0,9 \text{ fe}/2[$

4) Apprentissage de la défocalisation

L'apprentissage a été réalisé sur des images 512x512 pixels avec des configuration de RN différentes. La configuration retenue est celle qui donne les meilleurs résultats. Pour chacun des deux modèles précédemment mentionnés, il a été effectué de deux manières différentes. D'une part, il a été réalisé sur des images de nature diverses (cf. ANNEXE A -) et testé ensuite à la fois sur des images variées et sur des images SPOT, très spécifiques. D'autre part, il a été réalisé sur des images SPOT et testé sur des images SPOT mais aussi sur des images variées. Cela permet non seulement de vérifier le gain en précision que peut apporter un apprentissage spécifique à un contexte donné, mais aussi d'estimer la stabilité d'un apprentissage général ou spécifique.

Dans cette partie, le RN apprend à estimer le pas de focalisation du modèle avec lequel l'image défocalisée a été simulée.

a - Apprentissage sur la base générale

Les Figure 45 et Figure 46 présentent les résultats obtenus après un apprentissage sur la base d'images variées. On constate que les résultats sont légèrement moins bons pour un faible nombre de pas de focalisation. Dans ce domaine, la différence des dégradations introduites sur l'image pour deux pas successifs est bien moins importante que pour des fortes valeurs du pas. D'autre part, aux limites de l'intervalle considéré (pas proche de 0 ou de 10), le RN donne une estimation de la qualité moins précise que pour des pas moyens.

L'étude de sensibilité du RN vis-à-vis des composantes du vecteur caractéristique (cf. 5)) montrera que l'on peut améliorer ces résultats en supprimant les composantes les moins significatives physiquement, qui perturbent, peut-être, l'apprentissage du RN. Les résultats moyens restent cependant très satisfaisants.

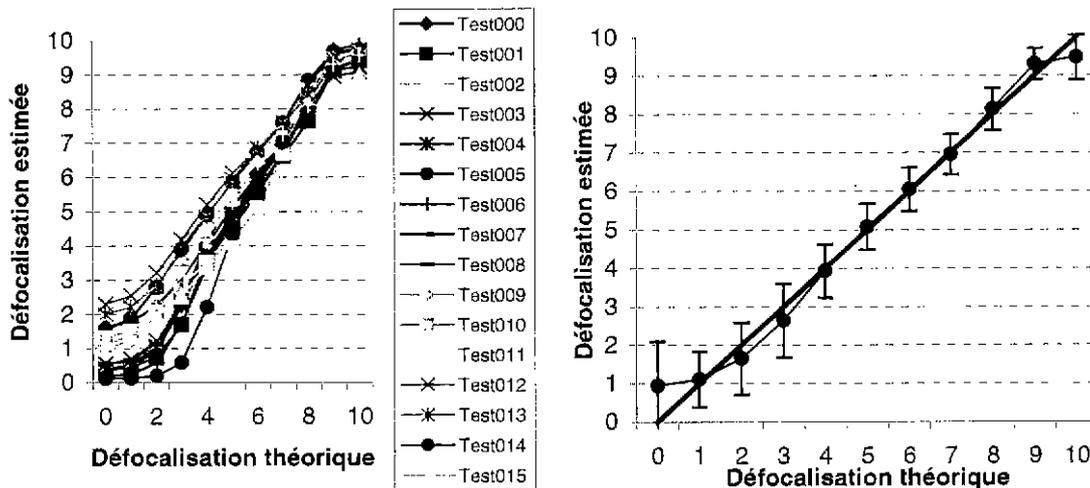


Figure 45 : Apprentissage sur la base générale, test sur des images variées. Le graphique de gauche montre les résultats de l'estimation de qualité obtenus pour chaque image de la base de validation, le graphique de droite montre le résultat moyen obtenu pour chaque qualité théorique ainsi que la dispersion quadratique autour de cette moyenne.

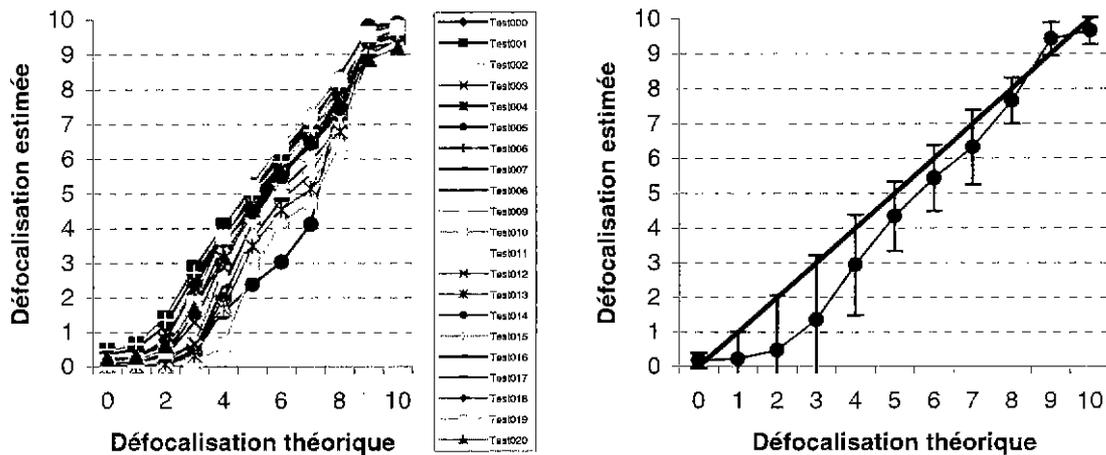


Figure 46 : Apprentissage sur la base générale, test sur des images SPOT. Le graphique de gauche montre les résultats de l'estimation de qualité obtenus pour chaque image de la base de validation, le graphique de droite montre le résultat moyen obtenu pour chaque qualité théorique ainsi que la dispersion quadratique autour de cette moyenne.

| Images SPOT | | |
|-------------|---------|-------|
| Image | EQM (%) | R |
| Spot000 | 3,7 | 0,995 |
| Spot001 | 3,2 | 0,995 |
| Spot002 | 7,7 | 0,988 |
| Spot003 | 8,7 | 0,987 |
| Spot004 | 5,2 | 0,993 |
| Spot005 | 19,1 | 0,937 |
| Spot006 | 13,7 | 0,975 |
| Spot007 | 12,0 | 0,978 |
| Spot008 | 13,7 | 0,974 |
| Spot009 | 11,3 | 0,980 |
| Spot010 | 19,1 | 0,951 |
| Spot011 | 14,3 | 0,971 |
| Spot012 | 9,9 | 0,979 |
| Spot013 | 15,5 | 0,968 |
| Spot014 | 8,0 | 0,988 |
| Spot015 | 3,4 | 0,995 |
| Spot016 | 7,0 | 0,989 |
| Spot017 | 7,0 | 0,991 |
| Spot018 | 5,5 | 0,991 |
| Spot019 | 6,9 | 0,988 |
| Spot020 | 7,5 | 0,990 |
| Moyenne | 9,6 | 0,981 |

| Images variées | | |
|----------------|---------|-------|
| Image | EQM (%) | R |
| Test000 | 5,9 | 0,991 |
| Test001 | 6,8 | 0,989 |
| Test002 | 4,5 | 0,991 |
| Test003 | 4,6 | 0,992 |
| Test004 | 5,1 | 0,993 |
| Test005 | 8,6 | 0,996 |
| Test006 | 6,9 | 0,989 |
| Test007 | 5,1 | 0,990 |
| Test008 | 4,7 | 0,992 |
| Test009 | 4,2 | 0,995 |
| Test010 | 6,4 | 0,980 |
| Test011 | 10,1 | 0,980 |
| Test012 | 11,8 | 0,996 |
| Test013 | 9,5 | 0,995 |
| Test014 | 11,2 | 0,976 |
| Test015 | 4,6 | 0,997 |
| Moyenne | 6,9 | 0,990 |

Tableau 6 : Erreurs quadratiques moyennes et coefficients de corrélation entre les qualités réelles et les qualités estimées par le RN.

b - Apprentissage sur la base d'images SPOT

Les Figure 47 et Figure 48 présentent les résultats obtenus après un apprentissage sur la base d'images spécifiques. Le test sur des images spécifiques donne de très bons résultats dans

l'estimation du pas de focalisation. En revanche les tests sur des images variées donnent des résultats de plus en plus fortement biaisés, voire aberrants, à mesure que la défocalisation augmente. C'est le cas quelle que soit la configuration de RN utilisée lors de la phase d'apprentissage.

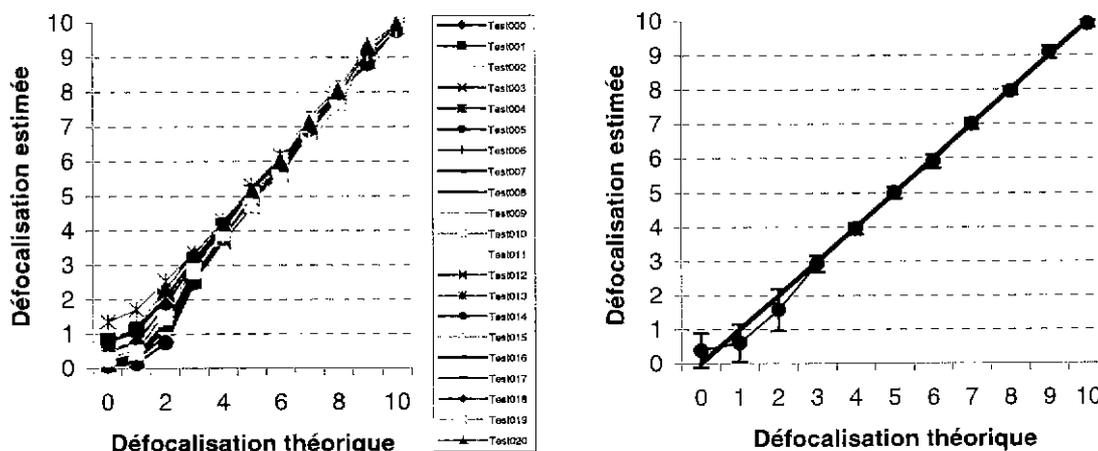


Figure 47 : Apprentissage sur la base SPOT, test sur des images SPOT. Le graphique de gauche montre les résultats de l'estimation de qualité obtenus pour chaque image de la base de validation, le graphique de droite montre le résultat moyen obtenu pour chaque qualité théorique ainsi que la dispersion quadratique autour de cette moyenne.

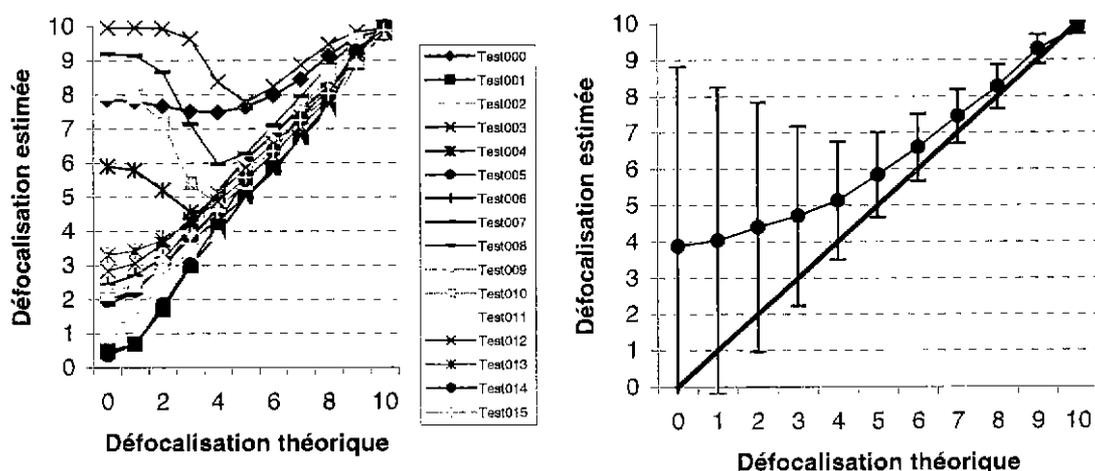


Figure 48 : Apprentissage sur la base SPOT, test sur des images variées. Le graphique de gauche montre les résultats de l'estimation de qualité obtenus pour chaque image de la base de validation, le graphique de droite montre le résultat moyen obtenu pour chaque qualité théorique ainsi que la dispersion quadratique autour de cette moyenne.

Le Tableau 7 montre le détail, pour chaque image, des résultats précédents. Hormis l'image Test000 qui est une image de champs, les plus grandes erreurs sont commises sur les images

représentant des portraits. Ces images sont effectivement très éloignées des images présentes dans la base d'apprentissage spécifique.

| Images SPOT | | |
|-------------|---------|-------|
| Image | EQM (%) | R |
| Spot000 | 2,4 | 0,998 |
| Spot001 | 2,8 | 0,998 |
| Spot002 | 4,4 | 0,996 |
| Spot003 | 4,5 | 0,996 |
| Spot004 | 3,7 | 0,997 |
| Spot005 | 5,0 | 0,993 |
| Spot006 | 2,0 | 0,998 |
| Spot007 | 2,8 | 0,997 |
| Spot008 | 1,9 | 0,998 |
| Spot009 | 1,8 | 0,999 |
| Spot010 | 4,0 | 0,997 |
| Spot011 | 1,8 | 0,998 |
| Spot012 | 2,2 | 0,998 |
| Spot013 | 5,3 | 0,998 |
| Spot014 | 3,0 | 0,997 |
| Spot015 | 2,6 | 0,997 |
| Spot016 | 3,4 | 0,997 |
| Spot017 | 3,6 | 0,996 |
| Spot018 | 2,1 | 0,998 |
| Spot019 | 2,8 | 0,997 |
| Spot020 | 3,2 | 0,999 |
| Moyenne | 3,1 | 0,997 |

| Images variées | | |
|----------------|---------|--------|
| Image | EQM (%) | R |
| Test000 | 41,2 | 0,833 |
| Test001 | 2,6 | 0,998 |
| Test002 | 8,1 | 0,993 |
| Test003 | 54,3 | -0,140 |
| Test004 | 25,6 | 0,831 |
| Test005 | 8,6 | 0,997 |
| Test006 | 10,1 | 0,983 |
| Test007 | 10,0 | 0,997 |
| Test008 | 44,8 | 0,185 |
| Test009 | 10,3 | 0,997 |
| Test010 | 36,6 | 0,457 |
| Test011 | 9,0 | 0,979 |
| Test012 | 13,2 | 0,995 |
| Test013 | 14,4 | 0,983 |
| Test014 | 2,8 | 0,997 |
| Test015 | 9,1 | 0,995 |
| Moyenne | 18,8 | 0,817 |

Tableau 7 : Erreurs quadratiques moyennes et coefficients de corrélation entre les qualités réelles et les qualités estimées par le RN.

L'apprentissage spécifique permet d'améliorer la précision dans l'estimation du pas de focalisation. Néanmoins sa stabilité étant bien moins importante que celle de l'apprentissage général, il impose de faire très attention aux images utilisées.

5) Etude de sensibilité

L'étude de sensibilité est réalisé de la manière décrite (cf. Chapitre 2 :III -C -).

La Figure 49 montre la sensibilité de l'estimation vis-à-vis des composantes du vecteur caractéristique dans le cas de l'apprentissage de la qualité d'images rendues floues avec le modèle de défocalisation.

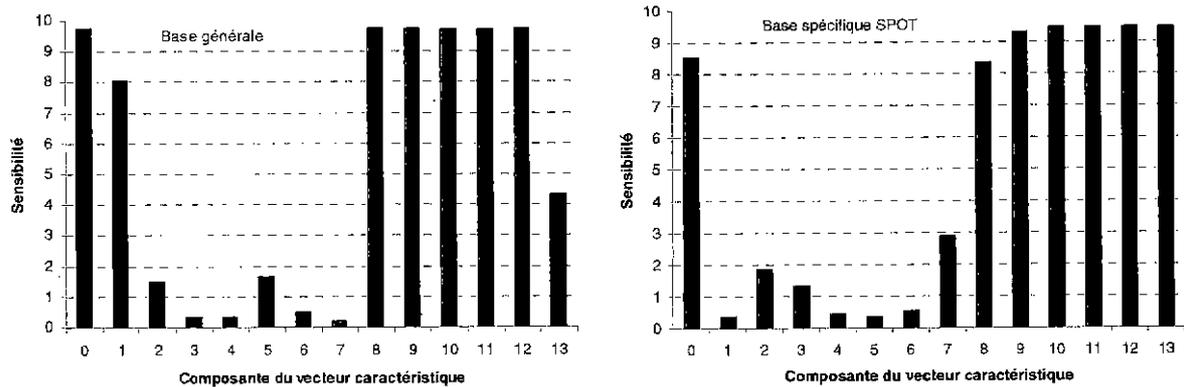


Figure 49 : Sensibilité aux composantes du vecteur caractéristique. Le graphique de gauche représente les sensibilité dans le cas d'un apprentissage sur la base générale, celle de droite représente les sensibilité dans le cas d'un apprentissage spécifique.

Les grandeurs importantes dans la prise de décision sont d'une part la composante n°0 du vecteur caractéristique, corrélation moyenne de chaque pixel avec ses voisins immédiats sur l'image, d'autre part les composantes fréquentielles moyennes.

Ces sensibilités s'expliquent facilement puisque l'effet de la défocalisation est d'augmenter la corrélation d'un pixel avec ses voisins, sur une distance de plus en plus grande au fur et à mesure que le flou devient important (jusqu'à 3 pixels pour $p=10$). Quant aux composantes fréquentielles, nous les avons calculées de manière à ce que leur variation caractérise la défocalisation.

Un apprentissage sur la base générale a été réalisé en ne conservant que les composantes du vecteur caractéristique dont la sensibilité est supérieure à 3. Les résultats pour chaque image sont présentés Tableau 8. La qualité de l'estimation du paramètre de focalisation se montre plus stable que celle obtenue avec toutes les composantes du vecteur caractéristique. Le RN converge plus facilement vers un extremum global dans un espace de dimension 6 que dans un espace de dimension 13 pour le même nombre d'exemples d'apprentissage.

| Images SPOT | | |
|-------------|---------|-------|
| Image | EQM (%) | R |
| Test000 | 6,3 | 0,994 |
| Test001 | 6,1 | 0,993 |
| Test002 | 6,3 | 0,986 |
| Test003 | 4,6 | 0,991 |
| Test004 | 5,4 | 0,990 |
| Test005 | 5,6 | 0,989 |
| Test006 | 7,9 | 0,988 |
| Test007 | 7,1 | 0,991 |
| Test008 | 7,0 | 0,984 |
| Test009 | 6,9 | 0,982 |
| Test010 | 7,1 | 0,979 |
| Test011 | 7,8 | 0,987 |
| Test012 | 9,5 | 0,986 |
| Test013 | 6,8 | 0,990 |
| Test014 | 7,9 | 0,980 |
| Test015 | 6,4 | 0,990 |
| Test016 | 4,9 | 0,994 |
| Test017 | 3,9 | 0,995 |
| Test018 | 5,3 | 0,996 |

| Images variées | | |
|----------------|---------|-------|
| Image | EQM (%) | R |
| Test000 | 12,0 | 0,989 |
| Test001 | 7,6 | 0,982 |
| Test002 | 5,9 | 0,988 |
| Test003 | 6,0 | 0,983 |
| Test004 | 6,4 | 0,987 |
| Test005 | 3,8 | 0,990 |
| Test006 | 4,8 | 0,990 |
| Test007 | 5,1 | 0,986 |
| Test008 | 6,0 | 0,986 |
| Test009 | 3,9 | 0,979 |
| Test010 | 7,1 | 0,993 |
| Test011 | 8,2 | 0,995 |
| Test012 | 5,7 | 0,989 |
| Test013 | 9,1 | 0,993 |
| Test014 | 5,5 | 0,994 |
| Test015 | 3,9 | 0,998 |
| Moyenne | 6,3 | 0,989 |

| | | |
|---------|-----|-------|
| Test019 | 4,2 | 0,995 |
| Test020 | 6,1 | 0,995 |
| Moyenne | 6,4 | 0,989 |

Tableau 8 : Erreurs quadratiques moyennes et coefficients de corrélation entre les qualités réelles et les qualités estimées par le RN.

En revanche, ce même apprentissage réalisé sur la base d'images spécifiques conduit à des résultats en tous points identiques à celui réalisé avec toutes les composantes du vecteur caractéristique, c'est à dire une très bonne précision sur des images spécifiques et une erreur moyenne de 20% sur les images variées.

On en déduit que la corrélation moyenne sur une distance de 1 pixel et le contenu fréquentiel de l'image sont nécessaires et suffisants pour estimer l'unique paramètre du modèle de focalisation avec la meilleure précision. La corrélation permet de juger de l'uniformité de l'image sur une distance courte, alors que les composantes fréquentielles permettent au RN de retrouver la coupure de la FTM.

Nous nous sommes aperçus par la suite que seules les composantes fréquentielles pouvaient être utilisées, la coupure de la FTM étant directement reliée au nombre de pas de défocalisation. En effet, la bande passante à -3 dB de cette FTM est donnée par la résolution de l'équation :

$$\frac{2J_1(\alpha p f_r)}{\alpha p f_r} = \frac{1}{2}$$

La solution numérique de cette équation est $\alpha p f_r = \pm 2,21$. La bande passante à -3 dB est donc donnée en fonction de p par : $\Delta f_r \approx \frac{6}{p}$. Toute l'information sur p est contenue dans les composantes fréquentielles.

6) Conclusion

Lorsque une image est rendue floue grâce à un modèle de défocalisation, il est possible d'estimer les paramètres de ce modèle sans image de référence. Ceci est réalisé en grande partie au regard du contenu fréquentiel de l'image et moyennant une phase d'apprentissage préalable adéquate. De plus, un apprentissage spécifique à un contexte donné permet d'améliorer sensiblement l'estimation de la qualité, ou des paramètres caractéristiques. Néanmoins, un tel apprentissage spécifique peut se montrer très instable dès lors que les images présentées sont éloignées du contexte d'apprentissage.

La phase d'apprentissage nécessite une connaissance a priori de la fonction qui transforme une image parfaite en une image floue. Dans la partie suivante, nous allons montrer qu'il est possible, dans certains cas, d'appliquer cette méthode pour estimer en vol la défocalisation éventuelle d'un instrument imageur, sans connaissance a priori du modèle de focalisation.

C - Application : étude de la défocalisation du satellite SPOT 1

1) Introduction – Présentation du problème

En juillet 1994, le Comité Directeur SPOT a donné son aval pour une expérimentation technologique de refocalisation des instruments HRV de SPOT1. Cette étude devait permettre non seulement de valider un mécanisme qui n'avait jamais été testé en vol, mais aussi des méthodes de mesure de la défocalisation et la procédure de refocalisation.

Cette étude a été menée de fin septembre à fin novembre 1994 sur 11 orbites du satellite. Un bilan de FTM relatives de HRV2 par rapport à HRV1 pour plusieurs focalisations de HRV2

devait permettre de mettre en évidence une défocalisation de HRV2 de deux pas environ [LeM95].

Rappelons brièvement les grandeurs connues et celles que l'on cherche à estimer.

Le système est réglé au mieux, avant le tir, pour être bien focalisé, mais le mécanisme permet une éventuelle refocalisation en autorisant un déplacement du plan image autour de sa position initiale. Le pas de focalisation p représente la position du plan image par rapport à la position de référence. Cette référence, dans le cas de SPOT1, est fixée au lancement à la position $p=p_0$ supposée de bonne focalisation, c'est à dire à la distance focale du centre optique du système. Le pas est entier et une variation de 1 pas de focalisation correspond à un déplacement $\Delta=10 \mu\text{m}$ du plan image.

Dans le cas d'un instrument mal focalisé (Figure 50), la position de meilleure mise au point est obtenue pour un pas $p=p_0+p_f$, avec p_f positif ou négatif selon la position par rapport à l'état initial.

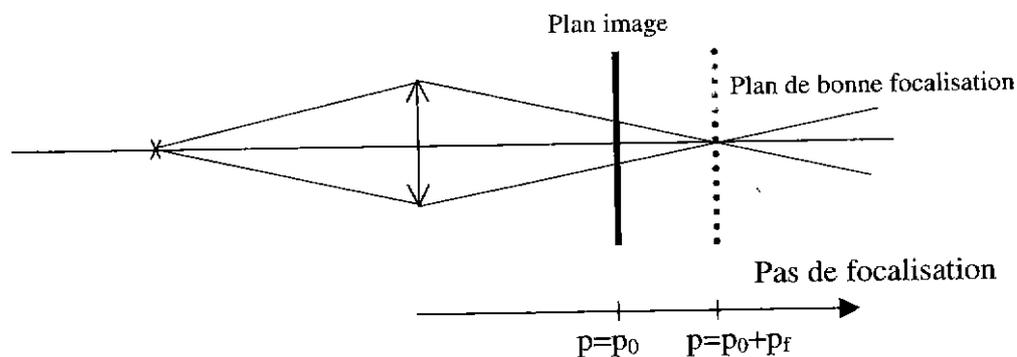


Figure 50 : Configuration optique d'un instrument mal focalisé.

Notre objectif est de montrer qu'il est possible d'obtenir une estimation de p_f à partir d'images acquises avec différentes valeurs du pas de focalisation. Nous allons utiliser les images SPOT1 de la campagne de 1994 pour montrer qu'il est possible d'utiliser notre méthode univariante afin d'estimer la valeur de p_f .

2) Estimation du pas de bonne focalisation

Nous avons dans un premier temps essayé d'utiliser les apprentissages réalisés sur le modèle de flou géométrique (cf. Chapitre 3 :II -B -1)). Pour différentes raisons, qui seront évoquées au Chapitre 3 :II -C -4), ce modèle s'est avéré insuffisant pour estimer la défocalisation des images SPOT 1 dont nous disposons. Nous avons donc réalisé un apprentissage directement sur les images SPOT 1 défocalisées issues de la campagne de 1994.

En raison du peu d'images SPOT1 défocalisées disponibles, 12 au total, les bases d'apprentissage et de test ont été réalisées de la manière suivante : chaque image a été découpée en imagerie 256x256 pixels. Une image panchromatique 6000x6000 pixels est découpée en 529 imagerie distinctes. 10 % de ces imagerie sont tirées au sort pour constituer la base de test, les autres constituent la base d'apprentissage. La base d'apprentissage comporte donc environ 5700 images pour une base de test de 630 images.

a - Estimation du pas de focalisation

Dans un premier temps, montrons que la méthode univariante est capable de retrouver le pas de focalisation qui correspond à une image.

L'apprentissage est réalisé sur le jeu d'images SPOT 1 défocalisées. Le vecteur caractéristique utilisé est le même que pour les modèles de défocalisation (cf. Chapitre 3 :II - B -3)). Le RN apprend à associer au vecteur caractéristique la valeur, connue, du pas de focalisation correspondant. Etant donné que le déplacement du plan image est très petit devant la distance focale, nous émettons, comme pour le modèle, l'hypothèse que la défocalisation ne dépend pas du sens de déplacement. Par conséquent une image acquise à $p_0 + p_f - p$ ($p \in [-10,10]$) est la même qu'une image acquise à $p_0 + p_f + p$. L'apprentissage a été réalisé de manière à estimer $|p - p_f - p_0|$. Nous avons, pour cette étude préliminaire, utilisé l'estimation $p_f = 1,8$ pas, qui avait été déterminée de manière bivariante lors de la campagne de 1994.

La Figure 51 montre sur la base de test que le RN est capable de différencier deux pas très proches l'un de l'autre. A titre de comparaison, l'œil humain ne peut guère différencier deux images acquises avec moins de 3 pas d'écart.

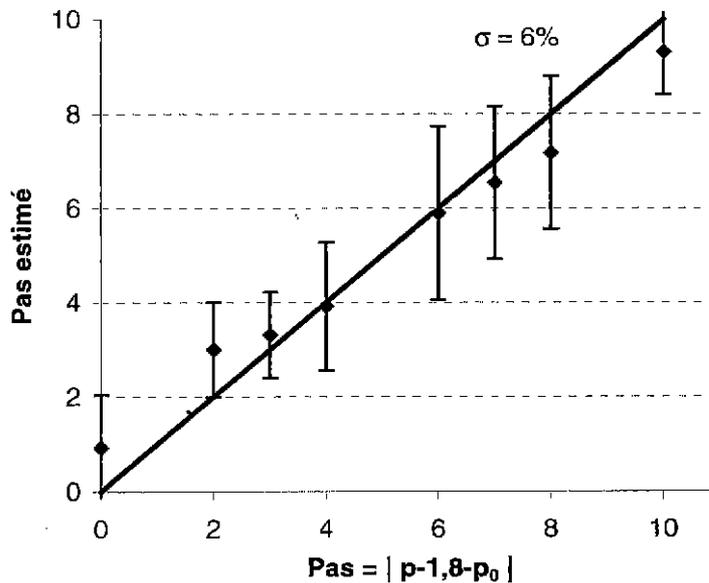


Figure 51 : Validation sur la base de test de l'apprentissage du pas de focalisation.

b - Estimation de la défocalisation

Le problème posé est le suivant : on désire estimer la défocalisation d'un instrument alors que l'on n'a pas réalisé d'apprentissage préalable. On suppose que l'on est capable d'acquérir plusieurs images pour plusieurs pas de focalisation différents.

Notre démarche repose sur les deux hypothèses suivantes :

- H₁) L'effet de la défocalisation sur deux images acquises respectivement avec des pas $p_0 + p_f - p$ et $p_0 + p_f + p$ est exactement le même.
- H₂) On est capable d'estimer une valeur maximale de $|p_f|$, par des observations ou par la connaissance du système par exemple.

L'estimation de la défocalisation se fait alors en trois étapes :

- 1) Un apprentissage du pas de focalisation identique à celui de a) est réalisé dans le domaine $p > p_0 + p_f$ et ce grâce à H₂ ;

- 2) Le RN, une fois l'apprentissage effectué, est ensuite utilisé pour évaluer le pas de focalisation d'images acquises dans le domaine $p < p_0 - p_f$ (de manière à ne jamais risquer d'être dans l'intervalle $p_0 \pm p_f$);
- 3) L'apprentissage et l'utilisation sont confrontés et la valeur de p_f est déduite grâce à H_1 .

i - Etape 1 : phase d'apprentissage

La phase d'apprentissage doit être réalisée pour des images acquises avec un pas de focalisation $p > p_0 + p_f$. Une étude visuelle des images SPOT 1 défocalisées de la campagne de 1994, ajoutée à une bonne connaissance de l'instrument, permet d'affirmer que la défocalisation, si elle existe, est inférieure à 2-3 pas. En effet, une telle défocalisation commence à peine à être visible, alors que dans le cas de ces images elle est totalement imperceptible.

Pour mener à bien notre étude, nous aurions besoin de plusieurs images acquises à des pas de focalisation régulièrement répartis sur les intervalles $[p_0+3 ; 10]$ et $[p_0-3 ; -10]$.

Cependant, nous disposons de peu d'images (douze au total) et quatre d'entre elles ont été acquises sur l'une des limites $p_0 \pm 2$. Etant donné que nous cherchons à valider le concept qui permettra d'estimer p_f et que nous avons besoin d'un grand nombre d'images, nous avons utilisé quand même ces quatre dernières images en rappelant que l'étude menée en 1994 pour estimer la défocalisation de HRV2 par comparaison avec HRV1, avait conduit à une estimation moyenne de 1,8 pas. Cela ne serait évidemment pas permis pour une véritable étude en vol de la défocalisation.

La Figure 52 montre, sur la base de test, les résultats obtenus après apprentissage du pas de focalisation pour des images acquises avec $p > p_0 + p_f$. Ces résultats sont d'une qualité tout à fait comparable à celle obtenue au a - . Un apprentissage parfait donnerait une estimation égale à la valeur attendue relativement à p_0 . L'erreur quadratique moyenne, σ , et le coefficient de corrélation R^2 sont donc calculés par rapport à cette valeur, dont les représentations sont situées sur la bissectrice des axes.

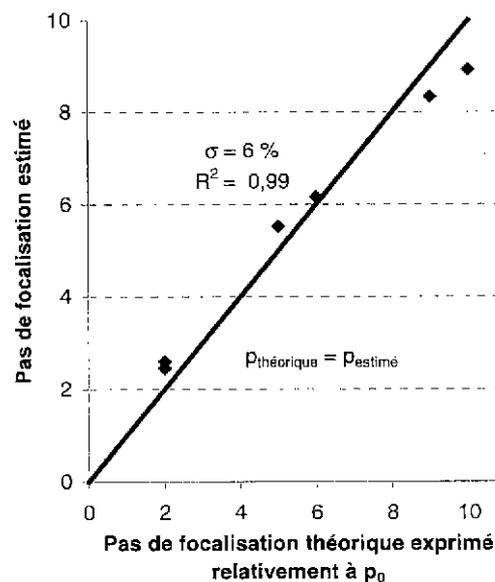


Figure 52 : Phase d'apprentissage du RN sur des images acquises avec $p > p_0 + p_f$. Résultats obtenus sur la base de test.

ii -Utilisation du RN pour des images acquises avec $p < p_0 - p_f$

L'apprentissage effectué, le RN est utilisé pour estimer le pas de focalisation d'images acquises dans le domaine $p < p_0 - p_f$. Pour ces images, le pas de focalisation est connu. La Figure 53 présente les valeurs des pas estimés par le RN sur ces images. Pour un instrument bien focalisé, l'estimation du pas est la valeur absolue de la valeur réelle (hypothèse de symétrie). Ce n'est pas le cas ici car l'instrument est mal focalisé. Ces estimations sont néanmoins sensées être situées sur une droite de pente -1 car elles sont juste translatées d'une valeur p_f par rapport à celles que l'on aurait obtenues sur un instrument bien focalisé. L'ordonnée à l'origine de cette droite est par ailleurs directement liée à p_f comme nous allons le montrer.

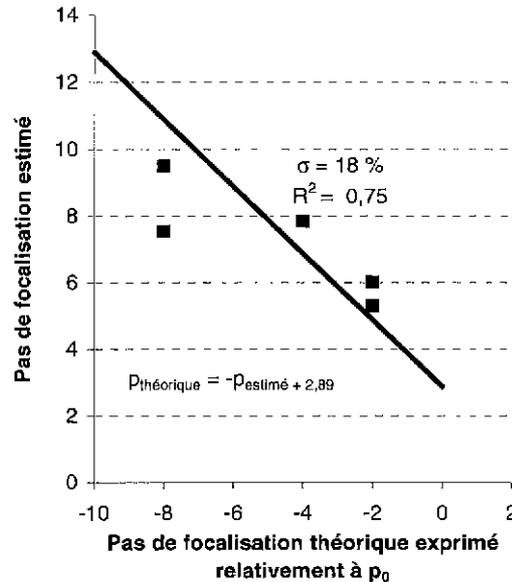


Figure 53 : Phase d'utilisation du RN sur des images acquises avec $p < p_0 - p_f$. Résultats obtenus sur la base de test

L'écart de certaines mesures par rapport à la droite théorique peut être expliqué par le peu d'images dont nous disposons pour réaliser l'apprentissage. Un tel apprentissage est forcément moins stable que lorsque il est réalisé sur un grand nombre d'exemples.

iii -Estimation de la défocalisation

Pour un instrument défocalisé de p pas (p positif ou négatif) et dont on connaît la position de bonne focalisation p_0 , la valeur \hat{p} du pas estimé par le RN est l'écart relatif par rapport à p_0 :

$$\hat{p} = |p - p_0|$$

Pour un instrument mal focalisé d'un nombre de pas p_f , on a de même :

$$\hat{p} = |p - p_f - p_0|$$

Appelons \hat{p}_+ la valeur du pas estimé pour une défocalisation $p_+ \in]p_0 + p_f, 10]$ et \hat{p}_- la valeur du pas estimée pour une défocalisation $p_- \in [-10, p_0 - p_f[$. On a alors :

$$\hat{p}_+ = |p_+ - p_f - p_0| = p_+ - p_f - p_0$$

$$\hat{p}_- = |p_- - p_f - p_0| = -p_- + p_f + p_0$$

La représentation de \hat{p} en fonction de p admet un axe de symétrie qui est alors la droite d'équation $p = p_0 + p_f$. D'où :

$$p_0 + p_f = \frac{p_+ + p_-}{2} \quad \forall (p_+, p_-) \text{ tels que } \hat{p}_+ = \hat{p}_-$$

En pratique, l'apprentissage est réalisé sur les (p_+) et testé sur les (p_-) , les estimations sont donc respectivement situées sur les droites $\hat{p}_+ = p_+$ et $\hat{p}_- = -p_+ + Cte$.

L'ordonnée à l'origine de cette deuxième droite, c'est-à-dire la valeur numérique de Cte , est obtenue par régression linéaire sur les estimations dans le domaine $p < p_0 - p_f$.

Dès lors on en déduit la bonne focalisation :

$$p_0 + p_f = \frac{Cte}{2}$$

La Figure 54 reprend cette démarche pour les mesures SPOT 1.

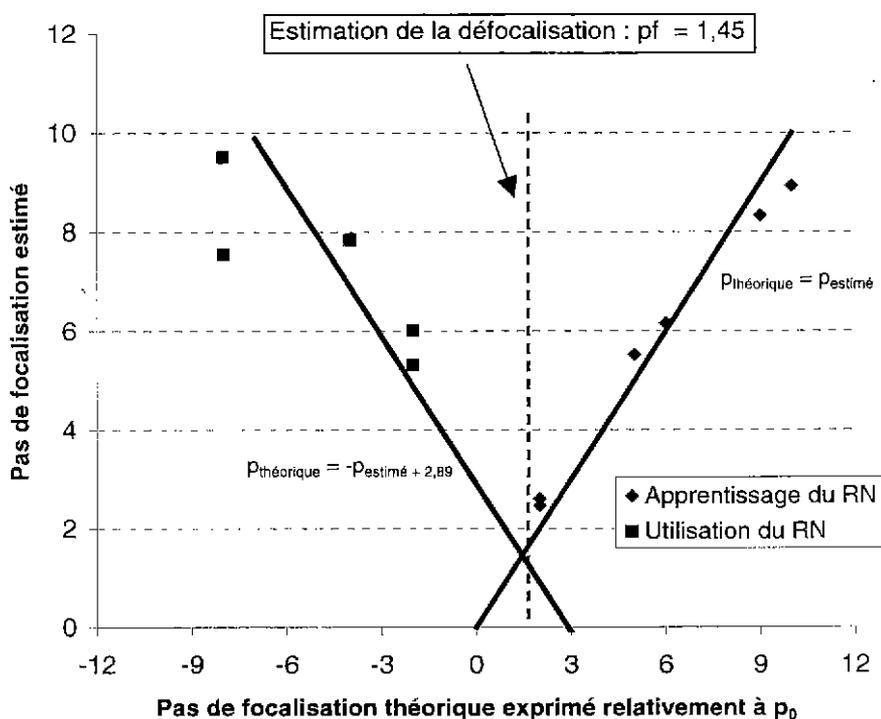


Figure 54 : confrontation des mesures et estimation de la défocalisation.

L'estimation de la défocalisation, 1,45 pas, est en accord avec celle qui avait été mesurée lors de l'étude de 1994. Malheureusement de telles expérimentations sont très rares et nous n'avons pu valider la méthode que sur ce jeu d'images SPOT1. C'est d'ailleurs aussi une des limitations de la méthode.

3) Etude de sensibilité

L'étude de sensibilité (Figure 55) montre que certaines composantes du vecteur caractéristique semblent avoir une influence mineure sur l'évaluation de la défocalisation. C'est le cas des composantes 2, 3, 4, 6 et 7. Les composantes 2 et 3 sont respectivement les valeurs du variogramme pour $h=3$ pixels et $h=20$ pixels. Elles représentent les corrélations moyennes de deux pixels séparés respectivement des distances 3 et 9 pixels. Il est compréhensible que cette dernière intervienne peu, le flou étant surtout caractérisé par une

augmentation locale de la corrélation entre des pixels voisins. L'étude sur les modèles avait donné des résultats comparables.

Cependant, la variance de l'image se montre importante dans le cas réel, alors qu'elle ne l'était pas sur les modèles. La suppression des composantes pour lesquelles la sensibilité est moindre permet de vérifier leur importance physique dans l'estimation de la défocalisation.

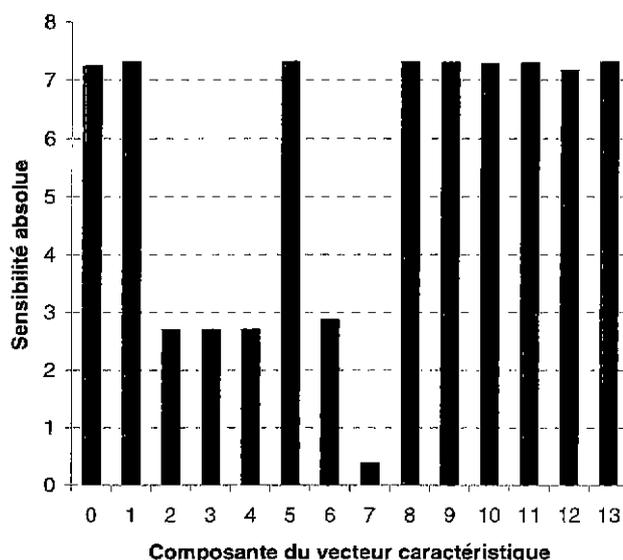


Figure 55 : Etude de la sensibilité absolue vis-à-vis de chacune des composantes du vecteur caractéristique.

Comme dans le cas des modèles, un apprentissage a été effectué en ne conservant que les entrées dont la sensibilité est supérieure à 3. La dégradation dans l'estimation de la défocalisation est de l'ordre de 5%. En revanche, si l'on supprime la variance comme on l'avait fait avec succès avec les modèles de défocalisation, la dégradation des résultats atteint 10%.

Une des raisons pour laquelle la variance est un paramètre important dans le cas réel alors qu'elle ne l'est pas avec le modèle peut être la suivante : lors du lissage d'une image SPOT avec le modèle, le bruit contenu dans l'image est lissé aussi. Il ne contient plus de composantes fréquentielles au delà de la fréquence de coupure.

Dans le cas réel, l'image est lissée par la défocalisation et tous les bruits électroniques se rajoutent ensuite. L'image floue contient les composantes hautes fréquences du bruit. La variance de l'image, ainsi que les autres paramètres dans une moindre mesure, apportent au RN une information complémentaire permettant de ne pas sous estimer la défocalisation au seul regard des composantes fréquentielles.

4) Stabilité

L'apprentissage réalisé à partir du modèle de défocalisation et celui réalisé sur les images SPOT 1 défocalisées ont été testés sur des images SPOT provenant de satellites divers (SPOT 1, 2, 3) et pour les deux instruments (HRV1 et HRV2). Certains instruments ont une focalisation imparfaite et connue :

- HRV1 de SPOT3 est défocalisé de +2 pas environ ;
- HRV2 de SPOT3 est défocalisé de +4 pas environ.

Chaque image est découpée en 529 imagerie 6000x6000 pixels. La défocalisation de chacune de ces imagerie est évaluée. La défocalisation de l'image initiale est calculée comme étant la

moyenne des 529 défocalisations des imgettes, la barre d'erreur étant l'écart type des défocalisations des imgettes.

Les résultats obtenus sont synthétisés Tableau 9. Les images portent un nom du type « S_KKK_JJJ_B_H » où S est le satellite concerné (SPOT1, 2, 3, 4), KKK et JJJ sont les coordonnées de la scène exprimées dans le référentiel utilisé par SPOT, B est le type d'image (P ou XS, cf. ANNEXE E -) et H l'instrument de mesure (HRV1 ou HRV2).

| Image | Défocalisation théorique | Défocalisation estimée d'après l'apprentissage sur le modèle | Défocalisation estimée d'après l'apprentissage sur SPOT1 |
|---------------|--------------------------|--|--|
| 1_541_282_P_2 | 2 | 3,3 ± 0,7 | 2,4 ± 1,7 |
| 1_541_282_P_1 | 0 | 3,3 ± 0,6 | 0,6 ± 2,3 |
| 1_137_280_P_1 | 0 | 4,1 ± 1,0 | 0,5 ± 1,4 |
| 2_147_300_P_1 | 0 | 4,0 ± 2,4 | 1,5 ± 1,9 |
| 3_054_236_P_1 | 2 | 1,2 ± 1,1 | 2,6 ± 1,8 |
| 3_280_269_P_1 | 2 | 7,5 ± 0,8 | 3,8 ± 3,4 |
| 3_305_273_P_1 | 2 | 4,2 ± 0,7 | 4,9 ± 1,8 |
| 3_675_379_P_1 | 2 | 2,0 ± 1,0 | 2,0 ± 2,2 |
| 3_677_382_P_1 | 2 | 1,9 ± 0,7 | 4,1 ± 2,9 |
| 3_033_268_P_2 | 2 | 4,0 ± 1,1 | 3,4 ± 2,1 |
| 3_695_410_P_2 | 4 | 1,7 ± 0,7 | 2,0 ± 1,5 |
| 3_033_246_P_2 | 6 | 7,9 ± 1,0 | 7,7 ± 3,0 |
| 3_054_262_P_2 | 6 | 6,7 ± 2,1 | 2,2 ± 3,1 |
| 3_054_269_P_2 | 6 | 6,4 ± 1,0 | 7,2 ± 1,8 |

Tableau 9 : Comparaison de la performance des différents apprentissages réalisés (sur le modèle et sur les images de la campagne SPOT1) sur des images acquises avec des instruments différents et parfois défocalisés.

L'apprentissage réalisé sur les images SPOT1 défocalisées donne de bons résultats sur des images SPOT 1 inconnues, mais des résultats un peu aberrants sur des instruments différents. Ceci est expliqué en partie par le fait que les instruments ont des niveaux de FTM et des niveaux de bruit différents et donc que l'apprentissage réalisé sur un instrument n'est pas transportable sur un autre

D'autre part, l'apprentissage réalisé sur le modèle de défocalisation permet de prédire le pas de focalisation des images SPOT 3 de manière convenable, mais se montre approximatif pour les autres instruments. Les niveaux de FTM et de bruit des instruments peuvent à nouveau en être une des causes. Une autre raison, qui renforce la nécessité de pouvoir faire un apprentissage sur des images défocalisées par l'instrument et non pas par le modèle, est que dans le cas réel, l'image est défocalisée, puis échantillonnée et bruitée, alors que le modèle « défocalise » une image déjà échantillonnée et bruitée.

On pourrait néanmoins procéder à une simulation plus réaliste en utilisant des images parfaites et en rajoutant le bruit après le défaut de mise au point.

5) Conclusion

Nous avons montré dans ce chapitre, que, pour une image rendue floue par un filtrage passe-bas, il est possible de retrouver directement sur l'image le paramètre du filtre utilisé. Ceci est réalisé grâce à un apprentissage neuronal qui met en jeu les seules composantes fréquentielles de l'image. Dans le cas d'un modèle de flou géométrique, il permet de retrouver le défaut de mise au point, exprimé en unité de longueur par exemple. Nous avons aussi, au début des travaux, utilisé un filtre gaussien pour montrer que l'on pouvait en retrouver la bande passante à partir de l'image dégradée.

La méthode a ensuite été adaptée, grâce à un apprentissage en vol, à l'étude du défaut de mise au point du satellite SPOT 1. Elle a permis, moyennant quelques hypothèses très réalistes physiquement, d'évaluer le défaut de mise au point de l'instrument HRV2 à 1,45 pas. Une étude bivariate menée à la suite de la campagne de 1994 et utilisant la comparaison des FTM des deux instruments HRV1 et HRV2 de SPOT1, supposés identiques, avait abouti à la conclusion que le défaut de mise au point était légèrement inférieur à deux pas.

L'avantage de notre méthode est de permettre une telle évaluation même si l'on ne dispose que d'un seul instrument. Néanmoins, elle nécessite de disposer de plusieurs images, acquises avec des nombres de pas de focalisation différents et régulièrement espacés.

Chapitre 4 :
Application à une image dégradée par des
défauts ponctuels

I - Introduction

L'objectif de cette partie est de montrer que la méthode univariante peut être utilisée pour détecter des singularités sur une image. De telles singularités, défauts quasi ponctuels et plus ou moins gênants selon la zone de l'image dans laquelle ils se trouvent, peuvent exister sur les images du satellite SPOT (cf. ANNEXE E -). Actuellement, le personnel du service qualité de SPOTIMAGE a pour charge de les détecter, visuellement, pour éventuellement les corriger. Un outil de détection automatique serait appréciable.

Une telle singularité n'est pas toujours visible. En effet, elle peut être masquée par son voisinage (Figure 56). Nous nous proposons ici de montrer que la méthode univariante peut simuler la détection visuelle. La méthode est étalonnée sur des tests visuels réalisés pour cette étude. De plus, pour montrer l'efficacité de cette détection, nous l'avons comparée à la détection faite par un modèle visuel bivariant complexe, le VDP (cf Chapitre 1 :III -B -1)).

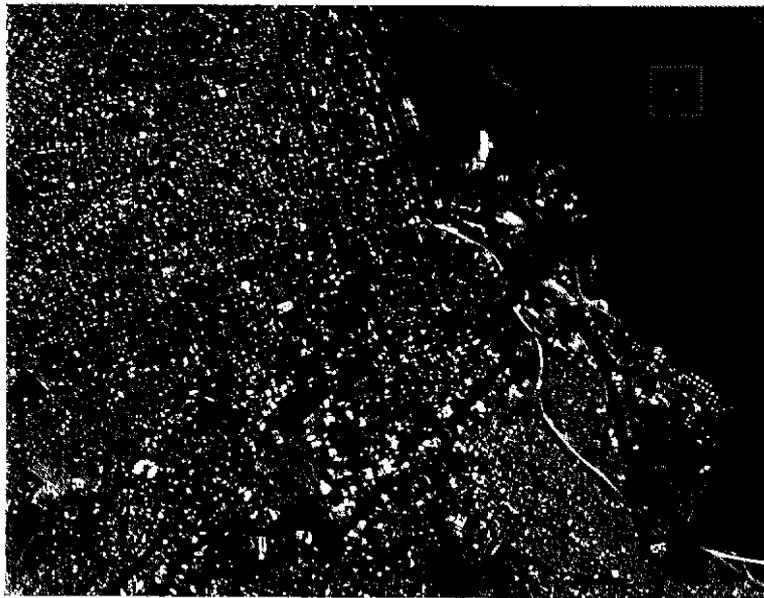


Figure 56 : Phénomène de masquage du système visuel.

II - Etude de la détection : tests visuels

L'objectif est ici de déterminer quelles sont les caractéristiques spatiales, fréquentielles, ou de contraste de l'image intervenant dans le mécanisme de détection visuelle de la singularité. Un protocole de tests visuels a été mis en place dans ce but.

A - Tests visuels

Une base importante d'images doit être constituée et présentée à des observateurs. La base est constituée de manière aléatoire :

- une image est tirée au sort dans un lot assez hétéroclite, constitué à l'avance ;
- un niveau de gris pour la singularité est tiré au sort ;
- la singularité est incrustée dans une image à une position aléatoire.

Les images utilisées sont des images de 512x512 pixels dont les niveaux de gris sont représentés par des valeurs entières codées sur un octet. Une transformation linéaire des

niveaux de gris permet un affichage à l'écran de l'image en modifiant la dynamique de visualisation de manière à ce qu'elle soit la même pour toutes les images. L'image est ensuite suréchantillonnée d'un facteur 2, par une interpolation bicubique, qui permet une présentation confortable à l'observateur tout en s'affranchissant de la fonction d'écran. Les images sont présentées pendant une durée de 15 s à l'observateur, qui, à l'aide de la souris, « clique » sur une éventuelle singularité. L'observateur n'a qu'un essai par image et les conditions d'observation sont imposées (mêmes conditions d'éclairage, même écran), de manière à laisser le moins de place possible au hasard.

Nous avons utilisé toutes les personnes disponibles pour réaliser les tests, soit une dizaine d'observateurs. Les résultats ont été considérés soit individuellement, soit en moyenne sur tous les observateurs. Etant donné la difficulté de mobiliser des observateurs pour un grand nombre de mesures, le comportement de l'un d'entre eux, appelé observateur n°1 (votre serviteur...) sera le seul dont le comportement individuel sera approfondi.

Les résultats des observations sont enregistrés de manière binaire, 1 pour un succès et 0 pour un échec, les différents observateurs étant différenciés. La probabilité de vision d'une singularité en fonction de l'un des paramètres que l'on aura choisi, sera la moyenne des notes de détection.

B - Recherche des paramètres significatifs

Il s'agit de rechercher les paramètres physiques qui peuvent avoir une influence sur la détection, ou l'absence de détection, de la singularité.

1) Le contraste local

a - Définition

Le contraste entre deux pixels voisins p_i et p_j de niveaux de gris n_i et n_j se définit par :

$$C_{p_i, p_j} = \frac{n_i - n_j}{n_i + n_j}$$

Sur une image $P(i, j)$, nous avons défini le contraste moyen autour du pixel (i, j) sur le voisinage V par :

$$\bar{C}(i, j) = \frac{1}{(2V + 1)^2 - 1} \sum_{k=i-V}^{i+V} \sum_{l=j-V}^{j+V} C_{p_{ij}, p_{kl}}$$

Lors de la transformation des niveaux de gris de l'image pour la visualisation, on a :

$$P(i, j) \rightarrow P'(i, j) = aP(i, j) + b$$

$$\text{avec } \begin{cases} a = \frac{M}{\text{Max}(P) - \text{Min}(P)} \\ b = -\frac{M * \text{Min}(P)}{\text{Max}(P) - \text{Min}(P)} \end{cases}$$

où M est le nombre maximum de niveaux de gris autorisés pour la visualisation.

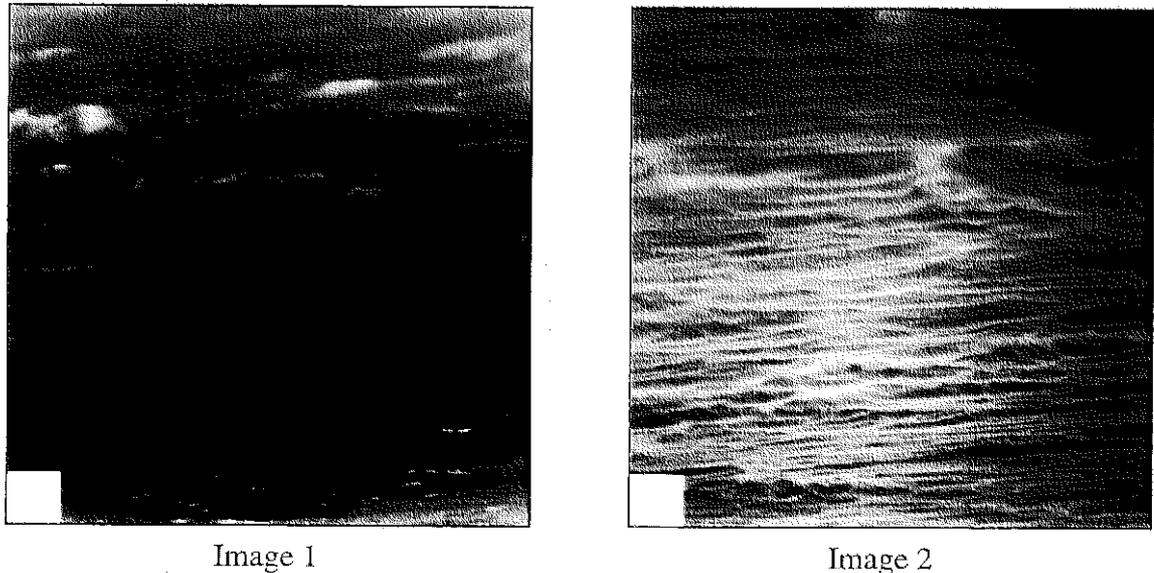
Le contraste devient donc :

$$C_{p_i, p_j} \rightarrow C'_{p_i, p_j} = \frac{n_i - n_j}{n_i + n_j - 2\text{Min}(P)} \quad (12)$$

Il ne dépend pas de M , mais uniquement des caractéristiques de l'image.

b - Influence sur la détection

Considérons le cas de 2 images différentes quant à leur contenu :



La première image contient de larges régions sombres et des zones claires localisées, alors que la seconde contient de larges régions claires.

Pour un observateur, la probabilité de détection d'une singularité en fonction de son contraste local, donnée par les tests visuels pour chacune des deux images, est représentée sur la Figure 57.

La différence entre les deux courbes pour d'importants contrastes positifs, c'est-à-dire pour des singularités sombres sur un fond plus clair, provient de la définition même du contraste : d'après l'équation (12), lorsque n_j tend vers $Min(P)$, le contraste tend vers 1. Autrement dit, un point sombre sur une zone sombre aura le même contraste, proche de 1, qu'un point sombre sur une zone claire. Le premier ne sera pas détecté alors que le second le sera. Ce phénomène sera appelé « ambiguïté sur le contraste » dans la suite du chapitre. Il sera donc intéressant de jumeler la mesure du contraste et la valeur du fond moyen sur la zone considérée.

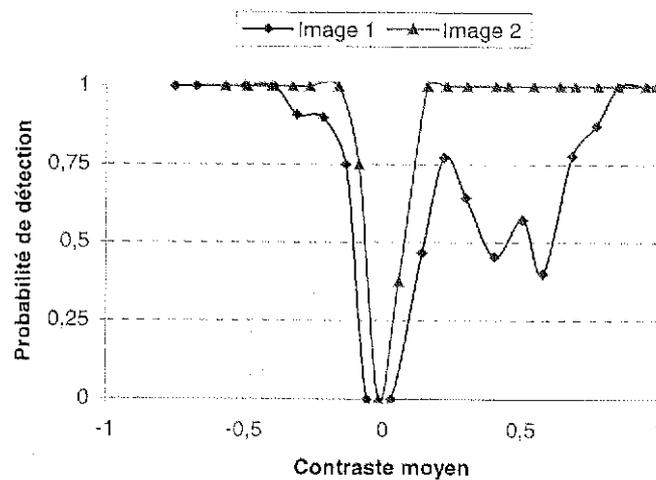


Figure 57 : Probabilité de détection d'une singularité en fonction du contraste par rapport au voisinage. Mise en évidence de l'ambiguïté sur le calcul du contraste.

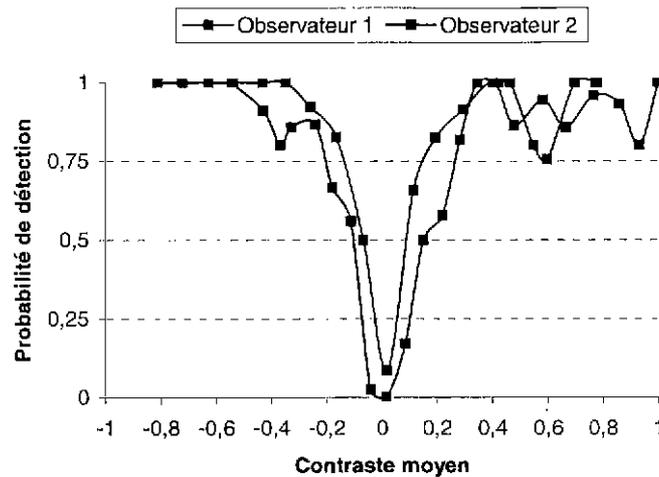


Figure 58 : Comparaison des résultats moyens de 2 observateurs pour la détection d'une singularité.

Notons aussi que la sensibilité au contraste diffère sensiblement d'un individu à un autre et que le seuil de détection donné par une probabilité de 0,5 varie en conséquence, comme le montre la Figure 58.

Globalement, il existe un seuil de contraste au delà duquel la probabilité de détection devient supérieure à 0,5. Ce seuil est de l'ordre de 0,2. Néanmoins, certains points de contraste inférieur sont détectés, alors que d'autres de contraste supérieur ne le sont pas. Dans toute la suite nous appellerons ces derniers des **points marginaux** pour la détection et nous essaierons de déterminer l'influence des paramètres autres que le contraste pour expliquer cette marginalité.

2) La distance de corrélation locale

Nous retiendrons comme information la distance de corrélation locale, déterminée à partir du variogramme local (cf. ANNEXE C -), calculé sur un voisinage de rayon 10 pixels autour de la singularité. Ce rayon a été choisi arbitrairement de manière à permettre un calcul statistique sur une zone suffisamment grande.

Les résultats obtenus sur un ensemble de 800 images, par un opérateur humain, sont matérialisés en portant sur un graphique les points marginaux :

- les cas de détection malgré un contraste local moyen faible (inférieur en valeur absolue à 0,1) ;
- les cas de non détection malgré un contraste local moyen suffisant (supérieur en valeur absolue à 0,1).

L'information portée en abscisse est le contraste local moyen, celle portée en ordonnée sera soit la distance de corrélation, soit la luminance moyenne locale.

Afin de faciliter les comparaisons de performances, les résultats seront présentés de la même façon pour deux modèles de simulation évalués dans la suite du chapitre.

Pour ces 800 images aux variogrammes sensiblement différents, les tests montrent que 13 % des singularités de contraste supérieur à 0,1 ne sont pas détectées par l'œil humain et que 12 % de celles qui ont un contraste inférieur à 0,1 le sont quand même. La Figure 59 montre le comportement de l'œil vis-à-vis de la distance de corrélation, uniquement pour les points considérés comme des marginalités.

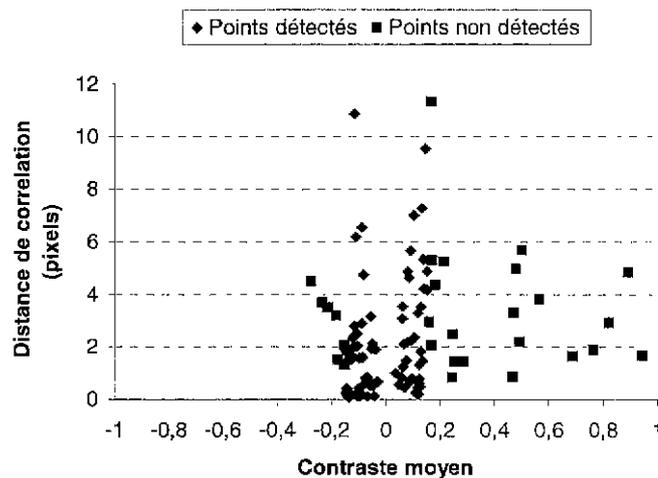


Figure 59 : Influences de la corrélation locale et du contraste sur la détection de l'observateur : points marginaux.

Contrairement à ce que l'on aurait pu attendre, nombre de points de faibles contrastes sont détectés alors que la distance de corrélation est faible, ce qui amène à conclure que les notions de distance de corrélation et de contraste local ne suffisent pas pour décrire le mécanisme de détection. On constate aussi que certaines singularités de contraste élevé ne sont pas détectées malgré une forte distance de corrélation (supérieure à 4 ou 5 pixels) : ceci se produit lorsqu'il y a ambiguïté sur le contraste (cf. 1)b -).

3) Les caractéristiques générales de l'image

Les caractéristiques de l'image pouvant intervenir dans la détection sont entre autres sa luminance moyenne et sa luminance maximale. Comme toutes les images sont visionnées avec la même dynamique de niveaux de gris, seule la luminance moyenne a une signification. Il semble même que la luminance locale est la seule qui puisse influencer la détection.

La Figure 60 montre, que les points de faible contraste sont détectés sur des zones claires (luminance locale forte), alors que des points de contraste local plus grand ne le sont pas sur des zones foncées (luminance locale faible). La moyenne a été calculée sur des voisinages plus ou moins grands ; il apparaît qu'un environnement de rayon 5 pixels autour de la singularité est celui qui donne la corrélation la plus significative entre le taux de détection et la moyenne.

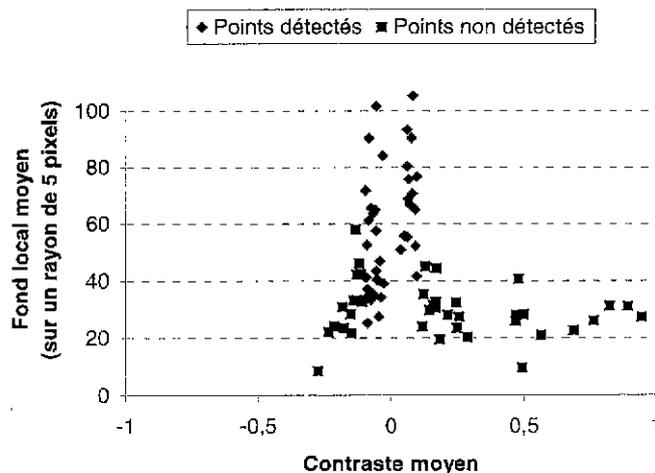


Figure 60 : Influences de la luminance locale et du contraste sur la détection de l'observateur : points marginaux

III - Adaptation du VDP pour la détection de singularités

Le VDP (cf. Chapitre 1 :III -B -1)) fonctionne de manière bivariante : il permet de donner une carte des différences visuelles entre deux images, mais en aucun cas de détecter des « points » sur une image. Dans cette partie, nous allons montrer que l'on peut tout de même l'adapter pour réaliser une détection de singularité.

A - Principe

Nous avons utilisé ce modèle de vision de la manière suivante, après avoir réalisé une première étape qui permet de ramener le problème à une situation bivariante (Figure 61) :

Etape 1 : La détection est faite « au mieux » : on suppose qu'il y a continuité en moyenne entre un pixel et ses voisins et tout pixel suspect (selon des critères de contrastes paramétrables) est retenu comme un candidat singularité.

Tous les candidats singularités sont ensuite effacés de l'image, c'est-à-dire mis au même niveau de gris que le voisinage moyen (8 encadrants).

Etape 2 : Le VDP compare l'image initiale et celle sur laquelle les candidats sont effacés pour ne retenir que ceux qui sont effectivement visibles par l'œil. Pour chacun des candidats singularité, la carte d'erreur donne la probabilité d'être vu.

Seuls les points détectés informatiquement et visibles à l'œil sont conservés par le VDP. On transforme donc une détection « au mieux » en une détection « humaine ». Il est entendu que le VDP réagit comme un observateur « standard » et sa conception rend difficile la prise en compte des caractéristiques visuelles de chacun. Il est cependant possible de le régler de manière à l'adapter à une population particulière d'individus.

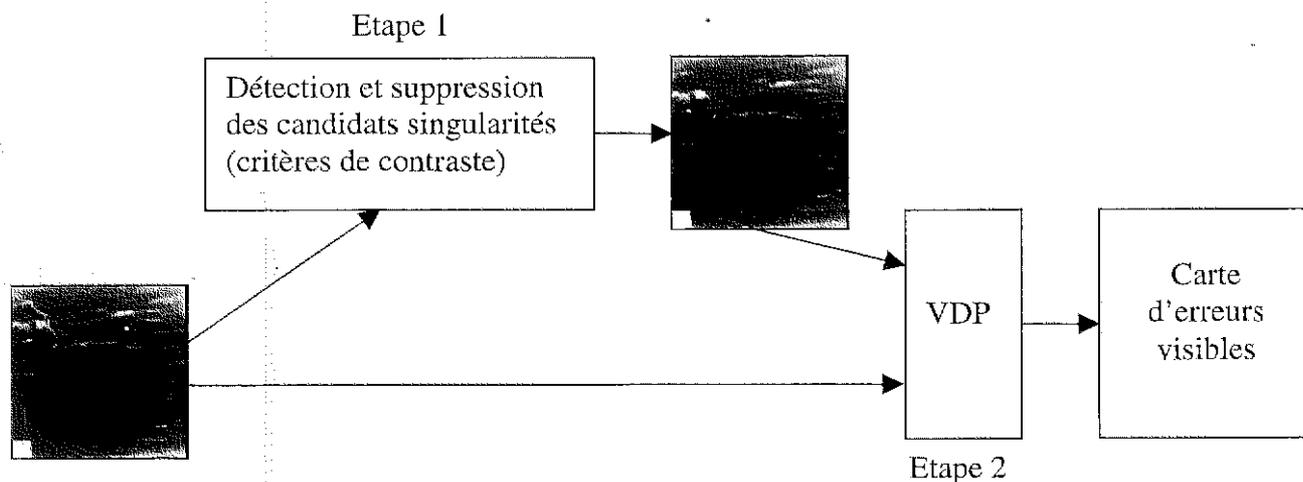


Figure 61 : Principe de l'adaptation du VDP à la détection d'un défaut ponctuel.

B - Résultats

La méthode a tout d'abord été testée sur l'image 2, ne présentant pas d'indétermination du contraste. La Figure 62 montre les résultats obtenus par le VDP avec un seuil de détection réglé de manière optimale pour avoir une réponse aussi proche que possible de celle de l'œil (minimisation de l'erreur quadratique moyenne).

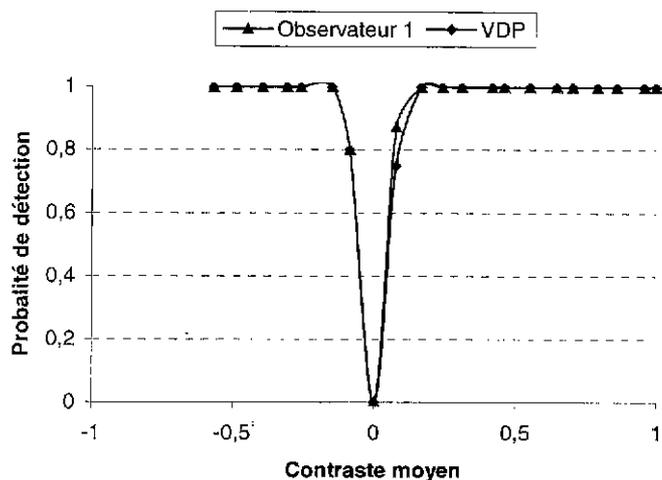


Figure 62 : Comportements comparés du VDP et de l'observateur sur l'image 2, pour laquelle il n'y a pas d'ambiguïté sur le contraste.

La Figure 63 montre maintenant le comportement du VDP sur l'image 1, lorsque le réglage du seuil est le même que pour l'image 2.

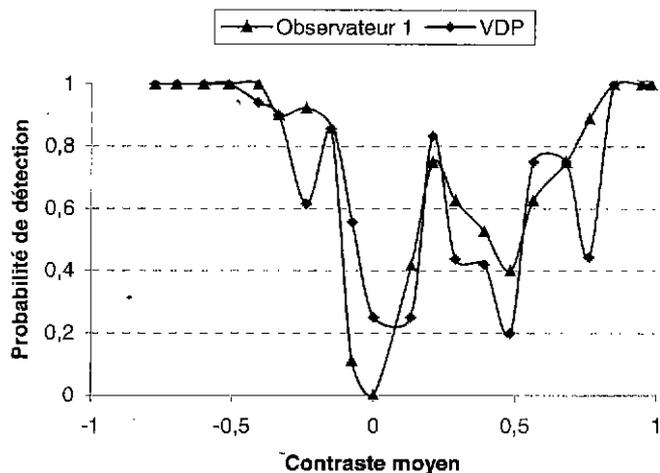


Figure 63 : Comportements comparés du VDP et de l'observateur sur l'image 1, pour laquelle il y a ambiguïté sur le contraste. Le VDP est réglé avec le même seuil que pour l'image 2.

Sur la zone critique (faible contraste), le VDP est plus performant que l'œil : ce n'est pas l'objectif recherché. La Figure 64 montre qu'un changement du seuil de détection du VDP permet de mieux reproduire le comportement de l'œil humain sur la zone correspondant aux faibles contrastes, mais on perd alors en précision pour des contrastes grands.

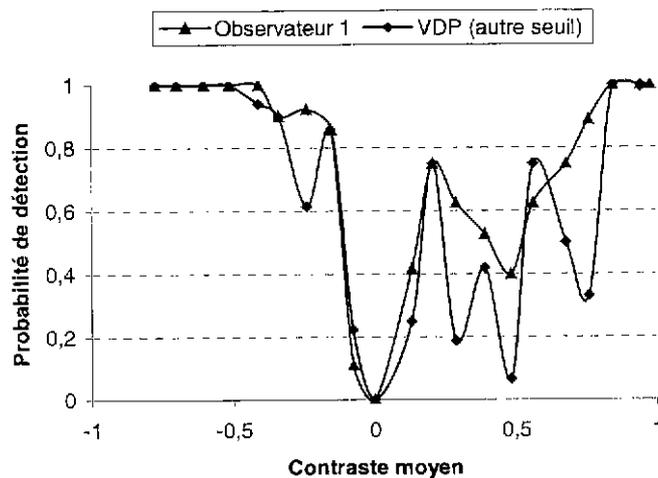


Figure 64 : Comportements comparés du VDP et de l'observateur sur l'image 1, pour laquelle il y a ambiguïté sur le contraste. Le seuil de détection du VDP est optimisé pour cette image.

Si l'on s'intéresse maintenant à l'ensemble des images et des observateurs, le VDP donne en moyenne un résultat très satisfaisant, comme le montre la Figure 65.

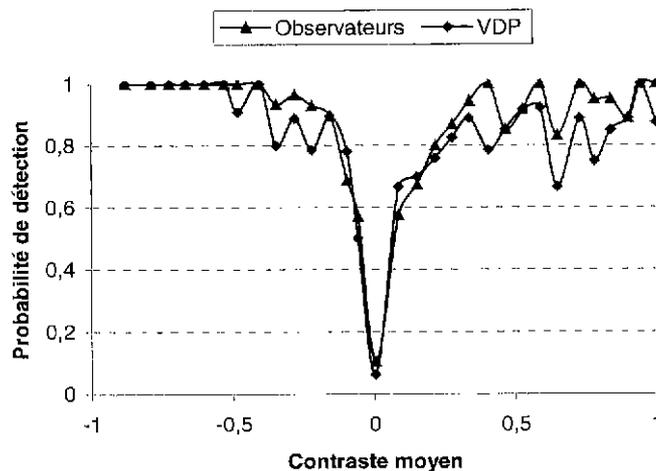


Figure 65 : Comportements comparés du VDP et des observateurs (en moyenne) sur l'ensemble des images.

Les Figure 66 et Figure 67 comparées aux Figure 59 et Figure 60 montrent cependant que le VDP, bien que globalement satisfaisant, réagit localement différemment de l'œil : 22 % des points de contraste inférieur à 0,1 sont détectés (contre 12 % pour l'œil) dont quelques points de contraste 0. Le comportement est sensiblement le même que celui de l'œil pour des contrastes importants.

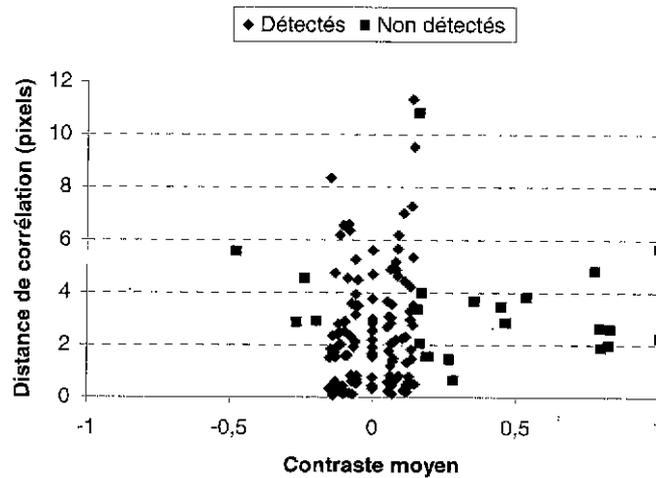


Figure 66: Influences de la corrélation et du contraste avec le voisinage sur la détection du VDP : points marginaux.

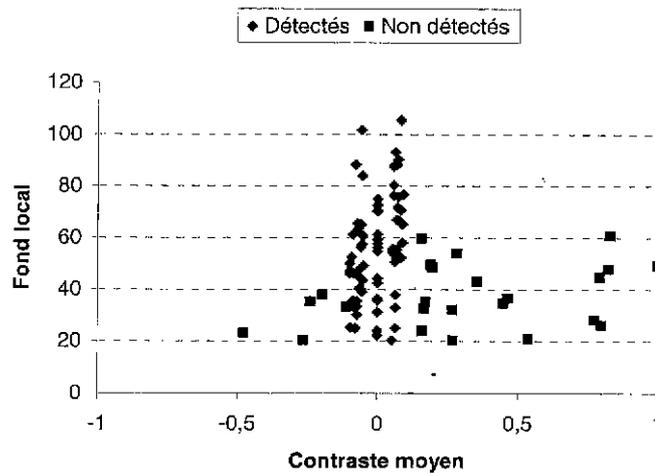


Figure 67 : Influences de la luminance locale et du contraste sur la détection du VDP : points marginaux.

C - Limitations

Il apparaît que le VDP, qui donne un résultat globalement satisfaisant sur un grand nombre d'images, peut avoir un comportement local sensiblement différent de celui du SVH. Ceci est certainement dû aux modèles de fonction de sensibilité au contraste et à la fonction d'amplification utilisés (cf. Chapitre 1 :III -B -1)). Les améliorations récentes apportées au VDP [CSB00] permettraient certainement d'améliorer ces résultats.

IV - Les réseaux de neurones

Dans cette partie, nous allons adapter la méthode univariante pour réaliser la détection de singularités sur l'image. Rappelons que l'on désire une détection représentative d'une détection visuelle, ni meilleure, ni pire.

Dans un premier temps nous avons réalisé une série de tests visuels de manière à déterminer quels sont les paramètres importants dans le mécanisme de détection. L'apprentissage est

ensuite réalisé sur les résultats des tests, le vecteur caractéristique étant constitué des paramètres significatifs.

A - Le vecteur caractéristique

Le vecteur caractéristique est constitué des grandeurs qui se sont révélées importantes lors des tests visuels. Ces grandeurs sont calculées jusqu'à des distances de 5 pixels de la supposée singularité, divers apprentissages ayant montré qu'une distance plus importante n'apportait aucune contribution significative à la simulation de la détection.

- 0) Niveau de gris du pixel ;
- 1) Fond moyen sur un voisinage de rayon 1 pixel ;
- 2) Fond moyen sur un voisinage de rayon 2 pixels ;
- 3) Fond moyen sur un voisinage de rayon 3 pixels ;
- 4) Fond moyen sur un voisinage de rayon 4 pixels ;
- 5) Fond moyen sur un voisinage de rayon 5 pixels ;
- 6, 7, ..., 13) Contrastes avec chacun des 8 pixels entourant le point étudié ;
- 14) Corrélacion moyenne avec un voisinage de rayon 1 pixel ;
- 15) Corrélacion moyenne avec un voisinage de rayon 2 pixels ;
- 16) Corrélacion moyenne avec un voisinage de rayon 3 pixels ;
- 17) Corrélacion moyenne avec un voisinage de rayon 4 pixels ;
- 18) Corrélacion moyenne avec un voisinage de rayon 5 pixels ;
- 19) Corrélacion avec le « carré » situé à une distance de 1 pixel ;
- 20) Corrélacion avec le « carré » situé à une distance de 2 pixels ;
- 21) Corrélacion avec le « carré » situé à une distance de 3 pixels ;
- 22) Corrélacion avec le « carré » situé à une distance de 4 pixels ;
- 23) Corrélacion avec le « carré » situé à une distance de 5 pixels.

B - Résultats

L'apprentissage a été réalisé sur 3000 exemples (cf Tests visuels Chapitre 4 : II - A -) constitués d'images dans lesquelles une singularité de niveau de gris aléatoire est incrusté à une position aléatoire. L'observateur dispose de 15 secondes et d'un unique essai pour détecter la singularité. S'il n'y a pas de détection dans ce délai, elle est considérée comme non gênante (réponse 0) et dans le cas contraire comme gênante (réponse 1). Comme précédemment, la probabilité de détection d'une singularité en fonction de l'une ou l'autre de ses caractéristiques est la moyenne des réponses obtenues pour cette caractéristique.

L'objectif est ici de simuler une détection visuelle, ni meilleure, ni pire. L'apprentissage a été réalisé pour plusieurs configurations de RN et testé sur 868 images différentes. Chaque image contenait une singularité plus ou moins visible. Lors de la phase d'apprentissage, le RN apprend à déterminer dans quelles conditions un point de l'image est une singularité gênante. Ensuite, lorsque le RN est utilisé pour rechercher les singularités, il scanne un à un tous les points de l'image et décide s'il est ou non une singularité.

Les configurations utilisées sont (24,22,8,1) pour les RN_1 et RN_2 et (24,8,1) pour le RN_3. Elle doivent permettre de mesurer le gain apporté, ou non, par un RN de taille importante. La différence entre le RN_1 et le RN_2 tient uniquement à l'initialisation des poids du RN avant l'apprentissage. Sur un RN de taille importante, l'initialisation (aléatoire) peut influencer d'une façon non négligeable sur la convergence de l'algorithme d'apprentissage.

Le Tableau 10 synthétise les résultats obtenus sur les 868 images pour les différentes configurations de RN.

| | Nombre de pixels détectés | Nombre de pixels non détectés |
|--------------------|---------------------------|-------------------------------|
| Observateur | 582 | 286 |
| RN_1 | 601 | 267 |
| RN_2 | 600 | 268 |
| RN_3 | 608 | 260 |
| RN_1 & observateur | 566 | 251 |
| RN_2 & observateur | 574 | 260 |
| RN_3 & observateur | 571 | 249 |

Tableau 10 : Comparaison des résultats obtenus par différentes configurations de RN pour la détection d'une singularité sur l'image.

On considère qu'il y a erreur de détection dès que le RN propose une détection différente de l'observateur. Le RN_1 a commis 5,8% d'erreurs, le RN_2 en a commis 3,9% et le RN_3 5,5%. Même si les résultats sont généralement bons et à peu près comparables d'une configuration à une autre, on constate néanmoins que, si l'apprentissage est réalisé sur un très grand nombre d'exemples comme c'est le cas ici, on peut gagner en précision en utilisant un RN de taille importante. A ce moment là, il faut réaliser plusieurs apprentissages pour plusieurs inialisations différentes et garder le meilleur.

La Figure 68 montre, en fonction du contraste local moyen, les probabilités de détection comparées pour le VDP (meilleur réglage du seuil), le meilleur RN (RN_2) et l'observateur. Elle permet de constater que le seuil de détection à faible contraste est le même pour l'observateur que pour le RN. Le VDP reste quant à lui plus sensible que l'œil. Le plus grand nombre d'erreurs commises par le VDP et le RN se situe au seuil de détection lorsqu'il y a ambiguïté sur le contraste, c'est-à-dire pour des singularités sombres sur un fond sombre.

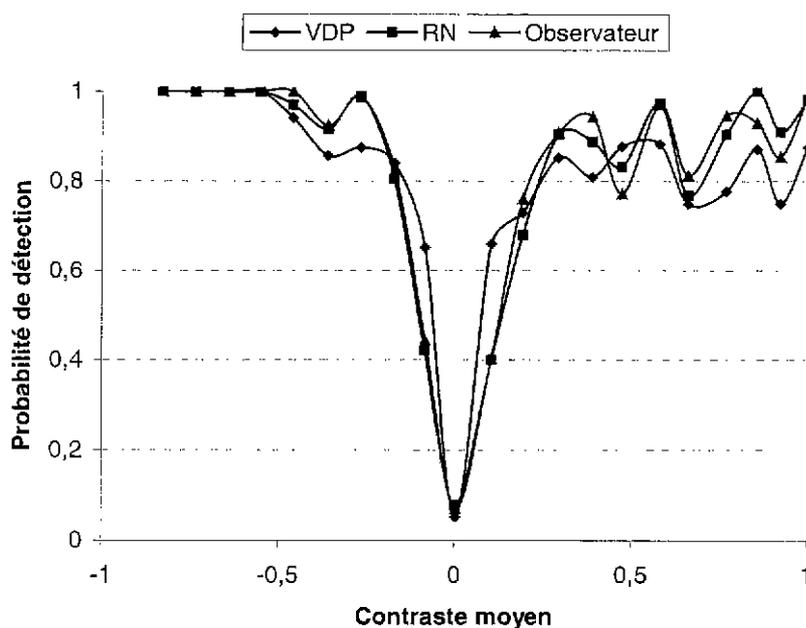


Figure 68: Comportement comparé VDP-RN-observateur.

La Figure 69 comparée à la Figure 59 montre que la structure du fond influence la décision du RN sensiblement de la même façon que l'œil.

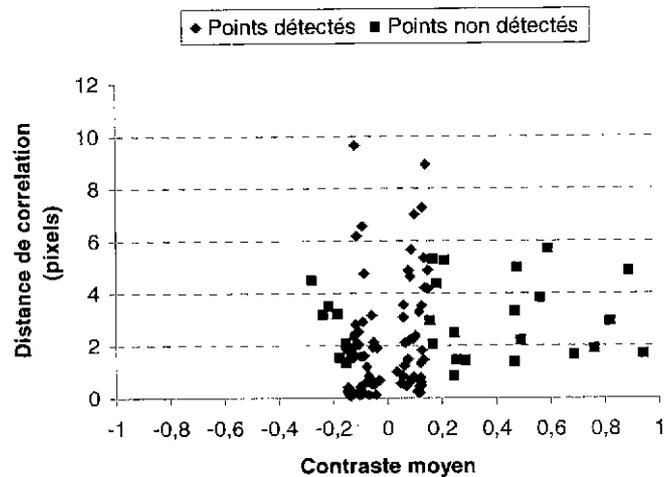


Figure 69 : Influences de la corrélation locale et du contraste avec le voisinage sur la détection du RN : points marginaux.

Le RN est moins sensible à la luminance de fond que le VDP comme le montre la Figure 70, comparée à la Figure 59.

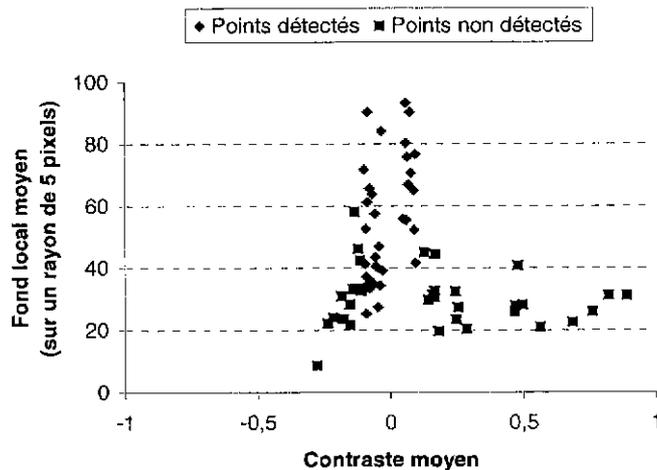


Figure 70 : Influences de la luminance locale et du contraste sur la détection du RN : points marginaux.

Notons que le RN n'a jamais détecté une singularité qui n'était pas présente sur l'image.

C - Etude de sensibilité

Pour la détection d'un défaut ponctuel, la sortie du RN ne peut prendre que deux valeurs : 0 en cas de non détection et 1 en cas de détection. La sensibilité du RN à la $j^{\text{ième}}$ composante du vecteur caractéristique est calculée de la manière habituelle (cf. Chapitre 2 :III -C -) en faisant varier sur son intervalle de définition, $[-1 ; 1]$, chacune des autres composantes restant constante à une valeur moyenne correspondant à une absence de défaut. La sensibilité est la déviation maximale observée en sortie du RN.

Si $(x_0, \dots, x_j, \dots, x_p)$ sont les composantes moyennes du vecteur caractéristique lorsque aucune singularité n'est présente et $f(x_0, \dots, x_j, \dots, x_p)$ est la sortie du RN correspondant à ce vecteur, la sensibilité $S(j)$ vis-à-vis de la $j^{\text{ième}}$ composante est définie par :

$$S(j) = \max_{x_j \in [-1,1]} \left[f(x_0, \dots, x_j', \dots, x_p) - f(x_0, \dots, x_j, \dots, x_p) \right]$$

Pour chaque composante, cette sensibilité est calculée en moyennant un grand nombre d'exemples. Elle permet de montrer quelles sont, parmi les composantes du vecteur caractéristique, celles qui peuvent transformer un pixel non détecté en un pixel détecté et réciproquement.

Les sensibilités pour les 3 configurations de RN utilisées sont présentées Figure 71. On constate que les sensibilités vis-à-vis des composantes dépendent de la configuration du RN utilisé. Ce n'était généralement pas le cas pour les dégradations précédemment étudiées, uniformes sur l'image (flou, JPEG). Il faut donc rester prudent dans l'explication de leur importance physique pour la modélisation du phénomène de détection. Plusieurs causes peuvent expliquer cette différence :

1) La convergence du modèle

Lors de la phase d'apprentissage, le RN se place dans une configuration minimisant l'erreur entre la sortie attendue et la sortie calculée. Cette minimisation est réalisée dans un espace à N dimensions (24 dans le cas qui nous intéresse ici). Le RN peut converger vers un minimum local, correspondant à des sensibilité vis-à-vis des composantes qui ne sont pas forcément toutes évidentes à interpréter, mais qui, par le jeu des redondances d'information qu'elle peuvent contenir, donnent au RN suffisamment d'information pour calculer la sortie. Plusieurs configurations peuvent donner des résultats identiques en moyenne.

2) La taille du RN

Un RN contenant peu de neurones, comme le RN_3, a une capacité d'apprentissage moins importante qu'un RN contenant beaucoup de neurones. Pour fournir la meilleure généralisation possible, les composantes fortement corrélées avec la sortie attendue prennent un poids très important devant les autres. Ce n'est pas le cas pour un RN de taille importante qui peut apprendre de nombreuses exceptions.

On constate néanmoins que parmi toutes les composantes, les contrastes avec les 8 pixels entourant le pixel considéré ont toujours une grande importance.

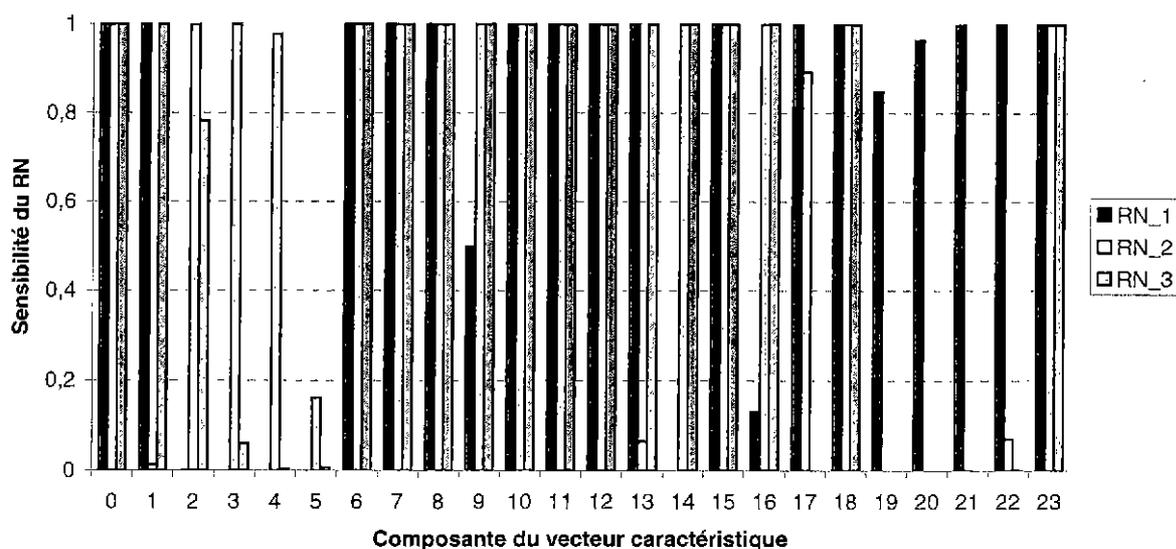


Figure 71 : Sensibilité vis-à-vis des composantes du vecteur caractéristique.

V - Conclusion

Nous avons montré dans cette partie qu'il était possible de simuler la détection visuelle d'une singularité sur une image. La comparaison d'une méthode de détection bivariante adaptée au problème, qui repose sur un modèle visuel complexe, et de la méthode univariante, prouve l'aptitude de cette dernière à reproduire le comportement du système visuel. Elle permet non seulement détecter les singularités avec une fiabilité supérieure à 95 %, mais le temps de calcul nécessaire est environ 100 fois plus faible qu'avec le VDP, qui calcule 33 FFT et 33 FFT inverses.

Ces résultats semblent prouver qu'il est possible de calibrer un modèle, à partir d'un nombre réduit de paramètres bien choisis, pour modéliser un comportement visuel simple.

La méthode a été mise en application sur des images SPOT contenant des OVNIS (cf. ANNEXE E -A4 -1)). Bien calibrée, elle permet de ne détecter que ceux qui sont visuellement gênants.

Chapitre 5 :
Application à une image dégradée par un
mélange de deux défauts

I - Introduction

Après avoir appliqué la méthode à des images dégradées par un défaut unique, nous allons montrer, dans un cas de mélange assez simple, que l'on peut retrouver les caractéristiques de chacun des défauts si l'on connaît le système dégradant, c'est-à-dire l'origine et la nature des dégradations.

La Figure 72 donne l'exemple d'une image dégradée par une combinaison « défaut de mise au point + compression JPEG », dans cet ordre. Les notes de qualité JPEG et de qualité flou sont définies de la manière suivante (cf. Chapitre 3 :) :

- note flou = 10 - nombre de pas défocalisation
- note JPEG = (facteur de qualité)/10

Une note 10 correspond à une dégradation inexistante vis-à-vis du défaut considéré, une note 0 correspond à une dégradation maximale dans l'intervalle considéré.

| Qualité JPEG \ Qualité flou | 9 | 7 | 5 | 3 | 1 |
|-----------------------------|---|---|---|---|---|
| 9 |  |  |  |  |  |
| 7 |  |  |  |  |  |
| 5 |  |  |  |  |  |
| 3 |  |  |  |  |  |
| 1 |  |  |  |  |  |

Figure 72 : Image dégradée par une combinaison « défaut de mise au point + compression JPEG »

Si l'on ne connaît pas le système dégradant, on peut considérer le mélange de dégradations comme une nouvelle dégradation, et la caractériser sur une échelle arbitraire (visuelle par

exemple). Un apprentissage est alors possible mais la distinction entre les différentes dégradations est perdue.

Nous nous placerons dans un premier temps dans le cas où l'on connaît le système dégradant, « défaut de mise au point + compression JPEG » et nous allons montrer dans quelle mesure on peut retrouver les caractéristiques de chacune des dégradations.

Dans un deuxième temps, nous montrerons comment l'apprentissage peut être réalisé si le système dégradant n'est pas connu.

II - Cas où le système dégradant est connu

Nous nous intéressons donc à partir de maintenant à des images dégradées par un défaut de mise au point suivi d'une compression JPEG. Le défaut de mise au point peut varier de 0 à 9 pas de focalisation (image parfaite à image très floue). Nous avons utilisé le modèle SPOT1 pour le reproduire (cf. Chapitre 3 :II -B -1)). La compression peut avoir un facteur de qualité variant de 100 à 0 (image parfaite à image très comprimée). Toutes les proportions sont considérées. L'objectif est de retrouver ces proportions sur l'image dégradée.

A - Une première approche

Il paraît naturel, dans un premier temps, d'appliquer telles quelles les méthodes développées pour l'évaluation du pas de focalisation et du facteur de qualité. On s'attend à ce que l'estimation du facteur de qualité de la compression, dégradation qui intervient la dernière, soit bonne. En revanche il est difficile de prévoir comment le flou de l'image sera apprécié après la compression. Il devrait néanmoins être estimé correctement sur des images faiblement comprimées.

La Figure 73 montre les résultats obtenus lors de l'estimation du pas de focalisation sur les images avec le RN qui avait été étalonné lors de l'étude correspondante (cf. Chapitre 3 :II -). Le défaut de mise au point varie de 1 à 9 (qualité flou de 9 à 1) et le facteur de qualité varie de 90 à 10 (qualité JPEG de 9 à 1).

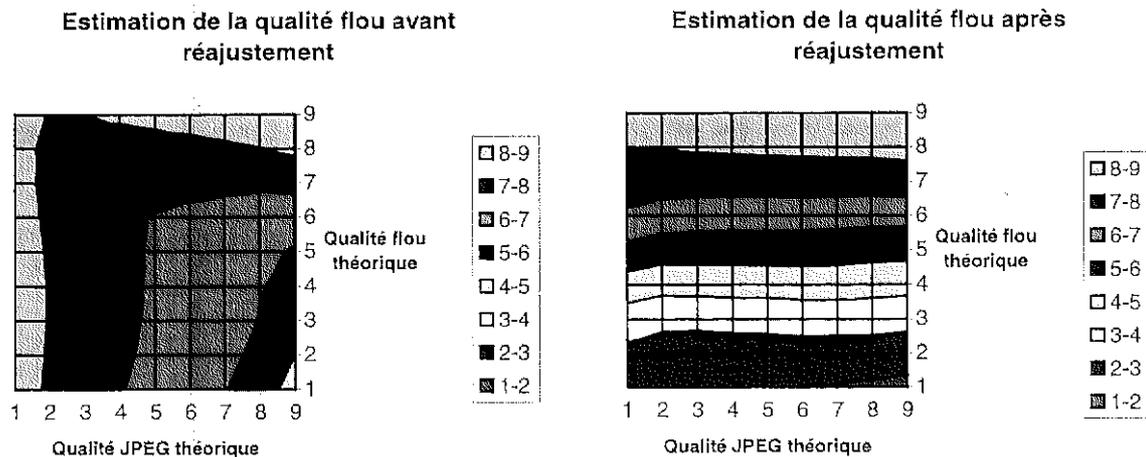


Figure 73 : Estimation du défaut de mise au point avec le RN spécifique, sans correction. La figure de gauche montre la note flou estimée en fonction de la qualité JPEG théorique et de la qualité flou théorique, la figure de droite montre l'écart moyen à la note théorique en fonction de la qualité JPEG théorique et de la qualité flou théorique.

Les estimations obtenues montrent effectivement que le flou est estimé correctement lorsque la qualité JPEG est supérieure à 8-9. En revanche, pour des compressions plus importantes, le RN devient incapable d'estimer le flou et commet une erreur d'autant plus importante que la

compression est forte (qualité JPEG faible). L'estimation du flou semble même ne dépendre que de la qualité JPEG dès lors que le rapport (qualité flou)/(qualité JPEG) devient faible (cf. Figure 73 gauche).

La Figure 74 montre, en revanche, que l'estimation de la qualité JPEG est pertinente.

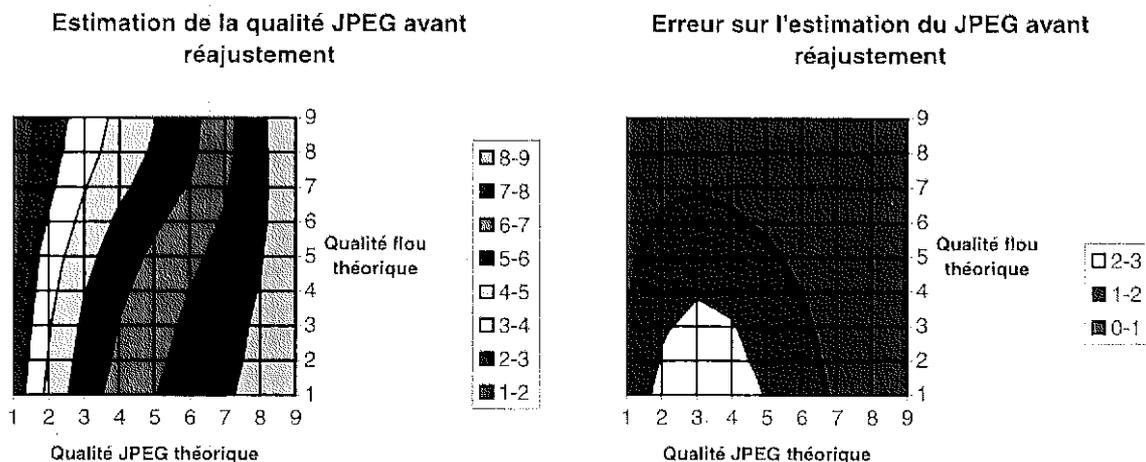


Figure 74 : Estimation de la qualité de compression avec le RN spécifique, sans correction. La figure de gauche montre la note JPEG estimée en fonction de la qualité JPEG théorique et de la qualité flou théorique, la figure de droite montre l'écart moyen à la note théorique en fonction de la qualité JPEG théorique et de la qualité flou théorique.

Ce résultat est conforme à celui qui était attendu car la compression intervient après le défaut de mise au point. On constate néanmoins un biais de plus en plus important lorsque la qualité flou diminue. Le RN qui est utilisé ici présentait déjà un biais plus important pour les images uniformes lorsqu'il avait été calibré (cf. Chapitre 3 :I -D -1)). C'est certainement ce qui explique la perte de précision dans l'estimation de la qualité JPEG quand le flou devient important. Comme le montre la Figure 75, quand le flou devient fort, le coefficient FF_8 du vecteur caractéristique JPEG (prépondérant dans l'estimation) diminue et donc la qualité JPEG est surestimée. En fait, quand une image relativement uniforme est comprimée fortement en JPEG, de nombreux blocs adjacents sont identiques. Comme le contraste entre les blocs adjacents diminue, FF_8 diminue aussi.

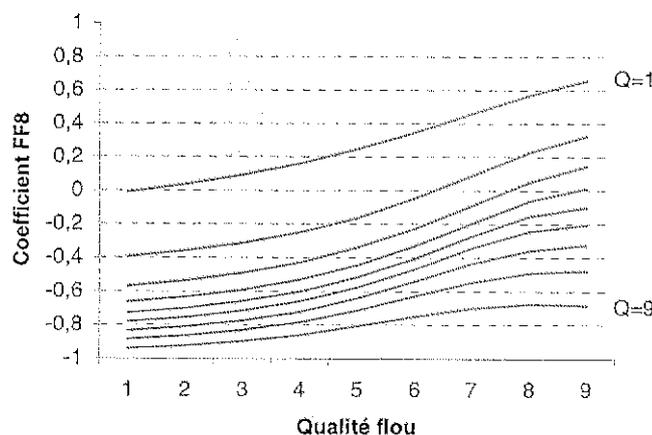


Figure 75 : Evolution du coefficient FF_8 du vecteur caractéristique JPEG en fonction de la qualité flou, pour une qualité JPEG donnée.

B - Réajustement des notes par un nouvel apprentissage : cas où le système dégradant est connu

Utilisées seules, les qualités flou et JPEG estimées avec les RN calibrés pour des dégradations uniques sont biaisées et ne permettent pas de retrouver les qualités réelles. Nous avons donc réalisé un nouvel apprentissage pour tenter d'approcher ces dernières plus précisément. Le vecteur caractéristique utilisé est constitué des deux notes estimées par le RN spécifiques et de toutes les composantes des vecteurs que nous avons définis pour estimer la qualité d'une image floue et celle d'une image JPEG. Les redondances ont néanmoins été supprimées :

- les composantes 0 et 1 sont les notes de qualité flou et JPEG estimée avec les RN spécifiques ;
- les composantes 2 à 29 sont les 28 composantes du vecteur caractéristique JPEG (cf. Chapitre 3 :I -C -) ;
- les composantes 30 à 35 sont les composantes fréquentielles (8 à 13) du vecteur caractéristique flou (cf. Chapitre 3 :II -B -3)), les autres ayant déjà été prises en compte.

Le RN utilisé comporte deux sorties, une pour l'évaluation de la qualité flou, l'autre pour l'évaluation de la qualité JPEG. L'apprentissage a été réalisé sur les images de la base générale, pour des qualités flou variant de 1 à 9 et des qualités JPEG variant de 1 à 9.

Les résultats moyens évalués sur la base de test sont présentés Figure 76 et Figure 77.

En ce qui concerne l'évaluation de la qualité JPEG, ils sont tout à fait comparables à ceux que l'on avait obtenus dans le cas d'une dégradation unique (cf. Chapitre 3 :I -). L'erreur absolue sur l'ensemble des images de la base de test ne dépasse pas 0,5 points de qualité.

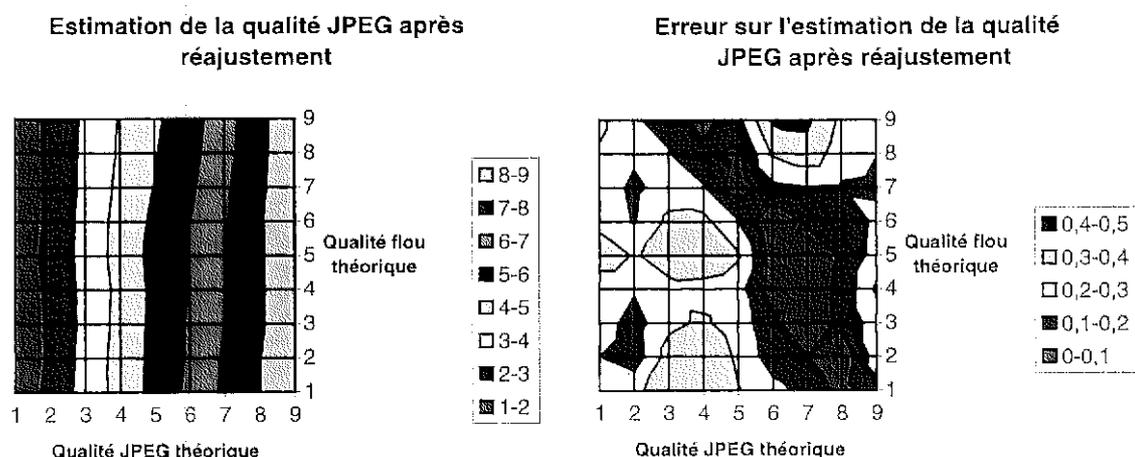


Figure 76 : Estimation de la qualité JPEG après correction. La figure de gauche montre la note JPEG estimée en fonction de la qualité JPEG théorique et de la qualité flou théorique, la figure de droite montre l'écart moyen à la note théorique en fonction de la qualité JPEG théorique et de la qualité flou théorique.

En revanche, l'estimation du pas de focalisation a été considérablement améliorée par rapport à celle que l'on obtenait sans la correction (Figure 77), et est très proche de la valeur théorique jusqu'à une qualité de compression assez faible, de l'ordre de 2 ou 3. Lorsque la compression existe, la précision dans l'estimation du flou reste tout de même légèrement moins bonne que celle que l'on avait obtenue pour le flou seul (cf. Chapitre 3 :II -). Au dessous d'une note de qualité JPEG de 2 ou 3, l'image est tellement comprimée que toute l'information a disparu. De plus le RN atteint sa limite d'extrapolation. Le large domaine pour

lequel l'estimation est correcte permet une précision absolue inférieure au demi pas de focalisation. L'erreur maximale est encore inférieure au pas dans la zone intermédiaire.

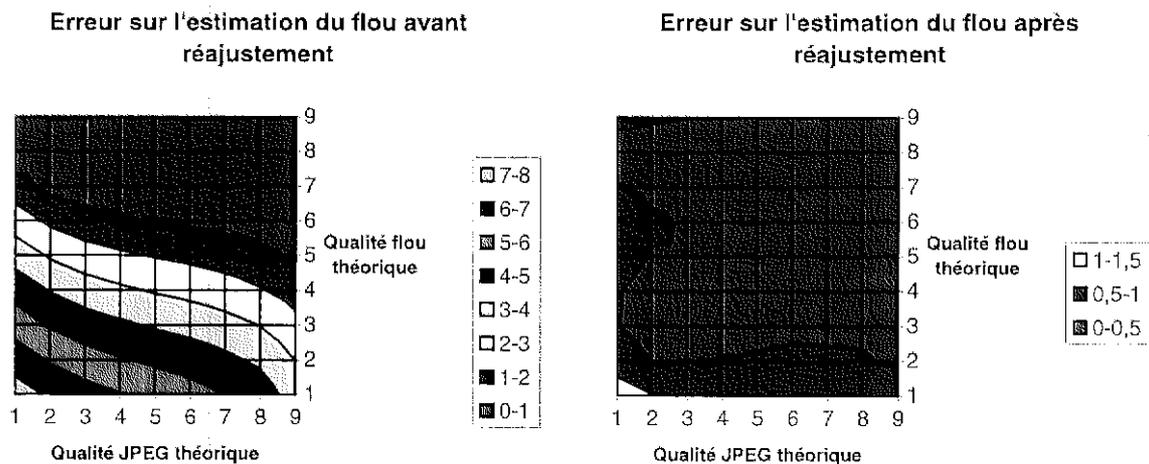


Figure 77 : Estimation du défaut de mise au point après correction. La figure de gauche montre la note JPEG estimée en fonction de la qualité flou théorique et de la qualité flou théorique, la figure de droite montre l'écart moyen à la note théorique en fonction de la qualité JPEG théorique et de la qualité flou théorique.

C - Etude de sensibilité

L'étude de sensibilité est réalisée de manière habituelle sur chacune des deux sorties du RN. Les sensibilités aux composantes du vecteur caractéristique pour chacun des deux cas sont présentées Figure 78 et Figure 79.

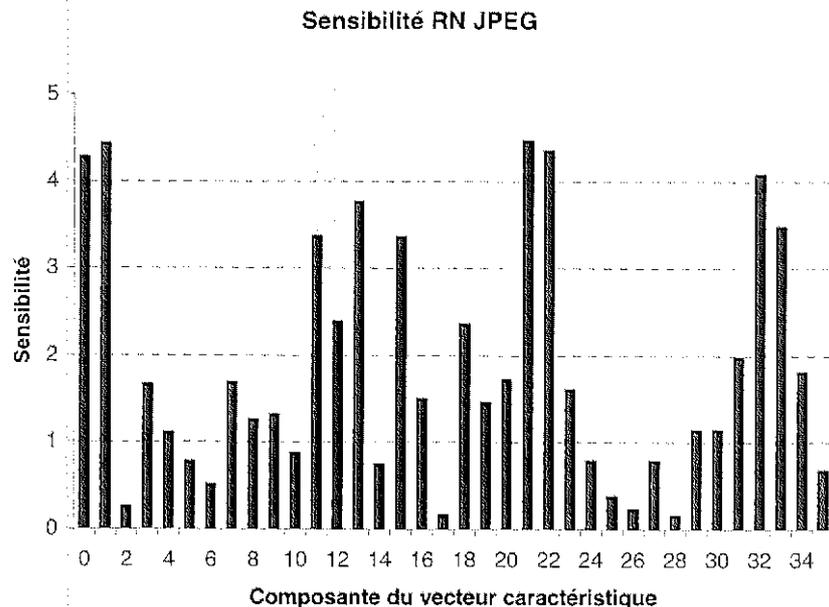


Figure 78 : Sensibilité de l'estimation de la qualité JPEG, après correction, vis-à-vis des composantes du vecteur caractéristique.

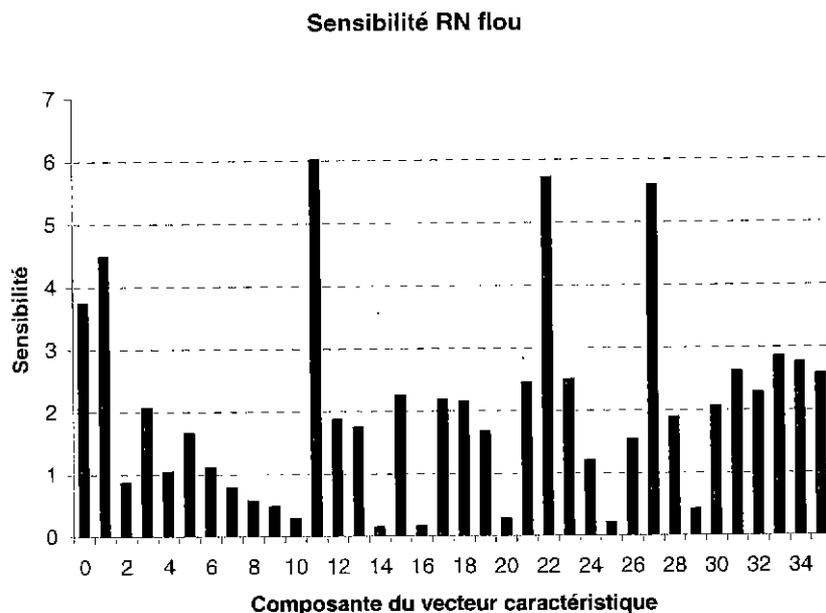


Figure 79 : Sensibilité de l'estimation de la qualité flou, après correction, vis-à-vis des composantes du vecteur caractéristique.

On constate que dans chacun des deux cas, les notes estimées avec les RN spécifiques (composantes 0 et 1) ont une grande importance. Néanmoins, elles ne sont pas suffisantes à elles seules pour permettre de retrouver les notes réelles. Ce point est d'ailleurs confirmé visuellement sur les Figure 73 et Figure 74.

On remarque aussi que les sensibilités vis-à-vis des composantes dans l'évaluation de la qualité JPEG sont sensiblement les mêmes que celles qui avaient été obtenues lors de l'apprentissage spécifique (cf. Chapitre 3 :I -). Les composantes fréquentielles, importantes elles aussi, permettent certainement au RN de décorréler le flou et la compression.

En ce qui concerne la note de qualité flou, on constate que les principales contributions sont les mêmes que pour l'évaluation de la note JPEG. Elles permettent certainement, dans ce cas aussi, de décorréler les deux dégradations. L'importance de la composante n°27 (variance de $V(k)$ du vecteur caractéristique JPEG), peut être expliquée parce qu'elle apporte une information sur le contenu fréquentiel à l'intérieur des blocs (cf. Chapitre 3 :I -C -1)).

On peut trouver curieux que les composantes 0 et 1, qui sont les notes estimées par les RN spécifiques, aient une importance aussi grande. En effet elles sont elles mêmes estimées à partir des autres et peuvent paraître a priori redondantes. Un apprentissage réalisé sans ces composantes montre une dégradation des estimations dès lors que l'un des deux défauts est très important (Figure 80). Le biais sur l'estimation de la qualité JPEG est inversé par rapport à celui que l'on avait trouvé dans l'apprentissage initial (Figure 74).

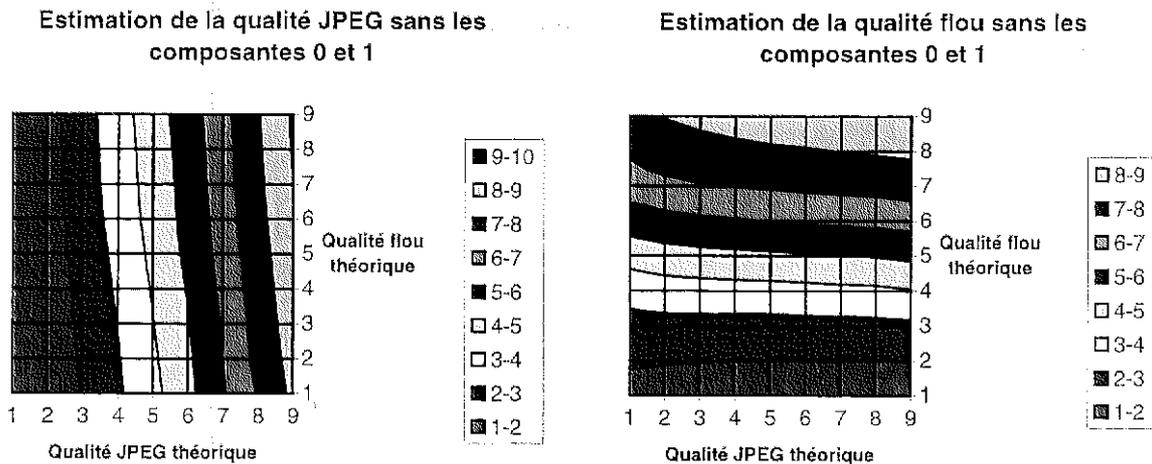


Figure 80 : Estimation des notes de qualité flou et JPEG après suppression des composantes 0 et 1 du vecteur caractéristique. La figure de gauche montre la note JPEG estimée en fonction de la qualité flou théorique et de la qualité JPEG théorique, la figure de droite montre la note flou estimée en fonction de la qualité JPEG théorique et de la qualité flou théorique.

Ces deux composantes, calculées par des RN spécifiques pour lesquels les apprentissages ont été différents, sont en fait indépendantes des autres composantes.

III - Cas où le système dégradant n'est pas connu

On suppose maintenant que l'on ne connaît pas la dégradation présente dans l'image. On peut se ramener au cas d'une dégradation unique en considérant celle-ci comme une nouvelle dégradation. S'il est possible de générer plusieurs images avec des niveaux de dégradation différents, on peut les noter, visuellement par exemple, pour réaliser un apprentissage.

A - Notation visuelle, modélisée par le VDP

Reprenons le cas du mélange flou + JPEG. Pour ne pas avoir à faire noter plusieurs milliers d'images par des observateurs, nous avons utilisé le VDP (cf. Chapitre 1 :III -B -1)). Le VDP propose une carte des différences visibles entre l'image dégradée et l'image parfaite. Nous avons fabriqué une note à partir de cette carte en additionnant les probabilités de détection de chacune de ces erreurs visibles. Les notes du VDP sont ensuite ramenées sur l'intervalle [0,10].

L'apprentissage est réalisé avec le même vecteur caractéristique que précédemment sans toutefois y inclure les composantes 0 et 1, composantes représentant les notes JPEG et flou estimées avec les RN spécifiques, puisqu'on suppose qu'on ne connaît pas le système dégradant :

- les composantes 0 à 27 sont celles du vecteur caractéristique JPEG (cf. Chapitre 3 :I -C -3)) ;
- les composantes 28 à 33 sont les composantes fréquentielles (8 à 13) du vecteur caractéristique flou (cf. Chapitre 3 :II -B -3)).

La Figure 81 montre que la précision dans l'estimation de la note de l'observateur (modélisé par le VDP) est très bonne : l'erreur est en moyenne de 4% sur les images de la base de test.

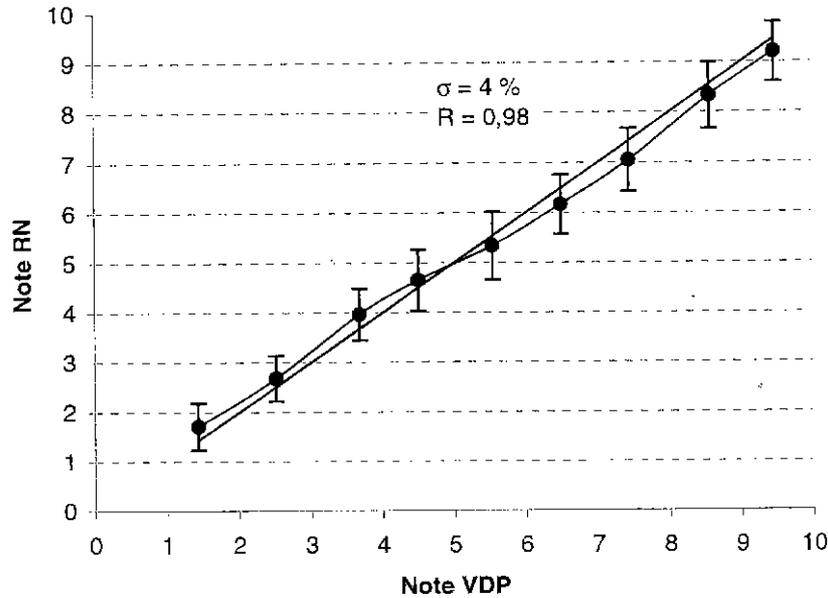


Figure 81 : Cas où l'on considère la combinaison de dégradation comme une nouvelle dégradation, unique : notation par un modèle représentatif d'une notation visuelle, le VDP.

B - Notation par calcul de distance à l'image parfaite

Les images sont maintenant notées différemment : la note de qualité est donnée sous forme de distance :

$$qualité = \frac{\sqrt{(9 - note_flou)^2 + (9 - note_jpeg)^2}}{\sqrt{9^2 + 9^2}}$$

Cette notation ne représente rien de concret, elle nécessite de plus de connaître les notes théoriques ce qui, on le suppose dans cette partie, n'est pas le cas. Néanmoins l'objectif est de montrer que l'on peut reproduire de manière univariante une notation de qualité différente de la notation visuelle précédente. La Figure 82 montre que dans ce cas, comme dans le précédent, les résultats obtenus par le RN sur la base de test sont très bons.

Il est donc possible de considérer la combinaison des dégradations comme un nouvelle dégradation, unique, et adopter un système de notation visuel ou autre, tel qu'on l'avait déjà fait précédemment.

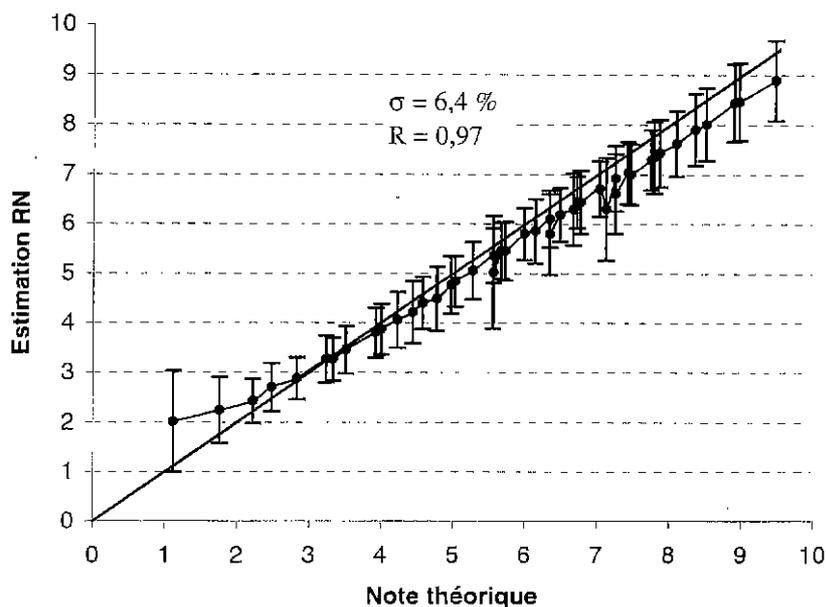


Figure 82 : Cas où l'on considère la combinaison de dégradation comme une nouvelle dégradation, unique : notation par calcul de distance à l'image parfaite.

C - Etude de sensibilité

L'étude de sensibilité, montre des tendances différentes pour la notation visuelle et la notation mathématique.

Les sensibilités obtenues pour l'évaluation visuelle de la qualité sont à peu près les mêmes que celles qui avaient été obtenues lors des études spécifiques, pour des défauts non mélangés (cf. Chapitre 3 :I -D -3) et Chapitre 3 :II -B -5)), comme le montre la Figure 83. On peut penser que le RN utilise les informations du vecteur caractéristique JPEG pour décorréler la compression et le flou. Ensuite les composantes fréquentielles du vecteur caractéristique du flou permettent d'en évaluer l'importance. Sur des images dégradées de cette manière (cf. Figure 72), l'œil fait aussi la différence entre la contribution du flou et celle de la compression.

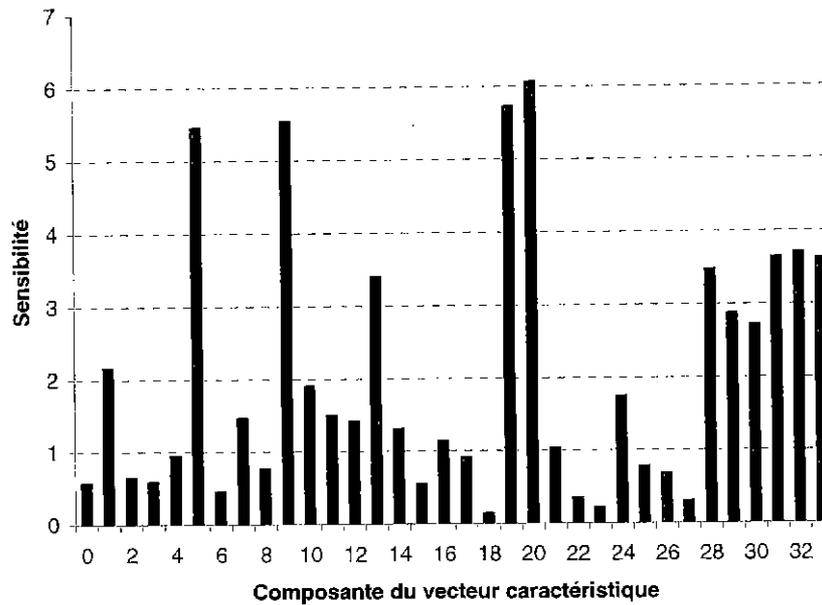


Figure 83 : Sensibilité vis-à-vis des composantes du vecteur caractéristique dans le cas où la combinaison de dégradations est considérée comme un défaut unique. Notation représentative d'une perception visuelle de la qualité.

En revanche, pour la notation mathématique, le RN privilégie essentiellement le coefficient D du variogramme (composante n°7 du vecteur caractéristique JPEG) qui représente la corrélation à courte distance sur l'image) et la valeur de FF_8 calculée sur l'image (composante n°22) qui caractérise les structures de période 8 pixels contenues dans l'image). Les autres composantes interviennent ensuite, sensiblement de la même manière (cf. Figure 84). Etant donné que la notation considérée ici n'a plus rien de visuel, l'explication de ce comportement est délicate. Ce n'est pas un point de convergence particulier du RN car plusieurs apprentissages donnent la même sensibilité. De plus, bien que la composante n°7 caractérise l'uniformité de l'image alors que la composante n°22 caractérise les blocs, elles n'avaient ni l'une ni l'autre d'importance pour caractériser chacune des dégradations individuellement.

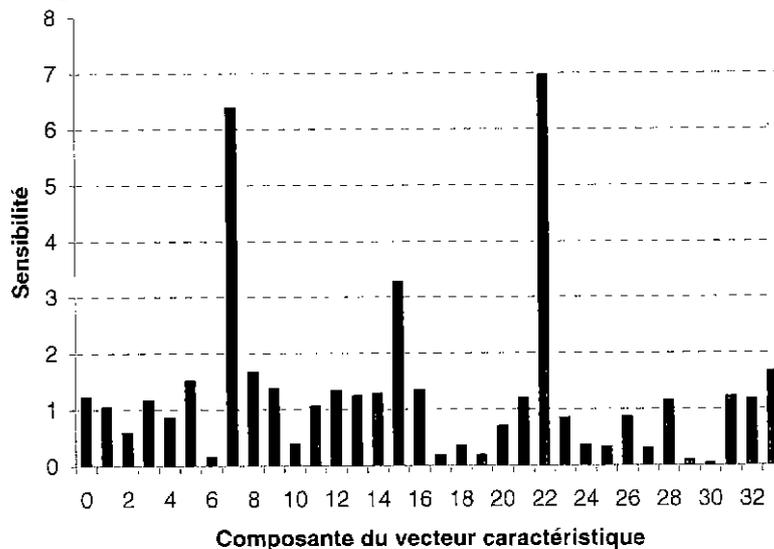


Figure 84 : Sensibilité vis-à-vis des composantes du vecteur caractéristique dans le cas où le la combinaison de dégradations est considérée comme un défaut unique. Notation par calcul de distance à l'image parfaite.

IV - Conclusion

Nous avons montré dans ce chapitre que la méthode univariante pouvait être appliquée à l'étude d'un mélange de dégradations.

Si le système dégradant, c'est à dire la cause des dégradations et l'ordre dans lequel elles apparaissent sur l'image, est connu, il est possible de les caractériser individuellement moyennant un apprentissage adéquat. Le cas de figure que nous avons étudié, une défocalisation suivie d'une compression JPEG, a permis de décorréler les deux dégradations. Ce résultat n'était pas forcément prévisible étant donné que la phase de compression inclut aussi un filtrage passe-bas.

Dans le cas où le système dégradant n'est pas connu, il est toujours possible de réaliser un apprentissage sur la qualité globale de l'image, sans pouvoir toutefois accéder à chaque dégradation individuellement. Nous avons montré, par exemple, que l'on pouvait estimer la qualité visuelle de l'image sans connaissance, a priori, de la nature des dégradations. Dans ce cas, l'apprentissage donne des résultats aussi satisfaisant que pour un défaut unique.

Ce chapitre ne constitue pas vraiment une généralisation de la méthode. Notre objectif est de montrer que la méthode univariante peut être appliquée même dans le cas, plus complexe, d'un mélange de défauts. Néanmoins, la limitation principale reste le nombre d'exemples requis pour réaliser l'apprentissage. Dans le cas d'un mélange encore plus complexe, il faudrait sélectionner ces exemples méticuleusement de manière à éviter les redondances pour en réduire au maximum le nombre.

Conclusion - Perspectives

Ce travail a permis de développer et valider une nouvelle méthode pour évaluer la qualité d'une image lorsqu'on ne dispose pas d'image de référence.

La manière de noter la qualité de l'image, qui dépend de l'application à laquelle elle est destinée, est préalablement apprise sur des exemples connus étalonnés par les personnes compétentes. Le modèle utilisé, un réseau de neurones, permet l'apprentissage de situations très complexes, moyennant un prétraitement de l'image adapté à la dégradation considérée. En particulier, si les exemples sont étalonnés sur une appréciation visuelle de la qualité de l'image, le modèle reproduit ensuite un comportement visuel sur un exemple inconnu et ce, sans modèle explicite de vision. De la même manière, le modèle peut être étalonné pour estimer, à partir de l'image dégradée, certaines caractéristiques du système dégradant (taux de compression, défaut de mise au point...).

La méthode a été dans un premier temps validée pour des défauts « simples », c'est à dire non mélangés les uns avec les autres, uniformes sur l'image ou bien ponctuels. Dans le cas de défauts ponctuels, l'apprentissage a été réalisé de manière à ne détecter que les défauts visuellement gênants. La détection obtenue, qui prend donc en compte les caractéristiques de masquage du système visuel, est très fiable. Les résultats sont même meilleurs que ceux obtenus avec le VDP, modèle bivariant que nous avons utilisé, il est vrai d'une façon un peu particulière, pour comparaison. En ce qui concerne les dégradations uniformément réparties sur l'image, nous avons montré qu'il était possible non seulement de donner une appréciation visuelle de la qualité, mais aussi de retrouver certaines caractéristiques du système dégradant. Une application concrète de la méthode a été réalisée sur un jeu d'images provenant du satellite SPOT1, qui avait un très léger défaut de mise au point. Elle a permis d'estimer ce défaut avec une précision inférieure au demi pas de focalisation.

Nous avons finalement montré comment on pouvait généraliser cette approche à un mélange de dégradations. Dans ce cas, la connaissance du système dégradant, c'est à dire des dégradations et de l'ordre dans lequel elles sont appliquées sur l'image, peut permettre de retrouver leurs caractéristiques propres. Si l'on ne connaît pas le système dégradant, on peut simplement, grâce à cette méthode, noter la qualité de l'image sur une échelle choisie en fonction de l'application visée.

Le principal avantage d'une telle approche est son adaptabilité d'un contexte à un autre. Il faut cependant réaliser un apprentissage pour chaque contexte. Un autre avantage est le temps de calcul nécessaire pour traiter une image. Le seul coût est celui dû au prétraitement de l'image, le réseau de neurones n'effectuant que des calculs élémentaires en nombre réduit. Un modèle bivariant complexe, comme le VDP par exemple, a besoin de calculer 33 FFT directes et 33 FFT inverses pour chacune des images soit 132 FFT, très coûteuses en temps de calcul.

En revanche, les limitations de cette méthode sont imposées par le réseau de neurones. Le nombre d'exemples nécessaires pour réaliser l'apprentissage doit approximativement être le carré du nombre d'entrées du modèle. Le prétraitement de l'image sert à réduire leur nombre, mais dans la plupart des cas, mille exemples au moins sont nécessaires. Lorsque l'on désire une notation représentative d'une perception visuelle, il n'est pas toujours facile de faire noter mille images par des observateurs...

* *
*

Plusieurs perspectives intéressantes se dégagent de ce travail, tant au niveau de la méthode, qu'au niveau des applications.

L'étude a été menée sans chercher à optimiser la taille de la base d'apprentissage du RN. Nous avons toujours utilisé un nombre d'exemples multiple du carré du nombre de neurones. Un apprentissage beaucoup plus supervisé, où les exemples seraient choisis méticuleusement de manière à apporter au RN des informations complémentaires mais disjointes, permettrait d'en réduire le nombre nécessaire de manière très importante.

D'autre part, nous n'avons utilisé qu'une architecture de RN (multicouche) et un seul algorithme d'apprentissage (rétropropagation du gradient). Ce type d'architecture permet de modéliser un phénomène à partir de caractères physiques le décrivant. En revanche il est difficile d'inverser le problème pour retrouver, après coup, quelles relations sur les caractères physiques sont réellement importantes dans la description quantitative du phénomène. Il existe des réseaux quelque peu différents, appelés réseaux de fonctions à base radiale, qui permettraient non seulement de modéliser l'évaluation de la qualité d'une image, mais aussi d'expliquer beaucoup plus précisément que l'étude de sensibilité que nous avons faite quels sont les caractères physiques mis en jeu.

Le domaine d'application de cette méthode est très vaste. Nous avons étudié le cas de quelques défauts ou mélange de défauts sur des images monochromes.

Il paraît difficile de retrouver les caractéristiques du système dégradant pour un mélange complexe, en revanche on peut envisager une étude de qualité visuelle globale. Un vecteur caractéristique permettant de décrire les principaux aspects de l'image, dans un espace perceptuel par exemple, permettrait peut-être d'établir de façon univariante une carte des dégradations gênantes sur l'image.

Bibliographie

- [AAE97] V.R. Algazi, N. Avadhanam and R.R. Estes, Quantifying Coding Performance for Pre-processed Images, *Proceedings of the SPIE, Very High Resolution and Quality Imaging II*, Vol. 3025, pp. 123-133, 1997.
- [ABP96] Anguita, I. Barbieri, F. Passagio, S. Ridella, Block distortion assessment for image compression through ANNS, *Proceedings of the SPIE – The international society for optical engineering*, vol. 2664, pp. 25-36, 1996.
- [Ama95] S. Mari, Learning and statistical inference, in M. A. Arbib, ed., *The Handbook of Brain Theory and Neural Networks*, MIT Press, Cambridge, Massachusetts, pp. 522-526, 1995.
- [AvA96] N. Avadhanam and V.R. Algazi, Prediction and Measurement of High Quality in Still Image Coding, *Proceedings of the SPIE, Very High Resolution and Quality Imaging*, Vol. 2663, pp. 100-109, 1996.
- [Bar90] P.G.J. Barten, Evaluation of Subjective Image Quality with the Square-Root integral Method, *Journal of the Optical Society of America*, Vol. 7, No. 10, pp. 2024-2031, 1990.
- [Bie94] J. Biedermann, Anwendungen Neuronaler Netze beim VLSI-CAD, Diplomarbeit, Institut für Numerische und Angewandte Mathematik, Geor-August Universität Göttingen, 1994.
- [Bes93] N. Beser, Image Data Compression Metrics, *Proceedings AIAA, Computing in Aerospace 9*, pp. 292-303, 1993.
- [BoC92] Bo Cai Gao, An Operational Method for Estimating Signal to Noise Ratios from Data Acquired with Imaging Spectrometers, *Remote Sensing of environment*, No. 43, pp. 23-33, 1993
- [Bra99] A. P. Bradley, A Wavelet Visible Difference Predictor, *IEEE Transactions on Image Processing*, Vol. 8, No.5 , pp717-730, 1999.
- [CaR68] F.W. Cambell and J.G. Robson, Application of Fourier Analysis to the Visibility of Gratings, *Journal of Physiology*, Vol. 197, pp.. 551-566, 1968.
- [Car93] Caruana, R, Generalization vs. Net Size, *Neural Information Processing Systems*, Tutorial, Denver, CO, 1993.
- [CCI90] Méthode d'évaluation subjective de la qualité des images de télévision, *Recommandation 500-4 du CCIR*, 1990.
- [CGC97] F.X.Coudoux, M.G.Gazalet, P.Corlay, J.M.Rouvaen, A perceptual approach to the reduction of blocking effect in DCT-coded images, *J. Visual Communication and Image Representation*, Vol.8, N°4, pp. 327-337, Dec. 1997.

- [ChL95] Chou et Y. C Li, A perceptually tuned subband image coder based on the measure of just-noticeable distortion profile, *IEEE transactions on circuits and systems for video technology*, vol. 5, n°6, pp. 467-476, 1993.
- [ChR90] B. Chitprasert and K.R. Rao; "Human visual weighted progressive image transmission" *IEEE transactions on communication*", Vol 38, n°7, pp 1040-1044, juillet 1990.
- [CKL66] F.W. Cambell, J.J. Kulikowski and J. Levinson, The Effect of Orientation on the Visual Resolution of Gratings, *Journal of Physiology*, Vol. 187, pp.. 427-436, 1966.
- [CNES1] CNES, Guide des utilisateurs de données SPOT : manuel de référence vol.1
- [Cor71] T.N. Cornsweet, *Visual Perception*. London. Academic Press, 1971.
- [CSB00] P. Le Callet, A. Saadane, D. Barba, Frequency and Spatial Pooling of Visual Differences for Still Image Quality Assessment, *Proceedings of the SPIE*, conference on Electronic Imaging, january 2000, to be published.
- [Dal92] S. Daly, The Visible Differences Predictor: an Algorithm for the Assessment of Image Fidelity, *Proceedings of the SPIE*, Human Vision, Visual Processing and Digital Display III, Vol. 1666, p 2-15, 1992.
- [Dal93] S. Daly, The Visible Differences Predictor: an Algorithm for the Assessment of Image Fidelity, in *Digital Images and Human Vision*, edited by A. B. Watson, MIT press; pp. 179-206, 1993.
- [Dal94] S Daly, A Visual Model for Optimizing the Design of Image Processing Algorithms, *International Conference on Image Processing*, ICIP Part 3, pp. 16-20, 1994.
- [DCK97] R.G. Driggers, P. Cox and M. Kelley, National Imagery Interpretation Rating System and the Probabilities of Detection, Recognition, and Identification, *Optical Engineering*, Vol. 36, No. 7, 1952-1959, 1997.
- [DeL62] J.J. DePalma et E.M. Lowry, Sine Wave Response of the Visual System, II, Sine Wave and Square Wave Contrast Sensitivity, *Journal of the Optical Society of America*, Vol. 52, pp. 328-335, 1962.
- [DiH90] M. Dinguirard and P. Henry, Expertise d'images SPOT acquises sur l'Amérique du sud : une nouvelle énigme, *rapport CNES S-NT-OB1-1362-CN*. du 02/08/90
- [EsF93] A.M. Eskicioglu, P.S. Fisher, A Survey of Quality Measures for Gray Scale Image Compression, *Proceedings AIAA*, Computing in Aerospace 9, Vol. 1/2, No. 939, p 304-313, 1993.
- [EuC96] T. Eude, H. Cherifi, Optimisation perceptuelle d'un codeur Jpeg, *Proc. of CORESA 96*, pp 220-227, 1996.

- [GoW77] R.C. Gonzalez and P. Wintz, Digital Image Processing, *Addison-Wesley Publishing Company*, 1977.
- [HaH77] C.F. Hall and F. Hall, A Nonlinear Model for the Spatial Characteristics of the Human Visual System, *IEEE Transaction on Systems, Man and Cybernetics*, Vol. SMC-7, No. 3, pp. 161-170, 1977.
- [Hal81] C.H. Hall, Subjective Evaluation of a Perceptual Quality Metric, *Proceedings of the SPIE, Image Quality*, Vol. 310, pp. 200-2004, 1981.
- [Hui75] C. Huijbregts, Regionalized variables and quantitative analysis of spatial data, *Display and analysis of spatial data*, Eds J. Davis and M. Mac Cullagh, wiley, New-York, 1975.
- [Jai89] A.K. Jain, Fundamentals of Digital Image Processing, *Prentice-Hall Information and System Sciences Series*, 1989.
- [JoB98] J.M Jolion, S. Bres, Influence du codage JPEG sur les descripteurs d'images, *Traitement du signal*, vol. 15 n°4, pp309-320, 1998.
- [KaK93] S. A. Karunasekera and N. G. Kingsbury, A Distortion Measure for Blocking Artifacts in Images based on Human Visual Sensitivity, *Proceedings of the SPIE, Visual Communications and Image Processing*, Vol. 2094, pp. 474-486, 1993.
- [Kak94] S. A. Karunasekera and N. G. Kingsbury, A Distortion Measure for Image Artifacts based on Human Visual Sensitivity, *Proceedings of ICASSP - IEEE*, Vol. 5, pp. 117-120, 1994.
- [KaK95] S. A. Karunasekera and N. G. Kingsbury, A Distortion Measure for Coding Artifacts in Images : Implementation Aspects, *Internal Report*, Signal Processing Group, Department of Engineering, University of Cambridge, 1995.
- [KBM98] : P. Kubik, E. Breton, A. Meygret, B. Cabrières, P. Hazane, D. Léger, SPOT4 HRVIR first in-flight Image Quality results, *Conférence Europto 98, Barcelone 21-24 sept. 98*.
- [KeH91] Keh-Shih Chuang, HK Huang, Assessment of Noise in a Digital Image Using the Join-Count Statistic and the Moran Test, *Phys. Med. Biol.*, Vol. 37, No 2, pp. 357-369, 1991
- [Kun93] M. Kunt, Traitement numérique des images, *Presses Polytechniques et Universitaires Romandes*, Traitement de l'information : volume 2, 1993.
- [LeF80] G.E. Legge and J.M. Foley, Contrast Masking in Human Vision, *Journal of the Optical Society of America*, Vol. 70, No. 12, pp. 1458-1470, 1980.
- [LeH89] J.S. Lee, K.Hoppel, Noise Modeling and Estimation of Remotely-Sensed Images, *Proceedings of IGARSS 89*, pp. 1005-1008, 1989

- [LeM95] : D. Léger, A. Meygret, Expérimentation technologique, bilan de refocalisation de SPOT1, *rapport CNES réf. S1-NT-OB1-2097-CN Ed.0 Rév.0*, 17/10/95.
- [Lim79] J.O. Limb, Distorsion Criteria of the Human Viewer, *IEEE Transaction on Systems, Man and Cybernetics*, Vol. SMC-9, No. 12, pp. 778-793, 1979.
- [Lin56] E.H. Linfoot, Transmission Factors and Optical Design, *Journal of the Optical Society of America*, Vol. 46, No. 9, pp. 740-752, 1956.
- [Lin58] E.H. Linfoot, Quality Evaluation of Optical Systems, *Optica Acta*, Vol. 5, No 1-2, pp. 1-13, 1958.
- [LoC86] I. D. Longstaff, J. F. Cross, A Pattern recognition Approach to Understanding the Multi-layer Perceptron, *RSRE Memorandum No. 3936*, 1986.
- [Maa95] Maass, W., Vapnik-Chervonenkis dimension of neural networks, in M. A. Arbib, ed., *The Handbook of Brain Theory and Neural Networks*, MIT Press, Cambridge, Massachusetts, pp. 522-526, 1995.
- [Mar86] H. Marmolin, Subjective MSE Measure, *IEEE Transaction on Systems, Man and Cybernetics*, Vol. SMC-16, No. 3, pp. 486-489, 1986.
- [MaS74] J.L. Mannos and D.J. Sakrison, The Effects of Visual Fidelity Criterion on the Encoding of Images, *IEEE Transactions on Information theory*, Vol. IT-20, No 4, pp. 525-536, 1974.
- [Mat71] G. Matheron, The Theory of regionalized variables and its application, *Les cahiers du centre de morphologie mathématique de fontainebleau*, 1971.
- [Mat96] T. Matews, Objective Image Quality Measures for Magnetic Resonance Images, Master's thesis, University of Calgary, 1996.
- [MEC98] A. Mayache, T. Eude, H. Cherifi, A comparison of image quality models and metrics based on human visual sensitivity, *Proceedings of the International conference on image processing*, vol.3, pp. 409-413, 1998.
- [MJR90] Peter MEER, Jean-Michel JOLION, Azriel ROSENFELD, A Fast Parallel Algorithm for Blind Estimation of Noise Variance, *IEEE transactions on pattern analysis and machine intelligence*, Vol.12, No.2, 1990
- [NiB92] N.B. Bouzas and B.H. Bouzas, Objective Image Quality Measure Derived from Digital Image Power Spectra, *Optical Engineering*, Vol. 31, No. 4, pp. 813-825, 1992.
- [Nil85] N.B. Nill, A Visual Model Weighted Cosine Transform for Image Compression and Quality Assessment, *IEEE Transactions on communications*, Vol. COM-33, No. 6, pp. 551-556, 1985.

- [NLS89] K.N. Ngan, K.S. Leong and H. Singh, Adaptive cosine transform coding of images in perceptual domain, *IEEE Transactions on acoustics, speech and signal processing*, Vol 37, n°11, pp 1743-1750, Novembre 1989.
- [NSR71] J. Nachmias, M. Sachs and J. Robson, Spatial Frequency Channels in Human Vision, *Journal of the Optical Society of America*, Vol. 61, pp. 1176-1186, 1971.
- [PhW84] G.C. Phillips and H.R. Wilson, Orientation Bandwidths of Spatial Mechanisms Measured by Masking, *Journal of the Optical Society of America*, Vol. 1, No 2, pp. 226-232, 1984.
- [Poy93] C.A. Poynton, "Gamma" and its Disguises: The Nonlinear Mappings of Intensity in Perception, CRTs, Film and Video, *Journal of the Society of Motion Picture and Television Engineers*, Vol. 102, No 12, 1993, available on <http://Home.InfoRamp.Net/~poynton/Poynton-articles.html>.
- [Oli98] J.L. Olives, Optimisation globale d'un système imageur à l'aide de critères de qualité visuelle, *thèse ENSAE*, 1998.
- [Pra78] W.K. Pratt, Digital Image Processing, *A Wiley-Interscience Publication*, J. Wiley & Sons, 1978.
- [RhW85] D. E. Rumelhart, G. E. Hinton, R. J. Williams, Learning Internal Representations by Error Propagation », *Parallel Distributed Processing : Explorations in the microstructure of cognition*, Vol.1 : Foundations, Eds. Rumelhart & McClelland, Bradford Books/MIT Press, 1985.
- [RoK76] A. Rosenfeld and A.C. Kak, Digital Picture Processing, *Academic Press*, 1976.
- [SaC89] J.A. Saghri, P.S. Cheatham and A. Habibi, Image Quality Measure Based on a Visual System Model, *Optical Engineering*, Vol. 28, No. 7, pp. 813-818, 1989.
- [Sch94] T. Schreiner, Ausdünnungsverfahren für Neuronale Netze, Diplomarbeit 1140, IPVR, Universität Stuttgart, 1994.
- [Sha56] O.H. Shade, Optical and Photoelectric Analog of the Eye, *Journal of the Optical Society of America*, Vol. 46, No 9, pp. 721-739, 1956.
- [SiD88] J. Sietsma, R. J. F. Dow, Neural Net Pruning – Why and How, *IEEE International Conference on Neural Networks*, Vol. 1, pp 325-333, 1988.
- [SNNS] A. Zell, G. Mamier, M. Vogt, N. Mache, R. Hübner, S. Döring, K. U. Hermann, T. Soye, M. Schmalzl, T. Sommer, A. Hatzigeorgiou, D. Posselt, T. Schreiner, B. Kett, G. Clemente, J. Wieland, SnnS (Stuttgart Neural Network Simulator) : User Manual, Version 4.1.
- [SNR71] M. B. Sachs, J. Nachmias and J.G. Robson, Spatial-Frequency Channels in Human Vision, *Journal of the Optical Society of America*, Vol. 61, No 9, pp. 1176-1186, 1971.

- [Sto72] T.G. Stockman, Image Processing in the Context of a Visual Model, *Proceedings of the IEEE*, Vol. 60, pp. 828-842, 1972.
- [Vas73] A. Vassilev, Contrast Sensitivity Near Borders: Significance of test Stimulus Form, size and duration, *Vision Research*, Vol. 13, April 1973, pp. 719-730, 1973.
- [Wal91] K Wallace, The JPEG still image compression standard, *Communication of the ACM*, 34, pp 30-44, 1991.
- [Wan95] B.A. Wandell, Foundations of Vision, *Sinauer Associates, Inc.*, 1995
- [Wat93] A.B. Watson, Visual Optimization of DCT Quantification Matrices for Individual Images, *Proceedings of AIAA Computing in Aerospace 9*, pp. 286-291, 1993.
- [WFP83] H.R. Wilson, D.K. McFarlane and G.C. Phillips, Spatial Frequency Tuning of Orientation Selective Units Estimated by Oblique Masking, *Vision Research*, Vol. 23, No. 9, pp. 873-882, 1983.

ANNEXE A - Base d'apprentissage générale



Image001.tif

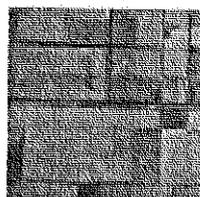


Image002.tif

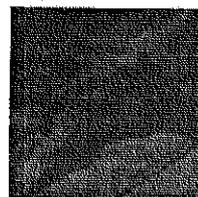


Image003.tif

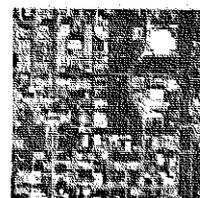


Image004.tif

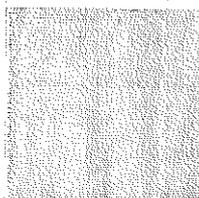


Image005.tif



Image006.tif



Image007.tif



Image008.tif



Image009.tif



Image010.tif

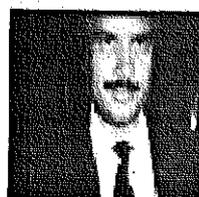


Image011.tif



Image012.tif

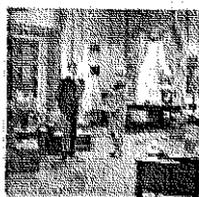


Image013.tif



Image014.tif

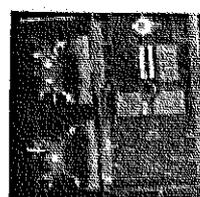


Image015.tif



Image016.tif



Image017.tif



Image018.tif



Image019.tif

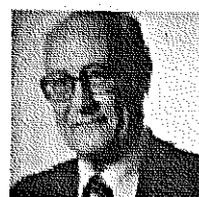


Image020.tif



Image021.tif

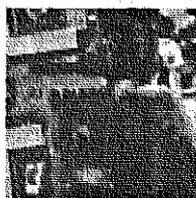


Image022.tif



Image023.tif

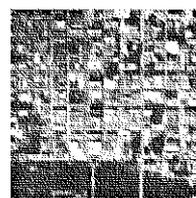


Image024.tif



Image025.tif



Image026.tif

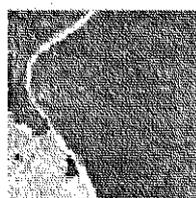


Image027.tif

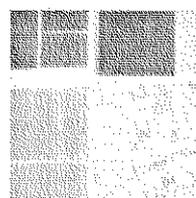


Image028.tif



Image029.tif

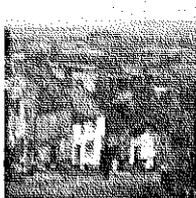


Image030.tif



Image031.tif



Image032.tif

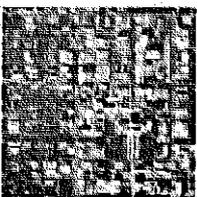


Image033.tif



Image034.tif

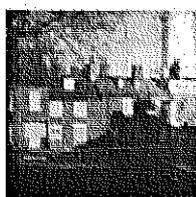


Image035.tif

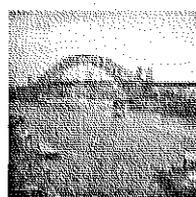


Image036.tif

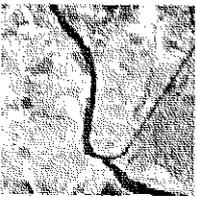


Image037.tif



Image038.tif

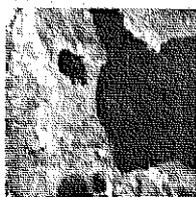


Image039.tif

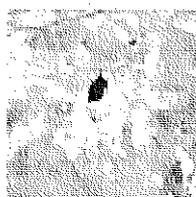


Image040.tif

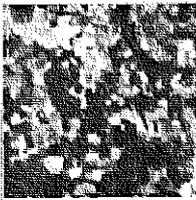


Image041.tif

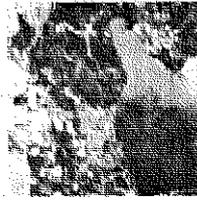


Image042.tif



Image043.tif

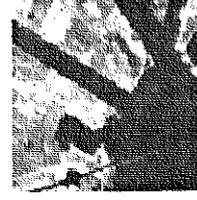


Image044.tif



Image045.tif



Image046.tif



Image047.tif



Image048.tif

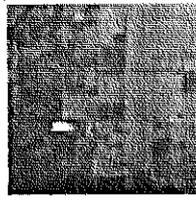


Image049.tif

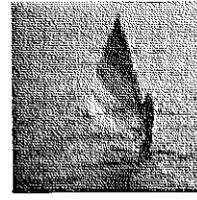


Image050.tif

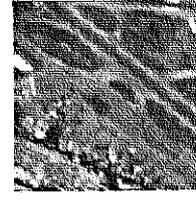


Image051.tif

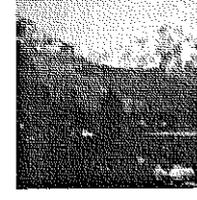


Image052.tif



Image053.tif

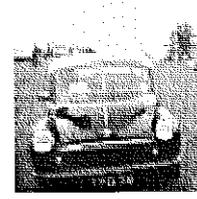


Image054.tif

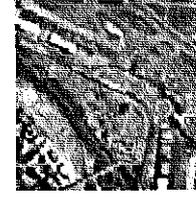


Image055.tif

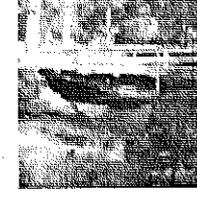


Image056.tif

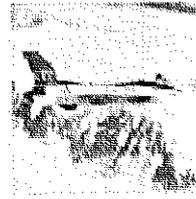


Image057.tif

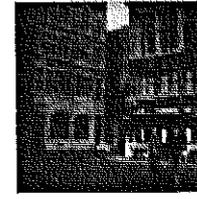


Image058.tif

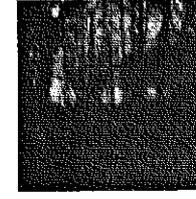


Image059.tif



Image060.tif

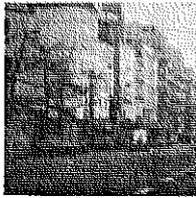


Image061.tif

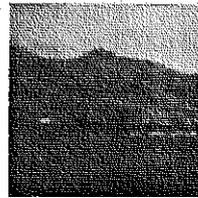


Image062.tif

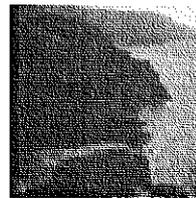


Image063.tif

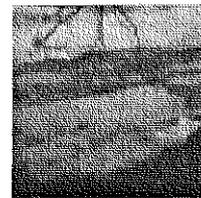


Image064.tif

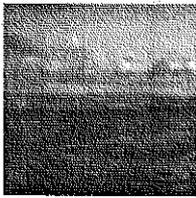


Image065.tif

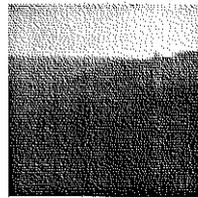


Image066.tif

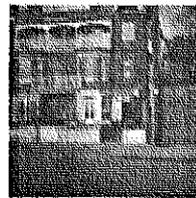


Image067.tif

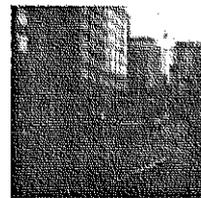


Image068.tif

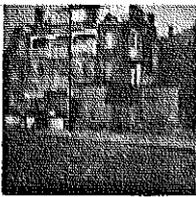


Image069.tif

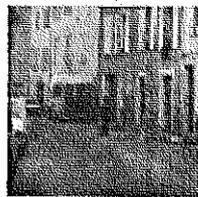


Image070.tif

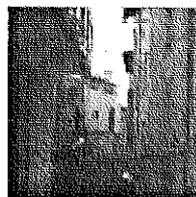


Image071.tif

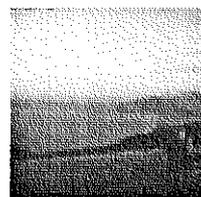


Image072.tif

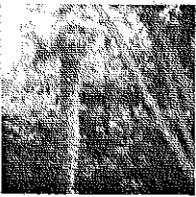


Image073.tif

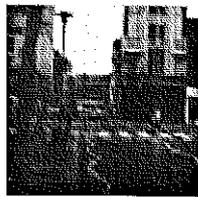


Image074.tif

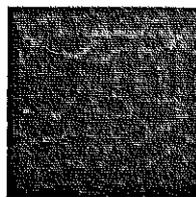


Image075.tif

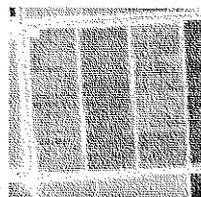


Image076.tif



Image077.tif

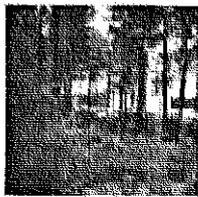


Image078.tif

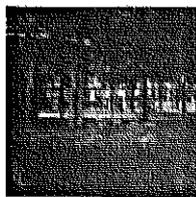


Image079.tif

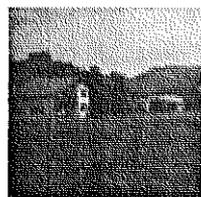


Image080.tif



Image081.tif



Image082.tif

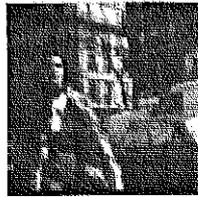


Image083.tif



Image084.tif

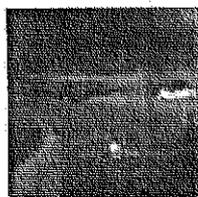


Image085.tif



Image086.tif

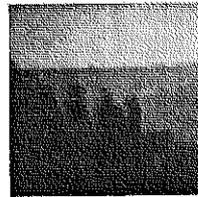


Image087.tif



Image088.tif



Image089.tif

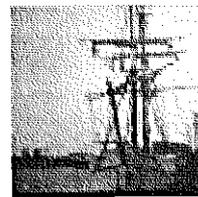


Image090.tif

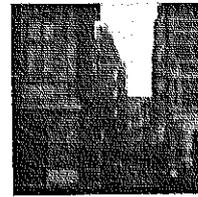


Image091.tif

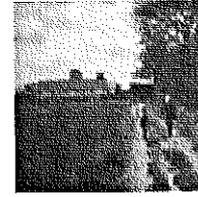


Image092.tif

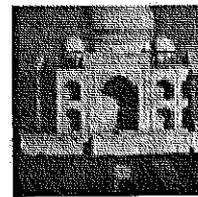


Image093.tif

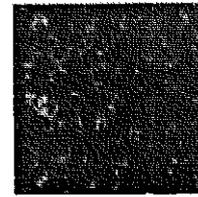


Image094.tif



Image095.tif

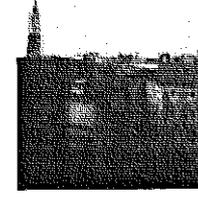


Image096.tif

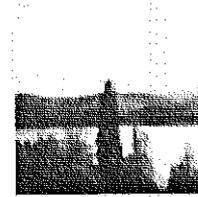


Image097.tif

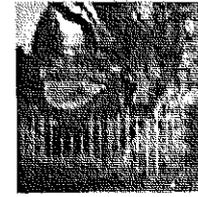


Image098.tif

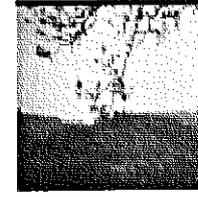


Image099.tif

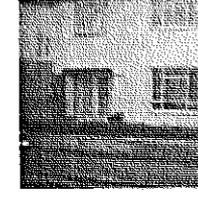


Image100.tif

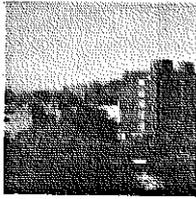


Image101.tif

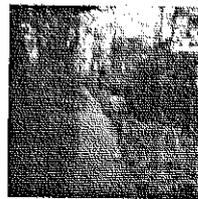


Image102.tif

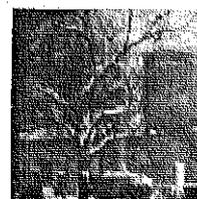


Image 103.tif

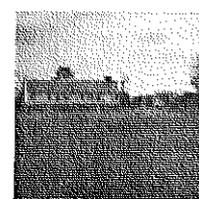


Image 104.tif

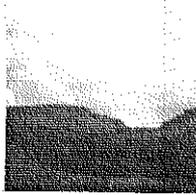


Image 105.tif

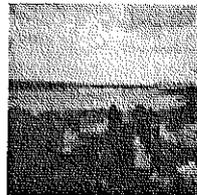


Image106.tif



Image107.tif

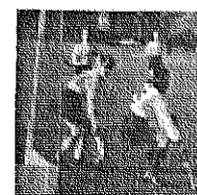


Image108.tif

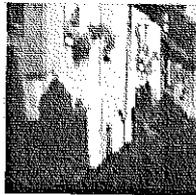


Image 109.tif



Image 110.tif



Image111.tif

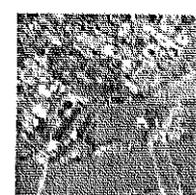


Image112.tif

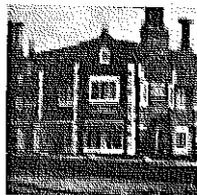


Image113.tif

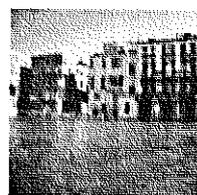


Image 114.tif

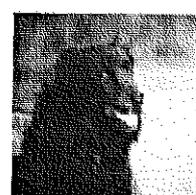


Image 115.tif



Image116.tif

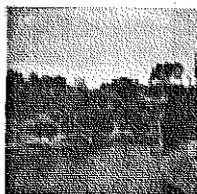


Image117.tif

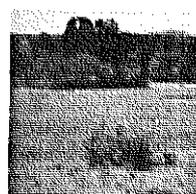


Image118.tif



Image119.tif



Image120.tif

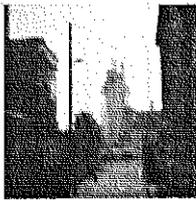


Image 121.tif



Image 122.tif

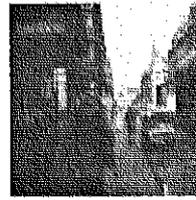


Image 123.tif



Image 124.tif

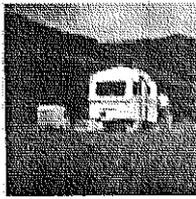


Image 125.tif

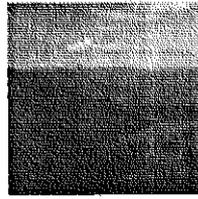


Image 126.tif

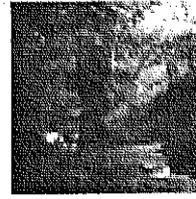


Image 127.tif



Image 128.tif

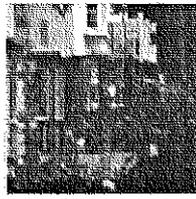


Image 129.tif

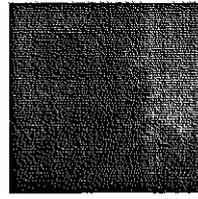


Image 130.tif

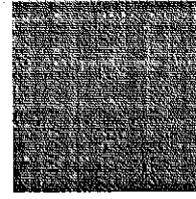


Image 131.tif

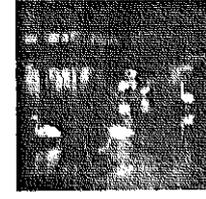


Image 132.tif



Image 133.tif



Image 134.tif



Image 135.tif

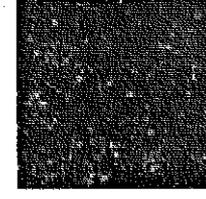


Image 136.tif

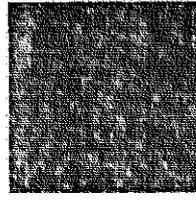


Image 137.tif

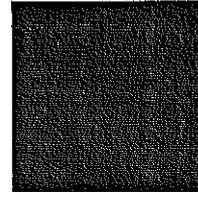


Image 138.tif

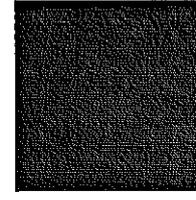


Image 139.tif

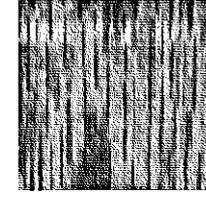


Image 140.tif

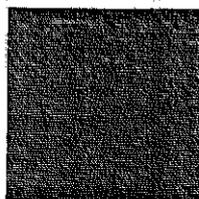


Image 141.tif

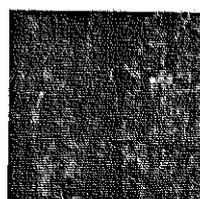


Image 142.tif



Image 143.tif

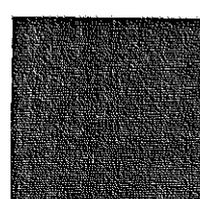


Image 144.tif

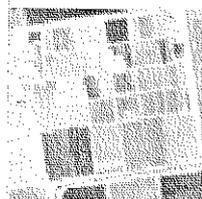


Image 145.tif

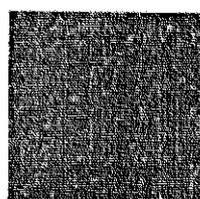


Image 146.tif

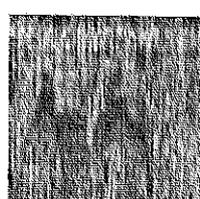


Image 147.tif

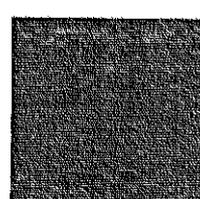


Image 148.tif



Image 149.tif

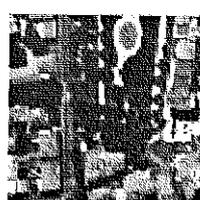


Image 150.tif



Image 151.tif

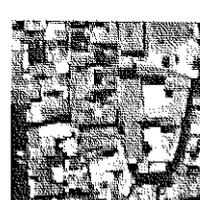


Image 152.tif

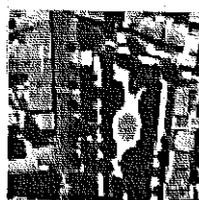


Image 153.tif

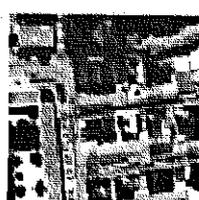


Image 154.tif

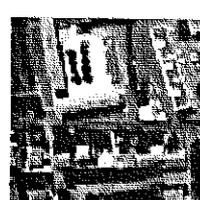


Image 155.tif



Image 156.tif

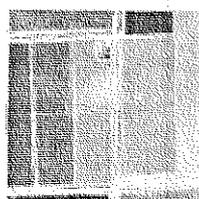


Image 157.tif

ANNEXE B - Le système visuel humain (SVH)

A1 - La perception visuelle

La perception visuelle est la faculté que l'homme a de détecter des structures ou des événements dans le milieu qui l'entoure. Elle nécessite une certaine sensibilité à l'énergie lumineuse.

La perception visuelle peut être étudiée de deux manières différentes :

- plan neurophysiologique : concerne l'étude du fonctionnement des éléments de base, par exemple les neurones ;
- plan psychophysique : on étudie la réponse globale du SVH à des stimulus (images) tests choisis de manière à mettre en évidence l'une ou l'autre de ses propriétés.

Le système visuel, très complexe, permet la perception des couleurs, du mouvement. Il s'adapte aux conditions de luminosité, d'éloignement... Nous nous limiterons volontairement dans cette annexe à l'étude de scènes ou d'images statiques, en niveaux de gris (c'est-à-dire monochromes).

L'anatomie du SVH sera décrite dans un premier temps avant d'aborder les aspects psychophysiques de la perception visuelle.

A2 - Anatomie du système visuel humain

1) Présentation générale

La Figure 85 montre l'organisation générale du SVH. La lumière incidente est captée par des photorécepteurs (cônes et bâtonnets), l'information est traitée au niveau de la rétine puis transmise le long du nerf optique jusqu'aux corps genouillés latéraux dont les axones servent de relais jusqu'au cortex strié. Au niveau du chiasma optique, la sortie de chaque œil est divisée en deux, l'information contenue dans la moitié gauche du champ visuel est traitée par la partie droite du cerveau et inversement.

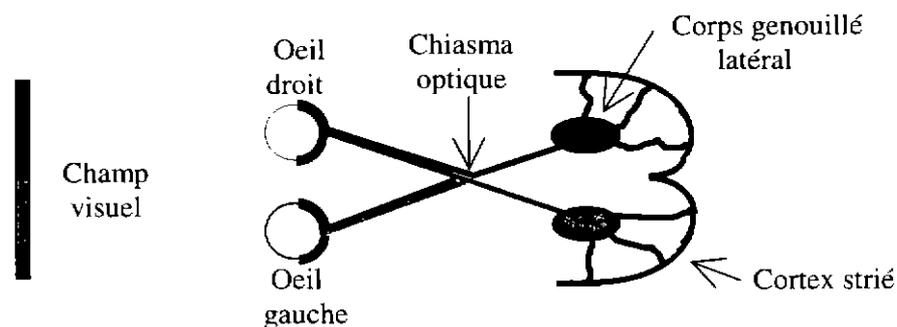


Figure 85 : Organisation schématique du système visuel humain.

Le système visuel humain utilise une partie du système nerveux. Celui-ci est composé de cellules nerveuses appelées neurones. Les transferts d'information entre les neurones se font de manière électrochimique au niveau de leurs jonctions, les synapses. Le neurone qui émet l'information est le neurone présynaptique, celui qui la reçoit est le neurone postsynaptique. L'action d'un neurone présynaptique sur un neurone postsynaptique peut être de deux types :

- Excitatrice : elle permet la création d'une impulsion dans le neurone postsynaptique ;
- Inhibitrice : elle neutralise cette impulsion.

Le système visuel représente environ 15 cm² de surface totale du cortex. Cette région est appelée « cortex visuel ».

2) L'œil

C'est l'organe du système visuel qui va permettre la transformation de l'information énergétique lumineuse en une information électrique interprétable par le cerveau (Figure 86).

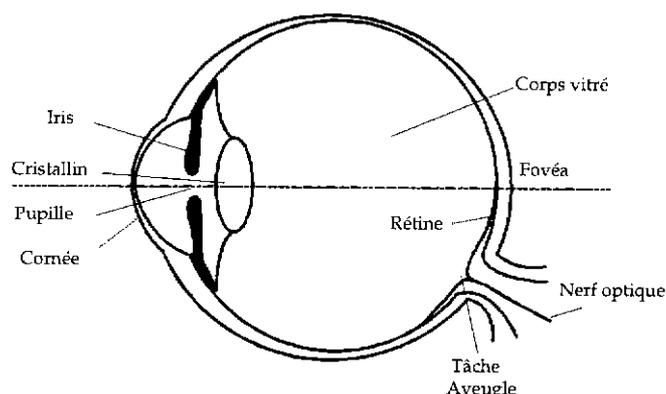


Figure 86 : Coupe schématique de l'œil.

La lumière traverse la cornée, la pupille, le cristallin et rencontre la rétine. La pupille agit comme un diaphragme et contrôle la quantité de lumière qui entre dans l'œil. Son diamètre peut varier de 2 mm à 9 mm : c'est un filtre passe-bas. Elle est commandée par l'iris. Le cristallin joue le rôle de lentille convergente pour former l'image sur la rétine. Ce n'est pas une lentille parfaite et cette imperfection est à l'origine d'une aberration sphérique qui se manifeste par un flou sur la rétine. Le cristallin joue donc aussi le rôle d'un filtre passe-bas.

3) La rétine

La rétine est la couche neurosensorielle. Elle transforme la lumière qu'elle reçoit en signaux électriques qui sont transmis vers le cortex visuel par le nerf optique. Elle est formée de cinq types de cellules, ordonnées régulièrement en trois couches comme le montre la Figure 87 [Cor71].

a - La couche de cellules photoréceptrices

La première couche contient les cellules photoréceptrices. Celles-ci sont de deux types :

- les cônes : au nombre de 6,5 millions, ils fonctionnent dans le domaine photopique, c'est-à-dire des moyennes et hautes luminances. Ils détectent la couleur (trois types de cônes) et permettent la distinction des détails ;
- les bâtonnets : au nombre de 130 millions, ils fonctionnent dans le domaine scotopique, c'est-à-dire pour les basses luminances. Ils permettent la distinction des formes.

Au voisinage de la fovéa, la densité de cônes est très importante et décroît rapidement lorsqu'on s'en éloigne. La densité maximale des cônes dans la fovéa est de l'ordre de 120 unités par degré, ce qui fixe la résolution visuelle à une minute d'arc environ. L'acuité visuelle est donc maximale quand l'œil est immobile et que l'image se forme au centre de la fovéa. Les bâtonnets sont plus nombreux dans la périphérie.

De plus, au niveau de cette couche, la lumière incidente subit une transformation non linéaire ($y = x^n$).

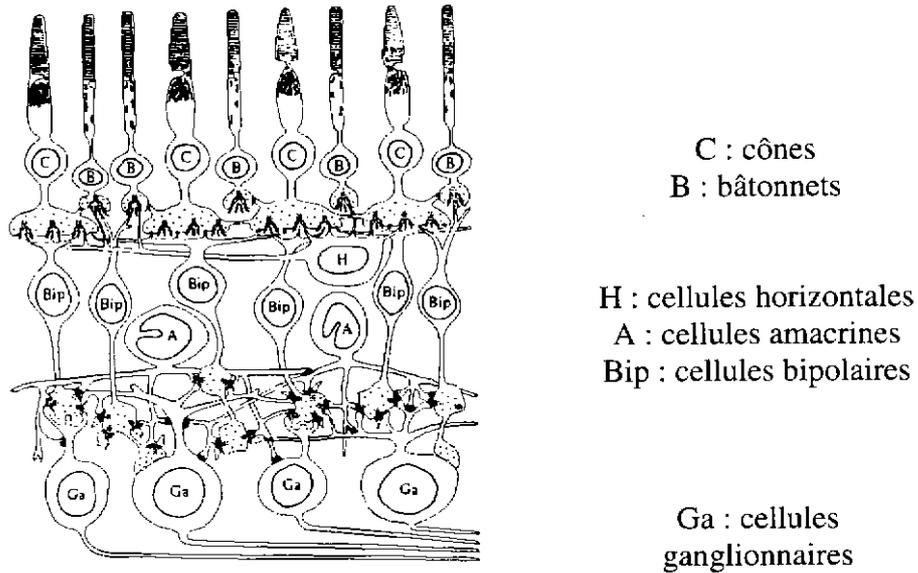


Figure 87 : Organisation en couches de la rétine.

b - La couche des cellules horizontales, amacrines et bipolaires

C'est la couche intermédiaire. Les cellules bipolaires sont connectées aux cellules photoréceptrices. Les cellules horizontales reçoivent les synapses des cellules photoréceptrices et bipolaires. Les cellules amacrines reçoivent les synapses des cellules bipolaires et peuvent agir sur les cellules bipolaires et ganglionnaires. Il existe donc à ce niveau des boucles de contre-réaction. Ce mécanisme est appelé « mécanisme d'inhibition latérale » : chaque cellule peut inhiber l'effet de celles qui l'entourent. L'intensité de cette inhibition est fonction de l'augmentation d'intensité de la lumière reçue par les photorécepteurs et elle diminue avec la distance entre photorécepteurs. Ce processus est en partie responsable de l'adaptation à la lumière.

c - La couche de cellules ganglionnaires

La dernière couche est composée des cellules ganglionnaires, connectées aux cellules bipolaires. Les axones des cellules ganglionnaires se prolongent dans le nerf optique.

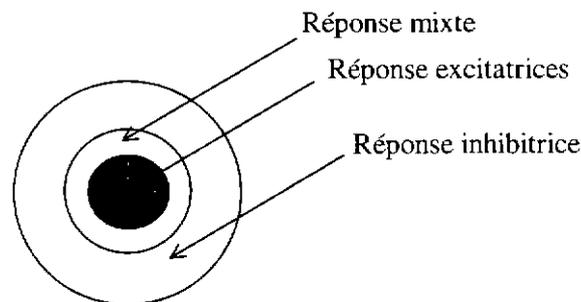


Figure 88 : Organisation de la réponse d'une cellule ganglionnaire.

A ce niveau, la luminance incidente n'a pas d'effet. Ces cellules mesurent une différence, dans leur champ réceptif, entre la lumière qui tombe au centre et celle qui tombe sur les bords. Le champ réceptif est divisé en plusieurs régions : excitatrices (ON), inhibitrices (OFF) et le mélange des deux (Figure 88).

L'information est analysée à travers les contrastes entre les réponses ON et OFF. Ces cellules permettent la perception des contrastes et la détection des contours.

Les cellules ganglionnaires ont des formes et des tailles différentes. Ces cellules ont été regroupées en deux classes : les cellules ganglionnaires « parasol », car elles possèdent de nombreuses terminaisons et les cellules « midget », qui en ont peu.

Après avoir été divisée au niveau du chiasma optique, l'information traitée par la rétine arrive au corps genouillé latéral.

4) Le corps genouillé latéral (CGL)

Après séparation par le chiasma optique, le CGL reçoit l'information provenant des cellules ganglionnaires. Le CGL possède une structure en couches (Figure 89), chacune ne recevant que les informations relatives à un seul œil. Les couches 1, 4, 6 pour un côté et 2, 3, 5 pour l'autre. Ces couches peuvent être parvocellulaires (couches 3, 4, 5 et 6) ou magnétocellulaires (couches 1 et 2) selon la taille des cellules qui les constituent.

Les structures qui sont observées depuis les cellules ganglionnaires jusqu'au CGL conduisent à penser qu'il existe des chemins visuels différents. Le chemin qui mène des cellules ganglionnaires « midget » aux couches parvocellulaires est appelé le chemin parvocellulaire. Celui qui mène des cellules ganglionnaires « parasol » aux couches magnétocellulaires est appelé le chemin magnétocellulaire. Les propriétés de ces chemins ont été étudiées sur les singes. Ils semblent jouer un rôle important dans la perception du mouvement (basse fréquence spatiale et haute fréquence temporelle).

L'information issue du CGL est envoyée dans le cortex visuel.

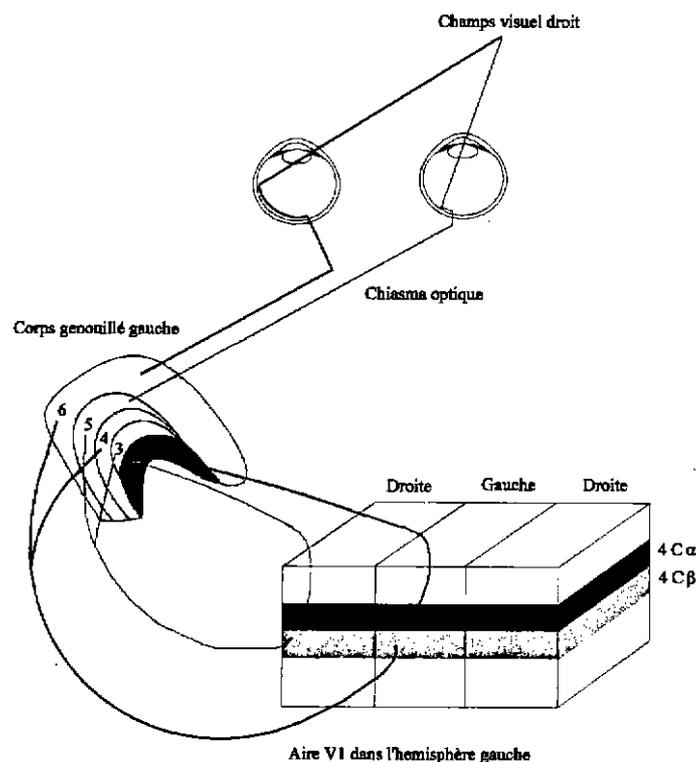


Figure 89 : Connexions des couches visuelles au CGL.

5) Le cortex visuel

Le cortex visuel ou aire V1, aussi appelé cortex strié en raison de son apparence, est situé à l'arrière du cerveau. Il reçoit l'information provenant du CGL. Il est constitué de 10^{10} neurones organisés en 6 couches, selon leur densité.

Le cortex visuel permet les mécanismes haut niveau de la vision : reconnaissance des formes, des textures, des couleurs, du mouvement...

A3 - Propriétés de la perception visuelle

Ces propriétés, spatiales ou fréquentielles sont issues d'expériences psychophysiques dans lesquelles on observe la visibilité d'un stimulus en fonction de son voisinage.

1) Perception de l'intensité lumineuse

L'observateur humain est sensible à une très large plage de luminances allant de $0,01 \text{ cd/m}^2$ dans le domaine scotopique (basses luminances) à plusieurs milliers de cd/m^2 en vision photopique (moyennes et hautes luminances). La luminance apparente d'un objet, bien que physiquement indépendante de la luminance des objets voisins, sera influencée par cette dernière. Notre perception est donc plus sensible à un contraste de luminance qu'à la valeur absolue de la luminance.

Le contraste entre deux niveaux de luminance L_{max} et L_{min} peut être défini par la formule de Michelson :

$$C = \frac{L_{max} - L_{min}}{L_{max} + L_{min}}$$

2) Lois de Weber

La perception de la variation de l'intensité lumineuse est donnée par l'expérience de Weber : Si un champ visuel de luminance uniforme L est divisé en deux selon une ligne et que l'observateur peut augmenter la luminance de l'une des deux parties d'une quantité variable ΔL , alors il ne remarque une variation que lorsque ΔL atteint une valeur précise non nulle. Cette valeur constitue le seuil de détection en luminance ou JND ("Just Noticeable Difference"). Si l'on répète cette expérience pour plusieurs valeurs de L , on remarque que le rapport $\Delta L/L$ est constant :

$$\frac{\Delta L}{L} = k$$

La loi de Weber traduit une perception logarithmique de l'œil à la lumière. La valeur de la constante k est de l'ordre de 0,01 à 0,02 suivant la méthode de test. Cependant cette loi n'est vraie que pour des luminances élevées.

3) Phénomène de Mach

Les liaisons latérales dans les diverses couches de la rétine sont à l'origine d'un phénomène étudié scientifiquement par E. Mach (1865) : lors de l'observation de deux régions uniformes de luminances physiques différentes juxtaposées (gris clair et gris foncé par exemple), on a l'impression que la luminosité augmente dans la région claire et diminue dans la région foncée autour de la ligne de séparation sur une distance que l'on appelle « bande de Mach » (Figure 90).

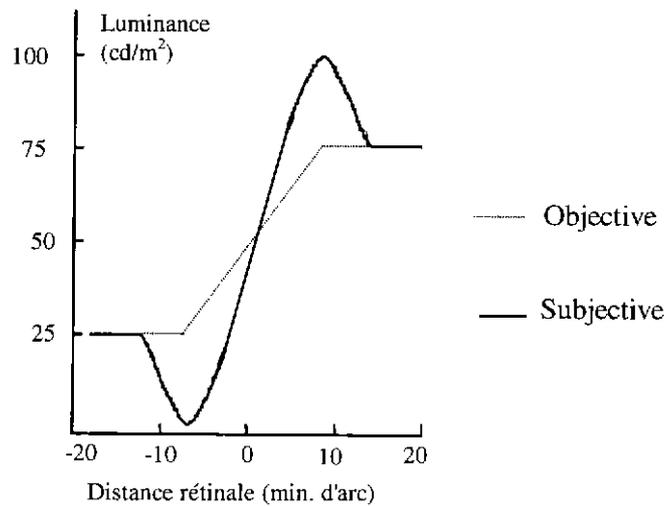


Figure 90 : Bandes de Mach, distribution subjective de la luminance le long d'un créneau.

4) Masquage spatial

Certaines expériences [LeF80], [Vas73] ont montré que la présence d'un fond non uniforme tend à augmenter le seuil de visibilité d'un stimulus visuel. Ce phénomène dit de masquage spatial traduit une diminution de la sensibilité visuelle à un stimulus au voisinage d'une variation importante d'intensité (la sensibilité est l'inverse du seuil). La Figure 91 montre la variation du seuil de visibilité d'une ligne en fonction de sa distance à un échelon et pour différents contrastes de celui-ci. On observe une augmentation du seuil de visibilité de chaque côté de l'échelon et un effet de masquage d'autant plus important que le contraste est élevé.

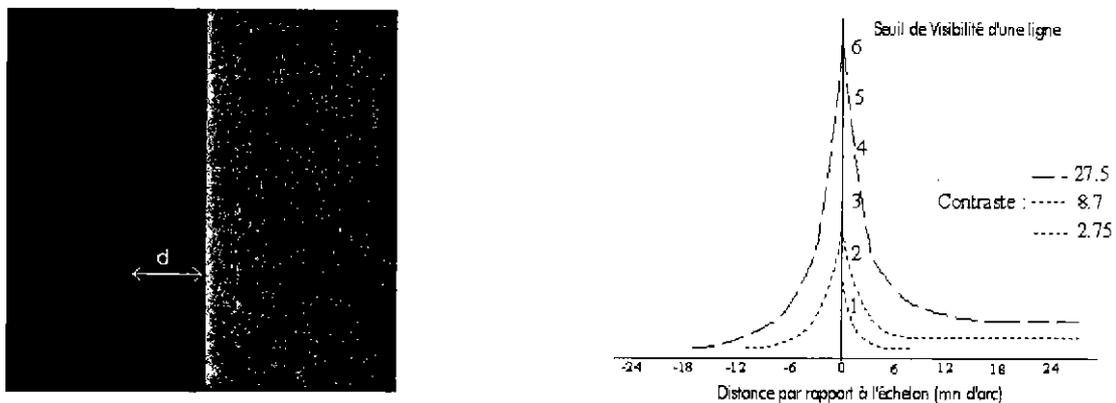


Figure 91 : Seuil de visibilité d'une ligne en fonction du contraste et de sa distance à l'échelon

5) Propriétés fréquentielles

a - Sensibilité aux fréquences spatiales

Une des premières expériences sur le comportement fréquentiel de l'œil est celle de Campbell et Robson [CaR68] qui ont étudié la détection par un observateur de mires sinusoïdales verticales fixes.

La mire sera détectée pour une valeur m_S (m seuil) ou seuil de visibilité. La courbe de sensibilité au contraste ou CSF ("Contrast Sensivity Fonction") est déterminée pour différentes valeurs de fréquences spatiales, elle est inversement proportionnelle à la valeur m_S .

La sensibilité visuelle est maximale pour des fréquences spatiales moyennes (de l'ordre de 3 cycles par degré), elle décroît pour les fréquences plus élevées. Ce filtrage est dû au comportement passe-bas du cristallin et de la pupille ainsi qu'au nombre limité de cellules rétiniennes par unité de longueur.

La Figure 92 représente la fonction de sensibilité au contraste. Les neurones ont la meilleure réponse pour les fréquences spatiales moyennes parce que les crêneaux lumineux tombent au centre du champ réceptif. Pour les hautes fréquences, le champ réceptif moyenne le stimulus tandis que pour les basses fréquences le stimulus au centre du champ réceptif s'oppose à celui du pourtour, diminuant la sensibilité.

La sensibilité fréquentielle se traduit par un comportement passe-bande du système visuel.

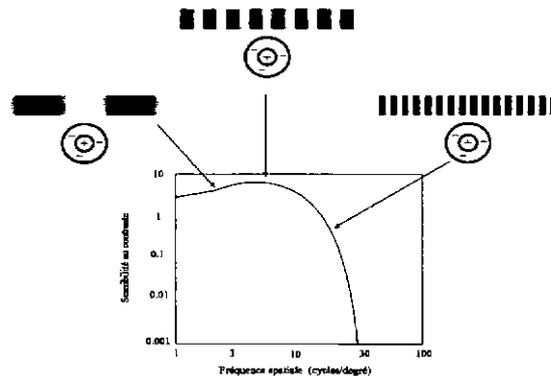


Figure 92 : Fonction de sensibilité au contraste.

b - Sensibilité à l'orientation du stimulus

Des mesures de sensibilité au contraste [CKL66] pour des mires sinusoïdales de différentes orientations, ont montré que l'œil est plus sensible aux directions horizontale et verticale. La sensibilité décroît jusqu'à une valeur minimale pour des directions à 45° [PhW84, WFP83].

c - Découpage en canaux fréquentiels

Plusieurs expériences psychophysiques et physiologiques ont montré l'existence de canaux fréquentiels dans le système visuel. [NSR71] observe que le système visuel détecte des fréquences spatiales de manière indépendante si elles sont différentes d'un facteur 1/2 ou 2. Ce type d'expérimentation est réalisé à l'aide de mires complexes, somme de deux mires sinusoïdales de fréquences différentes. On considère en général que les canaux fréquentiels ont une largeur d'une octave.

ANNEXE C - Quelques statistiques de l'image et leur interprétation

L'objectif de cette annexe est de rappeler brièvement les quelques statistiques élémentaires de l'image que nous avons utilisées ainsi que leur interprétation physique. A titre d'exemple, chacune d'entre elles sera calculée pour deux images différentes, l'une structurée de type « Ville », l'autre relativement uniforme de type « Champs » (Figure 93).

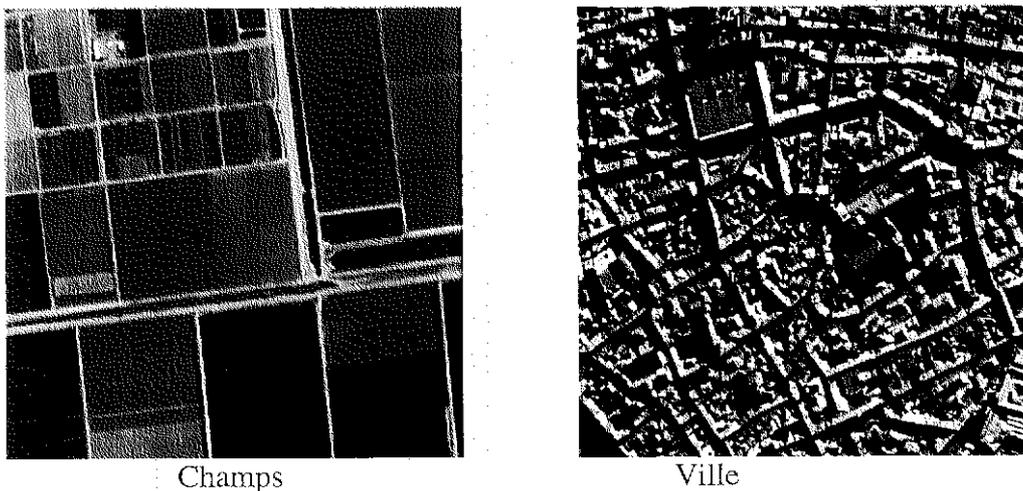


Figure 93 : Deux images de types différents : l'image de gauche est une image uniforme de type « champs », l'image de droite est une image structurée de type « ville ».

A1 - Histogramme

L'histogramme d'une image donne le nombre de pixels ayant un niveau de gris donné (Figure 94).

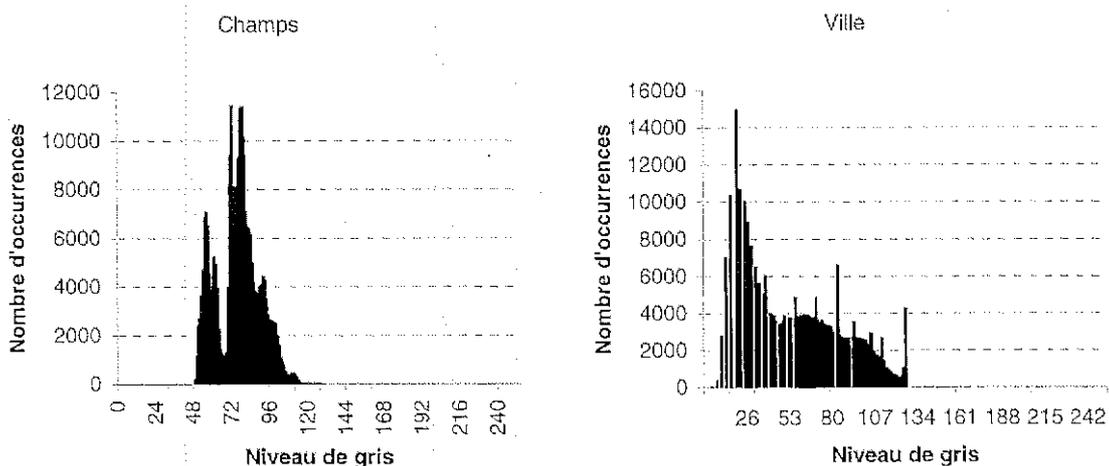


Figure 94 : Exemples d'histogrammes pour une image très structurée (Ville) et une image relativement uniforme (Champs).

Une image uniforme aura un histogramme resserré autour de la moyenne. L'histogramme d'une image contenant plusieurs structures uniformes comportera plusieurs pics. L'étalement

des niveaux de gris autour du pic correspond à la texture de la zone, au bruit contenu dans l'image...

Nous n'avons pas utilisé tel quel l'histogramme de l'image, inexploitable pour notre approche. En revanche, étant donné qu'il représente la fonction de répartition des niveaux de gris dans l'image, nous avons utilisé les moments de l'image.

A2 - Moments

Pour un vecteur X , on appelle espérance mathématique de X et on note $E[X]$ la moyenne des valeurs de X . On appelle moment d'ordre k de X la valeur $E[X^k]$ ou X^k désigne le vecteur constitué des toutes les composantes de X élevées à la puissance k .

L'image est une information spatiale bidimensionnelle échantillonnée. Les moments de l'image caractérisent la répartition des niveaux de gris dans l'histogramme. Etant donné qu'elle n'est pas spatialement infinie, on calcule des estimateurs des moments qui nous intéressent.

Dans notre travail, nous avons utilisé les quatre premiers moments de l'image :

La moyenne, dont l'estimateur sans biais est :

$$\text{moyenne} = \overline{\text{Im}} = \frac{1}{N} \sum_{i=1}^N \text{Im}(i)$$

La variance, dont l'estimateur sans biais est :

$$\text{variance} = \sigma^2 = \frac{1}{N-1} \sum_{i=1}^N [\text{Im}(i) - \overline{\text{Im}}]^2$$

On utilise aussi l'écart type $\sigma = \sqrt{\text{variance}}$. Cet indicateur caractérise l'écart moyen de la distribution par rapport à sa moyenne. En ce qui concerne les images, une image relativement uniforme a une variance faible, une image très structurée a une variance élevée.

Le moment d'ordre 3 ou skewness, dont l'estimateur sans biais est :

$$\text{skewness} = \frac{1}{N} \sum_{i=1}^N \left[\frac{\text{Im}(i) - \overline{\text{Im}}}{\sigma} \right]^3$$

Il caractérise la symétrie de l'histogramme des niveaux de gris autour de la valeur moyenne. Une distribution symétrique autour de sa moyenne a un skewness égal à zéro.

Le moment d'ordre 4 ou kurtosis, dont l'estimateur sans biais est :

$$\text{kurtosis} = \frac{1}{N} \sum_{i=1}^N \left[\frac{\text{Im}(i) - \overline{\text{Im}}}{\sigma} \right]^4 - 3$$

Le kurtosis caractérise la forme de pic ou l'aplatissement relatifs d'une distribution comparée à une distribution normale. Un kurtosis positif indique une distribution relativement pointue, tandis qu'un kurtosis négatif signale une distribution relativement aplatie.

A3 - Le variogramme

1) Définition

Le variogramme $\gamma(h)$ d'une image est une mesure de la variation moyenne des niveaux de gris (au carré) de deux points séparés d'un nombre de pixels h donné [Mat71, Hui75] :

$$\gamma(h) = \frac{1}{2} E \left[|\text{Im}(p+h) - \text{Im}(p)|^2 \right]$$

où $\text{Im}(p)$ est le niveau de gris du pixel p et E représente l'espérance mathématique.

La Figure 95 montre, pour les images « Ville » et « Champ », la représentation du variogramme.

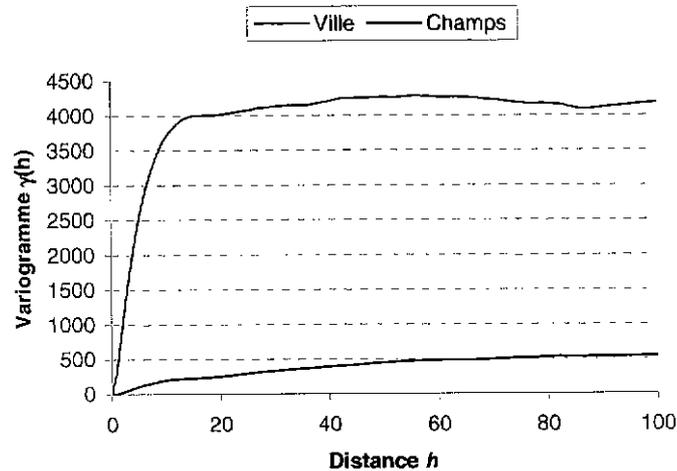


Figure 95 : Exemples de variogrammes pour un image très structurée (Ville) et une image relativement uniforme (Champs).

2) Propriétés

$\gamma(h)$ varie de 0 à une asymptote qui est la variance de l'image. En effet :

$$\begin{aligned} \gamma(h) &= \frac{1}{2} E \left[\text{Im}^2(p) + \text{Im}^2(p+h) - 2\text{Im}(p) \cdot \text{Im}(p+h) \right] \\ &= \frac{1}{2} E \left[\text{Im}^2(p) \right] + \frac{1}{2} E \left[\text{Im}^2(p+h) \right] - E \left[\text{Im}(p) \cdot \text{Im}(p+h) \right] \end{aligned}$$

Lorsque $h \rightarrow \infty$, $\text{Im}(p)$ et $\text{Im}(p+h)$ deviennent indépendant, donc :

$$E \left[\text{Im}(p) \cdot \text{Im}(p+h) \right] = E \left[\text{Im}(p) \right] \cdot E \left[\text{Im}(p+h) \right]$$

De plus, on a quel que soit h la relation :

$$E \left[\text{Im}(p) \right] = E \left[\text{Im}(p+h) \right]$$

On en déduit :

$$\lim_{h \rightarrow \infty} \gamma(h) = E \left[\text{Im}^2(p) \right] - E^2 \left[\text{Im}(p) \right] = \sigma^2$$

La tangente à l'origine du variogramme coupe l'asymptote pour une abscisse qui est la distance de corrélation statistique habituelle (Figure 96) :

$$d_c = \frac{\sigma^2}{\frac{d\gamma}{dh}(0)} \approx \frac{\sigma^2}{\gamma(1)}$$

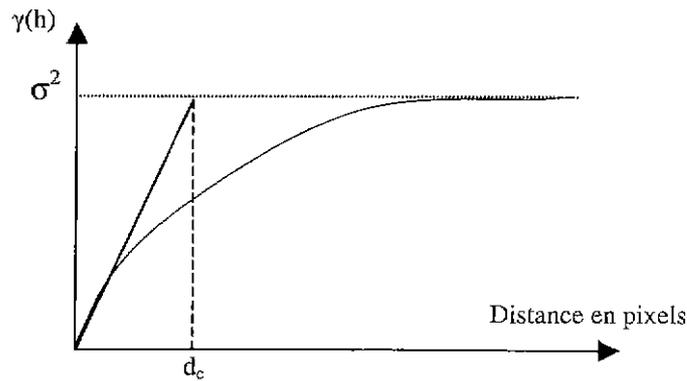


Figure 96 : Calcul de la distance de corrélation à partir du variogramme.

Pour l'image « Ville » cette distance vaut 11 pixels, alors que pour l'image « Champs » elle vaut 69 pixels.

D'autre part, certaines fluctuations locales peuvent être observées selon la structure de l'image.

3) Modélisation

Pour des images dont la texture est homogène, le variogramme est habituellement modélisé par une fonction du type [Rar89] :

$$\gamma(h) = A(1 - e^{-Bh})$$

où A représente la variance de l'image et B est caractéristique de la texture. Un tel modèle n'est pas suffisant pour décrire les variogrammes d'images quelconques, en particulier très structurées, dont la pente à l'origine et la variance sont élevées.

Nous avons choisi une modélisation à quatre paramètres, du type :

$$\gamma(h) = A(1 - e^{-Bh}) + C(1 - e^{-Dh})$$

En effet, elle permet de représenter aussi bien les variogrammes du type "ville" que ceux du type "champ", tout en conservant les informations pente à l'origine, asymptote, distance de corrélation (abscisse pour laquelle la tangente à l'origine coupe l'asymptote). Les paramètres A , B , C , D ont été calculés de manière à minimiser l'erreur quadratique moyenne entre le modèle et le variogramme. Les fonctions E_1 et E_2 étant permutables, nous avons choisi pour E_1 le couple (A, B) de paramètres correspondant à la variation lente, c'est-à-dire à longue distance sur l'image, et pour E_2 le couple (C, D) correspondant à la variation rapide, c'est-à-dire à courte distance.

Notons cependant que $A+C$ (asymptote du modèle) représente la variance de l'image et que $AB+CD$ (dérivée à l'origine du modèle) est inversement proportionnel à la distance de corrélation habituelle. La Figure 97 montre pour les image « Ville » et « Champ » que les fonctions E_1 et E_2 du modèle traduisent le comportement à l'origine et à l'asymptote du variogramme, c'est-à-dire justement la corrélation des pixels de l'image à courte et grande distances.

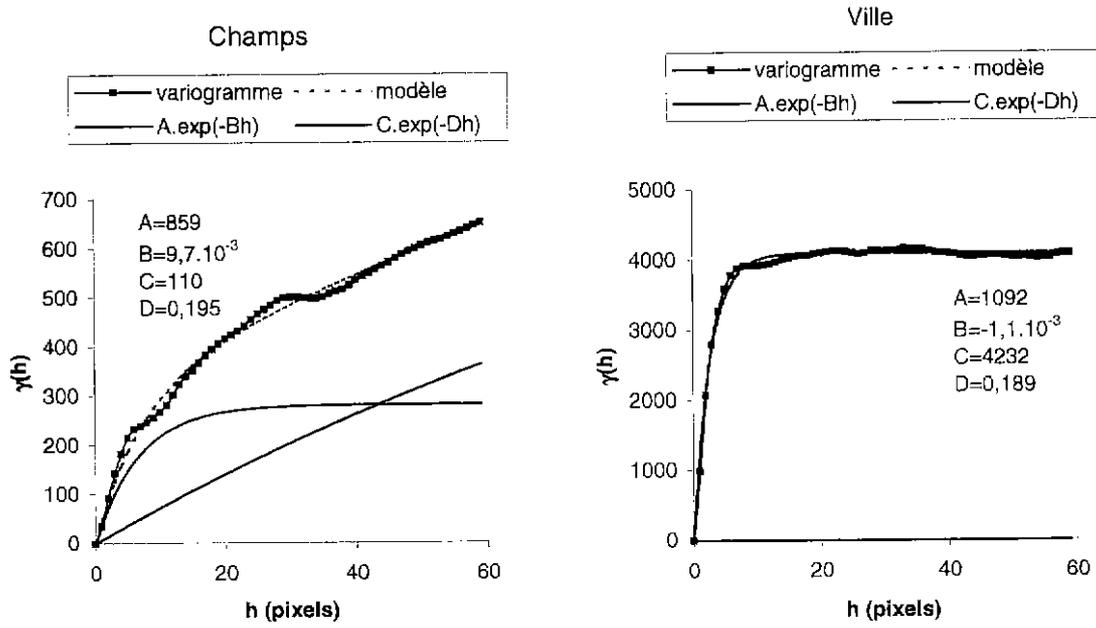


Figure 97 : Propriétés du modèle de variogramme.

ANNEXE D - Le codage JPEG

JPEG est un standard de compression d'image développé par le « Joint Photographic Experts Group ». Contrairement à beaucoup d'autres techniques de compression d'image (GIF, PGM, PPM,...), qui sont appelées « compressions entropiques » parce qu'elle s'effectuent sans perte d'information, le codage JPEG dégrade l'image de manière irréversible. Néanmoins, cette dégradation, adaptée aux caractéristiques du système visuel, permet de très forts taux de compression par rapport aux autres codages et ce, sans réelle gêne visuelle. La compression s'effectue en trois étapes, résumées sur la Figure 98 .

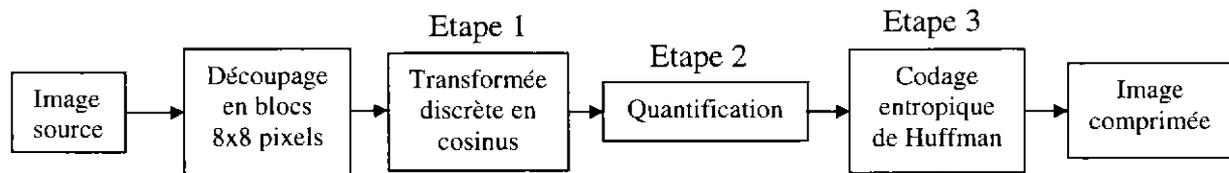


Figure 98 : Principe général de l'algorithme de compression JPEG.

Le découpage en blocs de 8x8 pixels ne constitue pas réellement une étape du processus de compression. En fait, dès le commencement du processus, l'image source est considérée comme un damier dont les cases font 8x8 pixels et chacun des trois traitements est appliqué sur chaque case de manière indépendante.

A1 - Etape 1: Transformée discrète en cosinus (DCT)

La transformée discrète en cosinus, de même que la transformée de Fourier discrète, est une bijection qui transforme un signal exprimé dans le domaine spatial (ou temporel) en sa représentation dans le domaine fréquentiel. Pour un bloc B de N lignes par N colonnes (N=8), où $B(i,j)$ est le niveau de gris du $j^{\text{ème}}$ pixel de la colonne i , la DCT est donnée en fonction des fréquences horizontale k et verticale l par :

$$TDC(k,l) = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} c(k)c(l)B(i,j) \cdot \cos\left[\frac{\pi k}{2N}(2i+1)\right] \cdot \cos\left[\frac{\pi l}{2N}(2j+1)\right]$$

avec

$$\begin{cases} c(k) = 1 & \text{si } k \neq 0 \\ c(k) = \frac{1}{\sqrt{2}} & \text{si } k = 0 \end{cases}$$

Lors de la première étape du codage JPEG, chacun des blocs 8x8 pixels est remplacé par sa DCT. Pour chaque bloc, le coin en haut à gauche de la DCT représente les basses fréquences contenues dans ce bloc, le coin en bas à droite représente les hautes fréquences. Pour une large majorité des scènes naturelles, une très grande partie des informations visibles se situe dans le domaine des moyennes et basses fréquences. C'est cette propriété qui est exploitée lors de la deuxième étape de la compression.

A2 - Etape 2 : quantification

L'étape de quantification est la seule étape irréversible du processus de compression JPEG. C'est à ce stade que le taux de compression, et donc la qualité de l'image comprimée, sont décidés.

Chaque coefficient de chaque bloc DCT (8x8 coefficients) est divisé par le coefficient correspondant d'une même matrice, 8x8 coefficients, dite matrice de quantification. Le résultat de la division est arrondi à l'entier le plus proche. Cette matrice contient des valeurs faibles dans son coin supérieur gauche et des valeurs plus grandes dans son coin inférieur droit. De cette manière les basses fréquences ne sont pas ou que peu altérées, alors que les coefficients correspondant à des hautes fréquences sont divisés par des nombres plus grands et sont souvent remplacés par des zéros. Ces zéros, les uns à côté des autres seront comprimés très efficacement lors de la dernière étape.

En pratique, lorsque on désire augmenter le taux de compression de l'image, l'algorithme utilise la même matrice de quantification multipliée par un coefficient adéquat. Le facteur de qualité, proposé par la plupart des programmes de compression JPEG, dépend directement de ce coefficient.

Néanmoins, ce facteur n'est pas prévu par la norme et chaque logiciel de compression le définit à sa manière !

A3 - Etape 3 : codage de Huffman

Une fois la matrice des coefficients DCT divisée par la matrice de quantification, chaque bloc de 8x8 coefficients va être comprimé de manière réversible par un codage entropique (sans perte).

Le coefficient DC de chaque bloc (en haut à gauche, représente la composante continue) a une valeur proche d'un bloc à un autre. Pour cette raison ils sont codés de manière différentielle.

Les coefficients AC (tous les autres) sont, pour chaque bloc, rangés par ordre de fréquence croissante et subissent un codage particulier (plage de zéros et code de Huffman).

Cette troisième étape n'influe pas sur la qualité de l'image.

ANNEXE E - Le système SPOT

A1 - Présentation de l'instrument

Le système SPOT [CNES1, <http://spot4.cnes.fr> ou <http://www.spotimage.fr>] est un système satellitaire d'observation de la Terre. Il est composé de satellites en orbite, d'installations au sol assurant la réception des images (au CNES) et d'un organisme assurant leur traitement et leur distribution (SPOTIMAGE).

Chaque satellite possède deux instruments de prise de vue capables d'acquérir des images en visée verticale ou en visée oblique.

Quatre satellites se sont succédés depuis le premier lancement en 1986 : SPOT2 en 1989, SPOT3 en 1993, et SPOT4 en 1998, le cinquième étant prévue pour 2001.

1) Le système imageur

L'instrument est de type push-broom : il est constitué d'une barrette de détecteurs et l'image est acquise ligne par ligne, la direction des colonne étant celle du déplacement du satellite.

a - Caractéristiques de l'optique

Caractéristiques communes aux instruments HRV des satellites 1, 2, 3 :

- Angle de champ de l'instrument : 4,13° ;
- Nombre d'ouverture : F/3,3 (focale 1082 mm) ;
- Largeur du champ au sol en visée verticale : 60 km. Lorsque les deux instruments HRV1 et HRV2 sont jumelés le champ couvert est de 117 km assurant ainsi un recouvrement de 3 km entre les champs individuels ;
- Largeur du champ au sol en visée oblique : 60 à 81 km. Les régions observés peuvent être situées dans un couloir de 950 km ;
- Bandes spectrales : en mode multibande (XS)
 - bande B1 (vert : 0,5 – 0,59 μm) ;
 - bande B2 (rouge : 0,61 – 0,68 μm) ;
 - bande B3 (proche IR : 0,79 – 0,89 μm) ;en mode panchromatique (P), bande unique (0,51 – 0,73 μm) ;
- Pas d'échantillonnage en visée verticale : 10 m en mode P, 20 m en mode XS.

Nouveautés de SPOT4 :

- Bande spectrale dans le moyen infrarouge (MIR : 1,58 – 1,75 μm) : instruments HRVIR de 20 m de résolution ;
- Protection des HRVIR contre les phénomènes de polarisation et d'éblouissement
- Suppression de la bande P, remplacée par le mode monospectral, réalisé dans la bande B2, avec une résolution de 10 m ;
- Un nouvel instrument : Végétation, qui donne une image mondiale quotidienne de résolution 1 km.

b - Caractéristiques électroniques

- Débit d'informations : 25 Mbits/s par canal soit 50 Mbits/s au total
- Codage des mesures correspondant à un pixel :
 - 3x8 bits en mode XS ;
 - DPCM en mode P.

c - Description fonctionnelle

La partie embarquée de la chaîne image comprend les fonctions de formation et de détection de l'image, ainsi que l'émission des données vers la Terre.

i - Optique

La Figure 99 décrit l'optique de la caméra HRV.

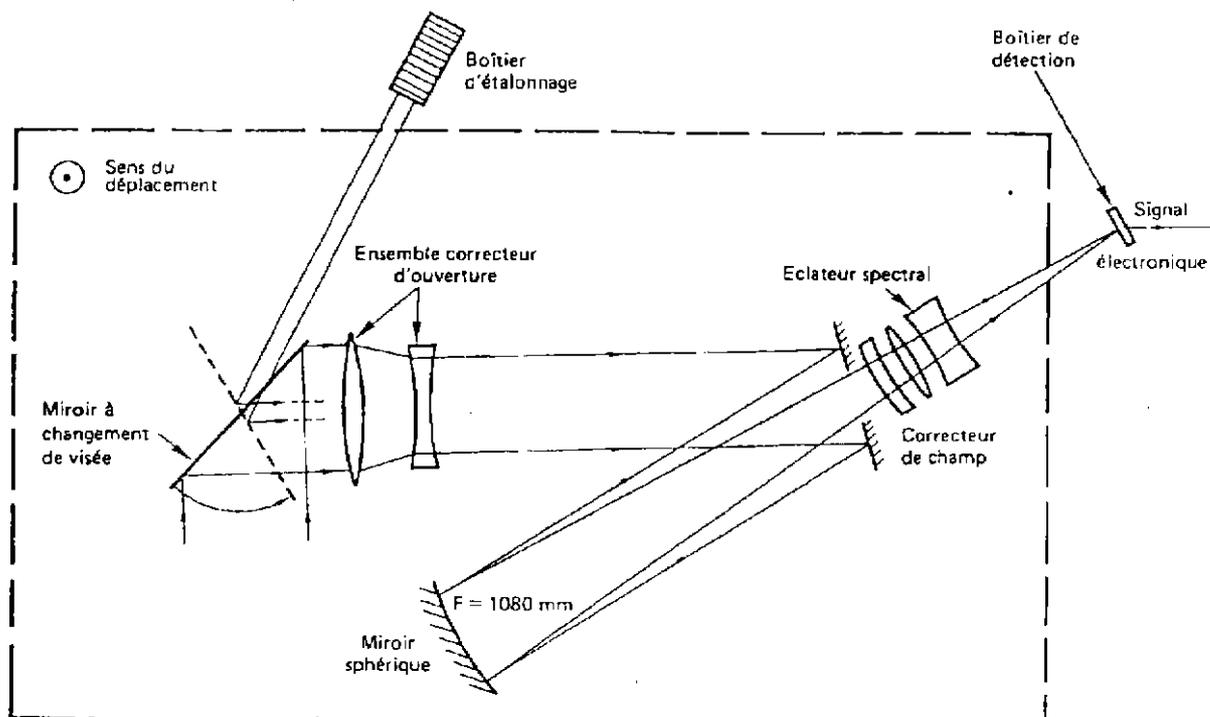


Figure 99 : Décomposition fonctionnelle du système optique

Les visées obliques sont effectuées au moyen d'un miroir orientable appelé miroir de changement de visée (MCV) pouvant prendre 91 positions différentes. Chaque position diffère de $0,6^\circ$ de la précédente. Le MCV possède en outre une position « étalonnage », qui permet de viser un boîtier d'étalonnage.

En mode multibande, une même ligne d'image est analysée simultanément dans chacune des trois bandes. Un éclateur spectral permet d'effectuer la séparation (Figure 100).

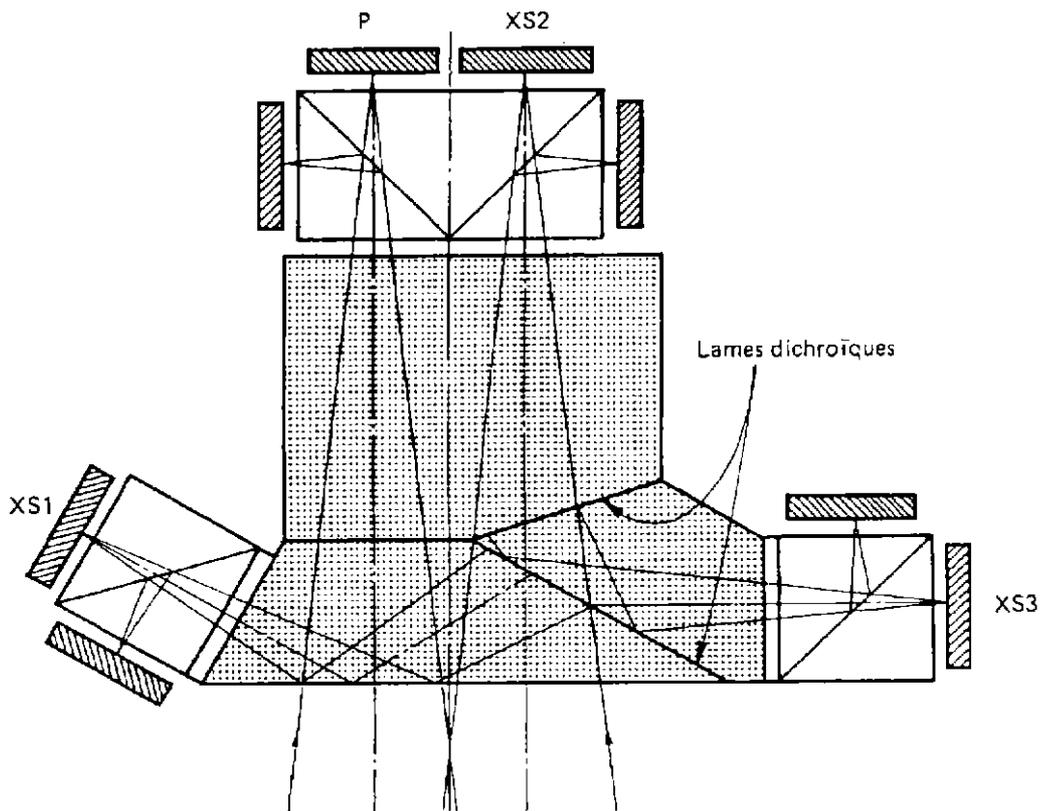


Figure 100 : Principe de la séparation spectrale pour SPOT 1,2 et 3.

ii -Détection

Les détecteurs CCD sont composés de 6000 photodiodes assemblées en ligne et espacées de $13 \mu\text{m}$. Ils permettent d'analyser d'un seul coup une ligne de paysage de 60 km de longueur orientée perpendiculairement à la trace au sol du satellite. En raison des limitations technologiques à l'époque de la conception, les photodiodes sont réparties sur quatre barrettes collées sur un diviseur optique de ligne (DIVOLI), de manière à éviter les discontinuités (Figure 101).

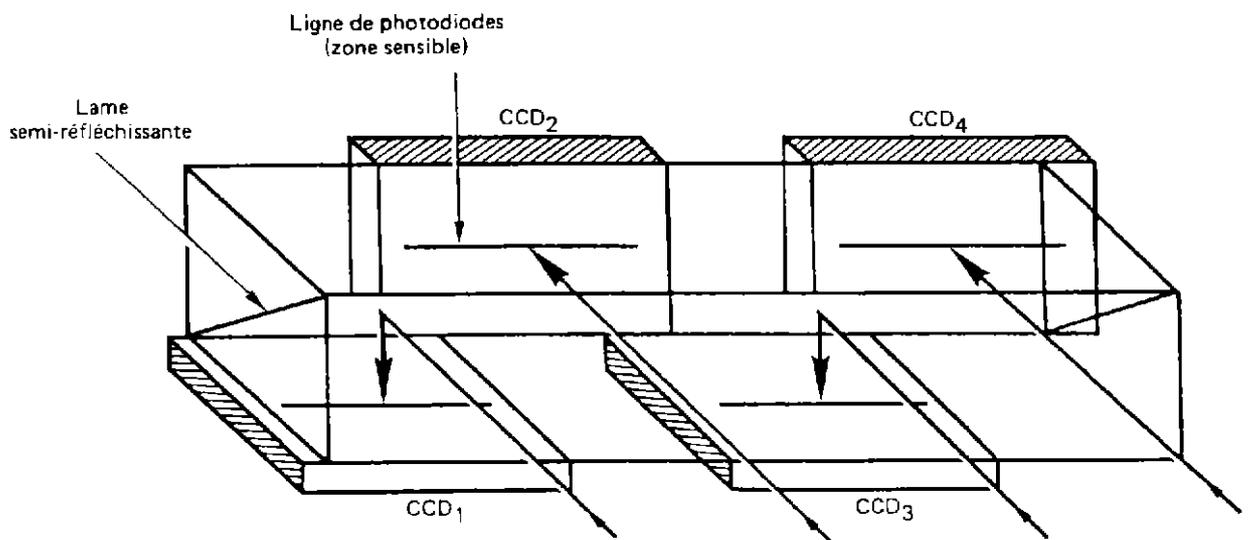


Figure 101 : Montage des 4 barrettes de détecteurs sur un DIVOLI

Pour chaque instrument, 4 diviseurs optiques de ligne permettent de réaliser 4 rangées de 6000 détecteurs sans discontinuité (Figure 102).

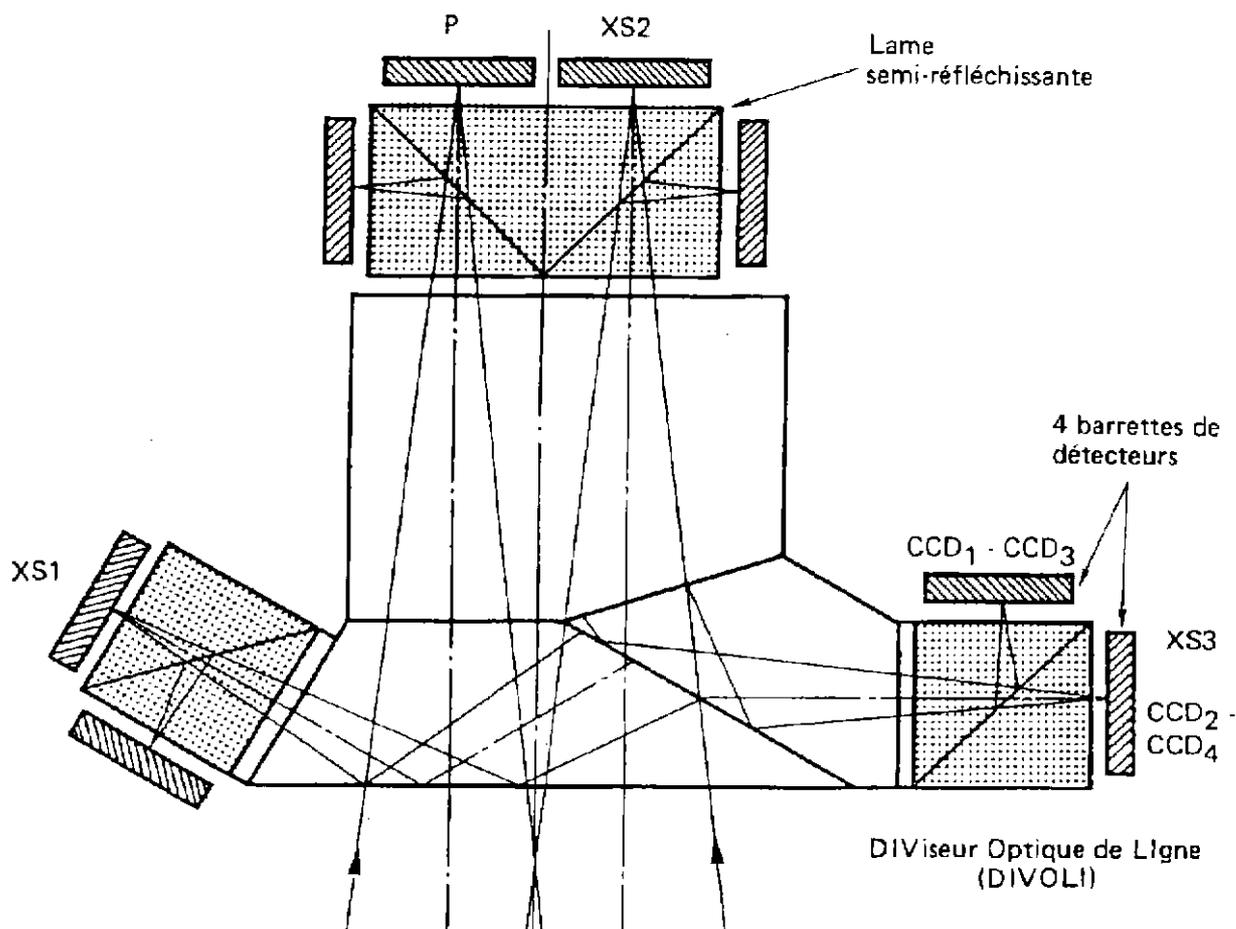


Figure 102 : Les 4 diviseurs optiques de ligne (DIVOLI)

iii -Amplification

Pour chaque bande spectrale, la voie d'amplification comprend huit valeurs discrètes de gain étagées selon une progression géométrique de raison 1,3.

iv -Numérisation

En mode XS, la numérisation est réalisée suivant un codage linéaire sur 8 bits. En mode P, deux possibilités existent : soit un codage linéaire sur 6 bits (peu utilisé), soit un codage DPCM (Differential Pulse Code Modulation) qui tire parti de la redondance de l'information.

v -Etalonnage

La Figure 103 présente le principe d'étalonnage de l'instrument.

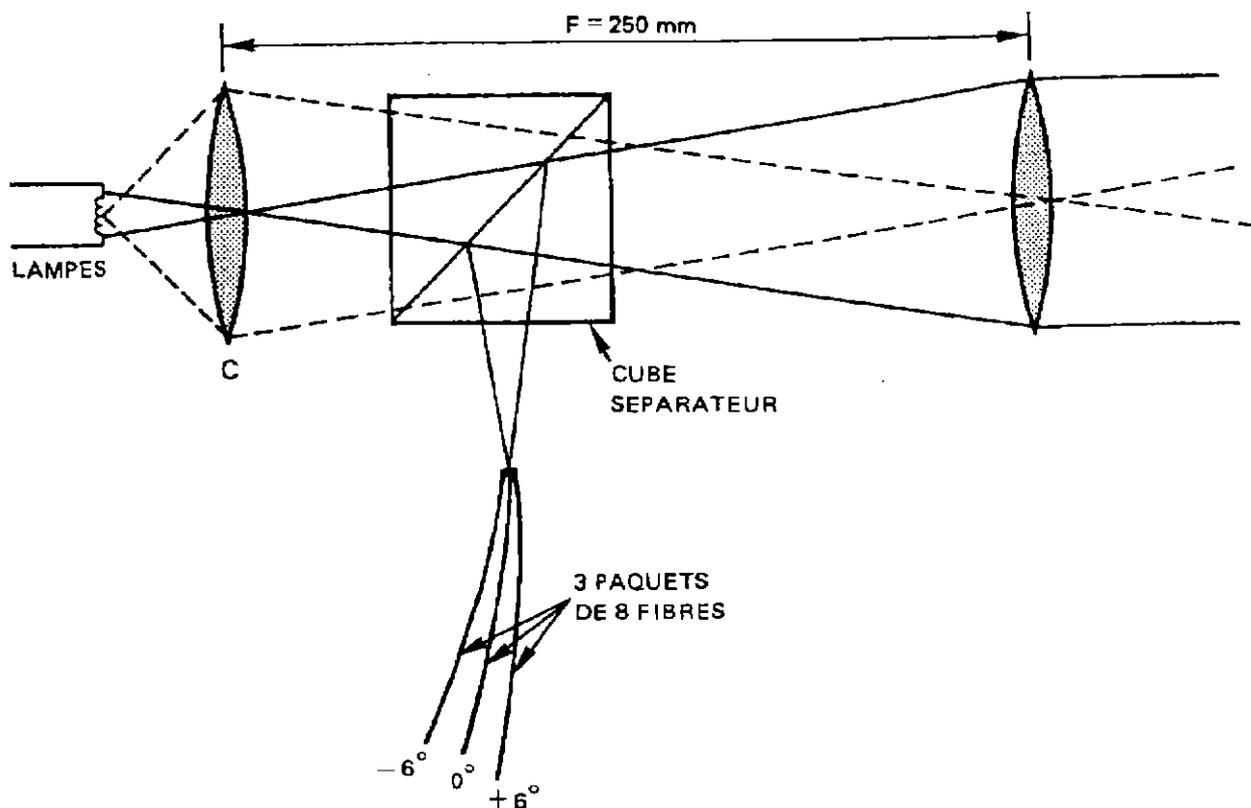


Figure 103 : Boîtier d'étalonnage

Deux sources lumineuse différentes peuvent être utilisées pour réaliser l'étalonnage : une lampe étalonnée, permettant d'éclairer l'ensemble des détecteurs pour un étalonnage relatif intrabande, ou un capteur solaire à fibres optiques permettant d'éclairer seulement quelques détecteurs dans chaque bande, pour un étalonnage absolu.

L'étalonnage relatif :

Chacun des détecteurs possédant sa propre sensibilité, il est nécessaire d'uniformiser le tout en apportant une correction en chaque point de la ligne. L'objectif de cet étalonnage est donc de déterminer les coefficients d'égalisation à appliquer en mesurant les courants d'obscurité et les gains interdétecteurs.

L'étalonnage absolu :

Il a pour but de relier la sortie numérique du détecteur à la grandeur physique correspondante, c'est-à-dire la luminance hors atmosphère de la scène observée. L'étalonnage interbande permet de retrouver la luminance de la scène sur les différentes bandes, alors que l'étalonnage multitemporel permet de comparer la luminance d'images prises dans la même bande à des dates différentes.

Les dispositifs bords sont complétés par l'exploitation d'images de sites uniformes (Groenland par exemple) ou de sites dont la réflectance est mesurée au sol simultanément à la prise de vue.

2) Conditions de prise de vue – La scène SPOT

L'image SPOT est une image haute résolution, en couleurs ou en noir et blanc. La prise de vue, que ce soit en visée verticale ou oblique est présentée sur les Figure 104 et Figure 105.

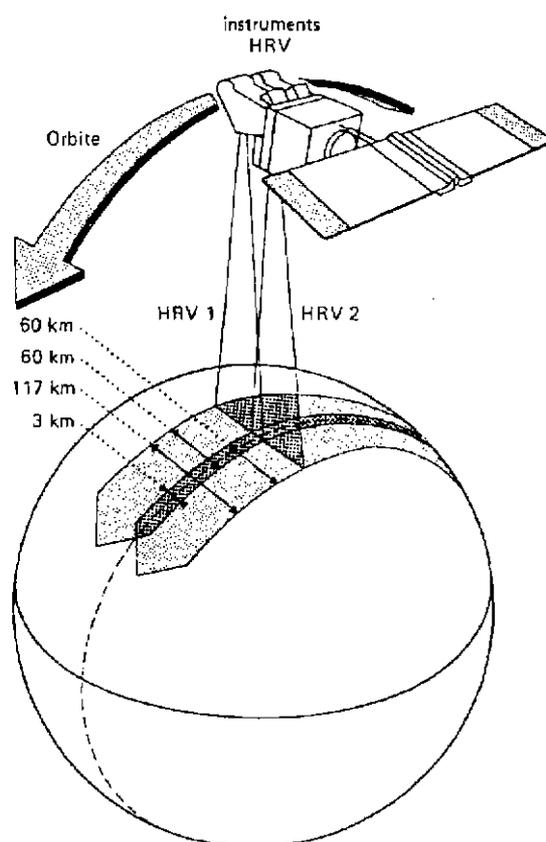


Figure 104 : Prise de vue en visée verticale

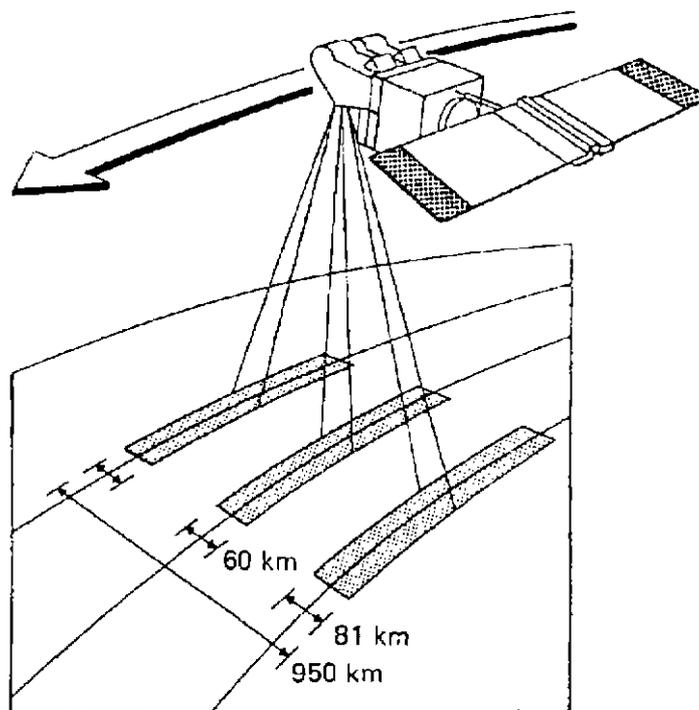


Figure 105 : Prise de vue en visée oblique

On considère que l'instrument fonctionne en visée verticale lorsque l'angle de prise de vue est compris entre $-7,5^\circ$ et $7,5^\circ$. Au cours du déplacement du satellite, chaque instrument produit

un ruban continu de largeur 60 à 81 km, qui est découpé en scènes de 60 km de longueur. Une scène brute comporte 6000*6000 pixels en mode P et 3000*3000 pixels dans chacune des bandes du mode XS.

Le cycle orbital est de 26 jours pour le satellite SPOT, mais la possibilité d'utiliser des angles de visée obliques permet une observation stéréoscopique d'une même scène à des dates proches, application particulièrement intéressante pour la restitution du relief, par exemple.

A2 - Modèle radiométrique

Pour les instruments HRV et HRVIR, la réponse de chaque pixels est donnée par [KMB98] :

$$X(k, m, p) = A(k).G(m, k). \gamma(k, m, b).g(k, p).L(k, p) + C(k, m, p)$$

où $X(k, m, p)$ est la réponse du pixel p de la barrette b dans la bande k à la luminance d'entrée $L(k, p)$ quand le gain sur l'image est m . $G(m, k)$ sont les gains (programmables), $A(k)$ sont les coefficients d'étalonnage absolu, $C(k, m, p)$ sont les courants d'obscurité, $g(k, p)$ sont les coefficients de normalisation inter-détecteurs et inter-colonnes, et $\gamma(k, m, b)$ sont des coefficients de normalisation interbarrette.

A3 - Prétraitement des images

L'image SPOT brute, dite de niveau 0, est entachée d'imperfections liées aux conditions de prise de vue et doit subir un prétraitement de manière à la rendre exploitable. Il existe quatre niveaux standardisés de correction : 1A, 1B, 2, S.

1) Niveau 1A

Une égalisation des détecteurs est réalisée sur l'image brute. Ce niveau est destiné à des utilisations nécessitant des informations ayant subi le minimum de traitements, comme la stéréorestitution et les études radiométriques fines, par exemple.

2) Niveau 1B

Ce niveau comprend le niveau 1A auquel sont rajoutées des corrections géométriques liées aux déformations systématiques introduites par le système (rotation de la Terre, effet panoramique).

La spécification quant à la précision de localisation sur l'image corrigée est de 1500 m en visée verticale, pour une distorsion interne est inférieure à 10^{-2} .

3) Niveau 2

C'est un niveau de prétraitement de précision, qui comprend le niveau 1B auquel sont rajoutées des corrections géométriques bidimensionnelles grâce à des points d'appui : l'image devient superposable à une carte. Ce niveau ne prend pas en compte les déformation dues au relief ; le produit est donc d'autant plus précis que l'angle de visée est proche de la verticale et l'altitude faible.

Les scènes de niveau 2 sont rééchantillonnées et orientées au nord cartographique.

4) Niveau S

Pour ce niveau, la scène devient superposable à une autre scène de référence avec une précision de 0,5 pixels lorsque l'angle de prise de vue est le même. Ce produit est utilisé pour des études multitemporelles.

A4 - Qualité des images

Deux types distincts de défauts peuvent affecter les images SPOT. Certains sont liés au système lui-même, à ses imperfections ou son environnement, alors que d'autres sont générés par la chaîne de production des images.

1) Défauts système

a - Drop-out bord

Le drop-out bord est une perte d'information sur une ligne de l'image, provenant d'un défaut de fonctionnement sur l'enregistreur à bord du satellite. Il apparaît sur l'image SPOT comme une succession de points alignés très visibles.

Il n'existe aucune possibilité de corriger cette anomalie.

b - OVNI [DiH90]

L'OVNI est une élévation anormale du signal de quelques détecteurs contigus (en moyenne 5, mais peut atteindre 20 détecteurs) et ceci dans une seule bande spectrale. Ces défauts sont quatre fois plus nombreux dans la bande XS2 que dans les autres bandes. Contrairement aux drops, les valeurs erronées ne sont pas distribuées aléatoirement mais varient de manière continue et peuvent atteindre 80 pas CAN.

Une étude systématique d'un grand nombre d'images a permis de montrer que ce défaut apparaît exclusivement sur des images prises au dessus de l'Amérique du sud, indépendamment de la date de prise de vue. Il est monochrome, existe sur toutes les bandes (surtout XS2), et à tous les niveaux de traitement (0, 1A, 1B).

Ce défaut apparaît en fait lorsque le satellite traverse la ceinture de Van Allen, ceinture de particules chargées (électrons, protons, quelques ions), probablement en provenance du vent solaire, qui ont été capturées par la magnétosphère terrestre et qui sont distribuées le long des lignes de champ. L'interaction de certaines particules avec le détecteur peut saturer certains pixels et déborder sur les pixels voisins.

c - Défaut résiduel d'égalisation sur des images de faible réflectance

Sur des images de faible réflectance (mer, forêts), ce défaut apparaît sous la forme de griffures aisément discernables à l'œil nu.

L'égalisation permet de corriger les différences de gain entre les détecteurs pour un niveau de luminance donnée, le plus souvent assez fort. Le défaut est dû à la non linéarité différentielle qui existe d'un détecteur à l'autre le long de la barrette. Les coefficients d'égalisation, calculés pour des fortes réflectances, ne sont plus exactement les mêmes pour des réflectances plus faibles.

d - Défaut résiduel d'égalisation sur des images de forte réflectance

Sur des images de forte réflectance, il se matérialise par des décalages des niveaux moyens des différentes barrettes compris entre un et quatre pas de quantification dans le canal XS1. C'est le cas par exemple lors de l'observation de paysages désertiques.

Ce défaut est essentiellement provoqué par la différence de polarisation qui existe entre la lumière provenant de la scène d'étalonnage (souvent de la banquise) et celle provenant d'un paysage différent.

e - Couplage parasite P+XS

Les images prises en mode P+XS présentent une structure régulière : tous les 7 pixels en mode XS, et tous les 16 pixels avec un décalage de 7 ou 9 pixels entre les deux registres en

mode P. Cet effet d'amplitude de 1 à 3 pas CAN provient d'un phénomène de couplage qui affecte les courants d'obscurité des modes P et XS d'un même instrument.

En appliquant des courants d'obscurité mesurés en mode P+XS on peut éliminer cette structure. Les images postérieures au 16 novembre 1986 sont automatiquement corrigées.

f - Bruit structuré pair impair selon les colonnes

Les détecteurs pairs et impairs transfèrent leurs charges vers deux registres différents. Si ces deux registres ne sont pas parfaitement égalisés, un bruit structuré pair impair, selon les colonnes apparaît sur l'image. Ce bruit est en général négligeable devant le bruit d'égalisation.

g - Inefficacité de transfert

Au cours du transfert des électrons de la barrette vers le registre (transfert vertical), certains d'entre eux peuvent être perdus et contaminer des pixels voisins. Ce phénomène apparaît sous la forme d'un colonnage plus ou moins gênant.

D'autre part, le même phénomène lors de la lecture du registre (transfert horizontal), peut occasionner une accumulation de charges en sortie du registre. Ce phénomène apparaît sous la forme d'une hausse anormale du niveau des pixels correspondants à l'extrémité de la barrette.

h - Bruit structuré pair impair selon les lignes

Une légère fluctuation de l'horloge peut modifier le temps d'intégration de chaque ligne. Ce défaut apparaît sous forme de lignage plus ou moins gênant sur l'image.

i - Fluctuation des courants d'obscurité

Les 9000 détecteurs permettant de générer l'image en mode XS ou les 6000 en mode P possèdent chacun leurs propres gain, courant d'obscurité (sortie pour une entrée nulle) et donc coefficient d'égalisation correspondant, déterminé pendant la phase d'étalonnage (cf. v -) effectuée régulièrement. Tous les coefficients sont archivés et permettent de corriger les images.

Toutefois, les coefficients utilisés pour corriger l'image ne sont pas forcément les mieux adaptés au type de scène et à l'instant de prise de vue. Il en résulte un bruit le long des lignes de détecteurs qui apparaît comme un colonnage dans l'image égalisée.

2) Défauts de production

a - Notation de couverture nuageuse

Au moment de l'inventaire des images, chacune d'entre elles reçoit une note de couverture nuageuse. A cette fin, chaque octant de scène est noté de 0 à 2 de la manière suivante :

- 0 si la couverture nuageuse est inférieure à 10 % ;
- 1 si la couverture nuageuse est comprise entre 10 % et 25 % ;
- 2 si la couverture nuageuse est supérieure à 25%.

La note de l'image est l'entier de huit chiffres constitué par les huit notes côte à côte.

Il peut arriver que le notateur fasse une mauvaise estimation de la couverture nuageuse, suite à des ambiguïtés.

b - Drop out sol

C'est une perte d'information sur une ligne provenant de défauts enregistreur au sol (mauvais état des têtes de lecture, mauvais état de surface des bandes...). Elle affecte les quatre barrettes de façon identique et apparaît comme une suite de points faux alignés très visibles.

Si un drop out sol touche deux pistes adjacentes, l'espacement entre les deux têtes étant de 1,5 pouces, le défaut sera visible sur deux lignes distantes de 7 pixels en mode XS et 14 pixels en mode P.

c - Perte de synchronisation

Il peut y avoir une perte de synchronisation lors de l'acquisition des scènes. Elle se traduit par une perte de lignes sur l'image.

Au niveau de prétraitement 1A ces lignes sont remplacées par des zéros, et le fait est signalé dans les fichiers d'accompagnement auxiliaires. Aux niveaux 1B et 2, dans la mesure où le nombre de lignes manquantes est inférieur à quatre, elles seront interpolées grâce à celles qui les encadrent. Si le nombre de lignes manquantes est supérieur à quatre, les points sont mis au niveau zéro.