



Codage hybride adaptatif d'images en vue de la transmission à bas débit : logiciel et matériel

Jamal Baina

► To cite this version:

Jamal Baina. Codage hybride adaptatif d'images en vue de la transmission à bas débit : logiciel et matériel. Sciences de l'ingénieur [physics]. Université Paul Verlaine - Metz, 1992. Français. NNT : 1992METZ004S . tel-01775962

HAL Id: tel-01775962

<https://hal.univ-lorraine.fr/tel-01775962>

Submitted on 24 Apr 2018

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

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



AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : ddoc-theses-contact@univ-lorraine.fr

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4

Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

http://www.cfcopies.com/V2/leg/leg_droi.php

<http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm>

THESE

soutenue le 23 janvier 1992
pour obtenir le titre de
DOCTEUR DE L'UNIVERSITE EN ELECTRONIQUE

CODAGE HYBRIDE ADAPTATIF D'IMAGES EN VUE DE LA TRANSMISSION A BAS DEBIT: LOGICIEL ET MATERIEL.

par

Jamal BAINA

Président du jury
Rapporteurs

M.
MM.

A. TOSSER-ROUSSEY ENIM Metz.
H. MAITRE ENST Paris.
A. DIOU IUT Le Creusot.
D. FLAENDER Directeur du CERLOR.

Examineurs

MM.

C. LHERMITTE Supélec Metz.
V. THOMAS CCETT.

Invités

MM.

R. KLEIM Université Metz.
M. BENNOUNA EMI Mohammed V Rabat.
H. CAUDRON TDF CERLOR.
P. JEAN TDF CERLOR.

S/M3 92/4

BIBLIOTHEQUE UNIVERSITAIRE - METZ	
N° inv.	1992 0845
Cote	S/M3 92/4
Loc	Magasin

THESE

soutenue le 23 janvier 1992
pour obtenir le titre de
DOCTEUR DE L'UNIVERSITE EN ELECTRONIQUE

CODAGE HYBRIDE ADAPTATIF D'IMAGES EN VUE DE LA TRANSMISSION A BAS DEBIT: LOGICIEL ET MATERIEL.

par

Jamal BAINA

Président du jury
Rapporteurs

M.
MM.

A. TOSSER-ROUSSEY ENIM Metz.
H. MAITRE ENST Paris.
A. DIOU IUT Le Creusot.
D. FLAENDER Directeur du CERLOR.
C. LHERMITTE Supélec Metz.
V. THOMAS CCETT.
R. KLEIM Université Metz.
M. BENNOUNA EMI Mohammed V Rabat.
H. CAUDRON TDF CERLOR.
P. JEAN TDF CERLOR.

Examineurs

MM.

Invités

MM.

Ce rapport présente un travail qui a été réalisé au Centre d'Etudes et de Recherche de Lorraine de TéléDiffusion de France (CERLOR) et au laboratoire Mécatronique Industrielle de l'Université de Metz.

Je suis très reconnaissant à Monsieur Didier Flaender, directeur du CERLOR, de m'avoir accueilli dans ce centre de recherche et d'avoir accepté d'être rapporteur de cette thèse.

Que Messieurs Hervé Caudron, directeur adjoint du CERLOR et responsable du laboratoire T.M.I. dans lequel j'ai fini mes travaux, et Philippe Jean, responsable du laboratoire T.E.C., qui m'a permis de commencer cette étude, soient remerciés pour les importants moyens et les facilités qu'ils ont mis à ma disposition. La confiance qu'ils me témoignent en siégeant en tant qu'invités à ce jury, me touche profondément.

Je souhaite que Monsieur le Professeur André Tosser-Roussey, professeur à l'Université de Metz et directeur du laboratoire Mécatronique Industrielle, reçoive ma profonde gratitude pour m'avoir intégré à son équipe de recherche et pour la confiance qu'il m'a accordée et qui se traduit aujourd'hui par la direction de ce jury.

Je suis très honoré de la participation de Monsieur Henri Maître, directeur du département Image à l'E.N.S.T., à mon jury et je le remercie très sincèrement pour ses conseils et pour son rapport.

Que Monsieur Michel Bruynoogh soit remercié d'être rapporteur de cette thèse.

Monsieur Vincent Thomas, chef de la cellule Codage d'Images au C.C.E.T.T., a bien voulu siéger dans ce jury, je l'en remercie vivement.

Je remercie Monsieur Claude Lhermitte, chef du service Informatique à Supélec, pour sa présence dans ce jury.

Monsieur Mustapha Bennouna, chef du département Traitement du Signal et Automatique à l'Ecole Mohammadia d'Ingénieurs, est venu de très loin

pour représenter un pays qui est cher à mon coeur, je l'en remercie beaucoup et j'espère que sa venue est l'aube d'une belle coopération entre le CERLOR et l'E.M.I.

Je remercie Monsieur Roland Kleim, Professeur à l'Université de Metz, de bien avoir voulu voulu siéger dans ce jury.

J'ai le vif plaisir de remercier Monsieur Arnaud Laprèvote avec lequel j'ai formé une équipe de choc tant pour la créativité que pour la bonne humeur.

Je désire que l'ensemble du personnel du CERLOR ainsi que chaque membre du laboratoire Mécatronique Industrielle soient remerciés pour l'excellente ambiance dans laquelle ce travail a été réalisé.

Les dispositifs technologiques et l'algorithme de traitement présentés dans ce mémoire sont cités dans deux demandes de brevets d'inventions, dont les intitulés sont :

- "***Une méthode et un système pour la mise au point automatique d'une caméra utilisant la TCD***". Brevet déposé le 11 avril 1991 au nom de TéléDiffusion de France S.A.. N°91 04420.
- "***Un procédé de codage de signaux d'image hybride adaptatif***". Brevet déposé le 24 juillet 1991 au nom de TéléDiffusion de France S.A.. N°91 09359.

"Entia non sunt multiplicanda praeter necessitatem * "
Guillaume d'Ockham philosophe franciscain 1270-1349.

*** Il ne faut pas multiplier les entités au-delà de ce qui est
nécessaire**

Chapitre I.....	1
INTRODUCTION	1
I - A Contraintes.....	1
I - B Solutions retenues	2
Chapitre II.....	6
TECHNIQUES CLASSIQUES DE CODAGE D'IMAGES	6
II - A Méthodes prédictives	6
II - A - 1 Le prédicteur des échantillons du signal image.....	8
II - A - 2 Le quantificateur de l'erreur de prédiction	8
II - A - 3 Défauts générés par la MICD	10
Le bruit granulaire.....	10
Le bruit de faux contour	10
Le bruit de traînage.....	10
Le flottement de contour.....	11
II - B Codage par transformations mathématiques	12
II - B - 1 Formulations mathématiques.....	12
II - B - 2 Transformée en cosinus discrète	14
Définition	14
Evaluation de la dynamique des composantes de la	
matrice transformée.....	15
II - B - 3 Codage par seuillage	17
II - B - 4 Images adaptées au codage par transformation	18
II - C Codage par quantification vectorielle.....	18
II - C - 1 Principe.....	18
II - C - 2 Formation des vecteurs.....	20
Techniques de normalisation pour la quantification	
spatiale	21
II - C - 3 Construction du dictionnaire.....	23
Nuées dynamiques.....	23
Algorithme de classification du dictionnaire LBG	25
Algorithme de Linde Buzo Gray.....	27
Méthode de Kohonen.....	28
II - C - 4 Distance déterminant la distorsion entre les	
vecteurs.....	29
Chapitre III.....	30
METHODES ADAPTATIVES DE CODAGE D'IMAGES	30

III - A Méthode prédictives	30
III - A - 1 MICDA-ER	31
III - B Techniques adaptatives de codage par transformation	
TCD	34
III - B - 1 Adaptation par variation de la taille des blocs	
transformés	34
III - B - 2 Stratégies adaptatives de codage de coefficients	
transformés	35
III - B - 3 Classification basée sur des paramètres	
énergétiques	35
III - C Approches adaptatives pour la quantification vectorielle.....	43
III - C - 1 Accommodation des tailles de bloc aux images	
codées	43
III - C - 2 Quantification vectorielle multi-étages	45
III - C - 3 Quantification vectorielle à dictionnaire	
rafraîchi	45
Chapitre IV	47
TECHNIQUES DE CODAGE PAR CLASSIFICATION VISUELLE	47
IV - A Quantification vectorielle	48
IV - B Codage par transformé en cosinus discrète.....	55
IV - C Codage Différentiel	58
Chapitre V	61
METHODES HYBRIDES DE CODAGE.....	61
V - A La quantification vectorielle appliquée dans le domaine	
transformé.....	61
V - B Codage hybride; adaptation visuelle.....	63
V - B - 1 Quantification vectorielle dans l'espace	
transformé.....	63
V - B - 2 Quantification vectorielle dans l'espace image.....	68
Classification visuelle	68
Résumé	70
Codage par transformée en cosinus discrète	71
Codage par quantification vectorielle.....	71
Chapitre VI.....	76
CORRESPONDANCE IMAGE - ESPACE TRANSFORME	76
VI - A Introduction.....	76

VI - B Correspondance pour les transformées mathématiques.....	77
VI - B Espace Transformé en Cosinus Discrète	77
VI - B - 1 Blocs homogènes	78
VI - B - 2 Transitions primaires	79
VI - B - 3 Manipulations du bloc image	80
Transformation qui donne à un pixel la valeur complémentaire à 255 de sa luminance.....	80
Symétrie par rapport à la médiane verticale	80
Symétries par rapport à la médiane horizontale.....	81
Symétrie par rapport à la première diagonale du bloc image.....	81
Symétrie par rapport à la seconde diagonale	82
Translation des luminances d'un bloc image	82
Chapitre VII	84
CODAGE PAR QUANTIFICATION VECTORIELLE DANS L'ESPACE TRANSFORME AVEC CLASSIFICATION VISUELLE.....	84
VII - A Classification des blocs image.....	84
VII - A - 1 Paramètres de classification.....	85
Activité spatiale totale	85
Activité partielle matérialisant les transitions verticales	86
Activité partielle correspondant aux transitions horizontales	86
Bloc présentant une symétrie par rapport à la première diagonale	87
Bloc image symétrique par rapport à la deuxième diagonale.....	87
Bloc image ayant une symétrie par rapport à son centre.....	88
VII - B Arbre de décision	88
VII - B - 1 Classes d'activité spatiale.....	88
VII - A - 2 Classes de composantes visuelles particulières.....	90
VII - A - 3 Formation de vecteurs de coefficients transformés	93
VII - C Quantification vectorielle des coefficients transformés	95
VII - C - 1 Distorsion entre deux vecteurs	96

VII - C - 2 Création du dictionnaire	101
VII - D Description du codage temporel	101
VII - D - 1 Estimation de mouvements	102
VII - D - 2 Codage de la composante continue par prédiction temporelle	103
VII - E Tests d'évaluation	103
VII - F Décodage rapide	104
Chapitre VIII	108
REALISATION MATERIELLE	108
VIII - A Modules de traitements parallèles	108
VIII - B Acquisition	111
VIII - C Le calculateur TCD	112
VIII - D Le codeur	114
VIII - D - 1 Description du schéma synoptique du système de transmission d'images animées	116
VIII - D - 2 Modules de codage	117
VIII - E Le décodeur	118
VIII - F Architecture du système de codage	119
Chapitre IX	120
PRESENTATION DES PERFORMANCES	120
IX - A Classification visuelle des blocs	121
IX - A - 1 Activité spatiale	121
IX - A - 2 Classification suivant les formes de transitions	122
IX - A - 3 Quantification vectorielle	125
IX - B Taux de compression	126
IX - C Comparaisons avec les techniques connues	127
IX - D Extensions et perspectives pour le procédé de codage	133
CONCLUSION	135
Liste des références	138
ANNEXES	142

Chapitre I

INTRODUCTION

L'image est devenue un support incontournable de communication. La richesse des informations qu'elle rassemble est sans comparaison; formes, couleurs et signes en déplacement. L'impact visuel que provoque l'image chez l'observateur n'a d'égale que la complexité de la symbolique qu'elle véhicule. En effet, les messages visuels sont les plus persuasifs et les plus riches. C'est une des principales raisons pour lesquelles nous constatons la multiplication d'équipements de communications visuelles. De nombreux services basés sur la transmission d'images ont vu le jour : la vidéoconférence, le visiophone, le photo-vidéotex, l'audio-vidéographie, le multimédia etc..

Ces nouveaux modes de communication nécessitent la création, la manipulation, l'archivage et l'échange d'images. Ils nécessitent aussi l'exploitation des voies numériques de transmissions nationales et internationales. L'extension du marché potentiel de ces services passe par la minimisation de leur prix de revient. Il s'agit de baisser le coût des terminaux, ainsi que le temps d'occupation des canaux de transmission.

I - A Contraintes

Les systèmes de communication basés sur l'image sont soumis à de sévères contraintes découlant de la nature même des images et des scènes qu'elles représentent.

L'image représentée numériquement est composée d'une matrice de pixels. Ces éléments picturaux sont représentés par les valeurs de leurs

luminances les différentes structures ou formes contenues dans l'image. Le volume des informations constitué par une image noir et blanc fixe de (512 x 512) est de 256 Koctets, il requière plus de 32 secondes de temps de transmission sur le Réseau Numérique à Intégration de Services (le RNIS possède deux canaux à 64 Kbits/s de débit). Pour les séquences animées la quantité d'informations est multipliée par le nombre imposé d'images par seconde.

Tout système de transmission d'images doit donc supporter une cadence soutenue, acquérir, traiter, transmettre et visualiser des séquences d'images numérisées dans un temps raisonnable. La notion de vitesse, ici, est liée à la capacité de l'équipement à traiter le flot de données imposé par le service étudié, pour la télévision 25 images par seconde sont nécessaires, mais pour le visiophone quelques images par seconde suffisent.

I - B Solutions retenues

La quantité d'information contenue dans une image est telle qu'elle pose de très sérieux problèmes. Les conséquences sont ressenties sur le temps de traitement et de transmission de l'image et sur la taille mémoire nécessaire pour son stockage.

Afin d'extraire les informations pertinentes et de réduire la quantité globale des données, nous appliquons à l'image une compression ou un codage. Depuis que la numérisation est technologiquement possible, plusieurs techniques de codage existent. Elles exploitent la représentation matricielle des pixels de l'image, la redondance et la corrélation entre les pixels voisins. Mais ces méthodes restent étroitement liées aux applications auxquelles elles sont destinées.

Pour les systèmes de transmission d'images, la contrainte essentielle est leur rapidité de traitement en temps réel et leur capacité à supporter une cadence soutenue d'informations. Ce qui impose souvent une limitation au niveau de la complexité de l'algorithme de compression utilisé en vue d'une implémentation matérielle.

Trois aspects comparatifs sont exploités pour les codeurs d'images :

- Coût des équipements de codage - décodage.
- Gain en débit.
- Qualité subjective des image restituées à la réception.

A la réception la restitution de l'image initiale est faite. Cette reconstruction doit être d'une grande fidélité et doit présenter un minimum de dégradations perceptibles par l'oeil de l'observateur.

Le taux de défauts tolérés dans les images reconstituées est indissociable de l'application étudiée. Une déformation d'un contour sur une radiographie archivée est grave de conséquences, donc intolérable. Cependant, la disparition des pois sur la cravate d'un interlocuteur par visiophone est tout à fait acceptable.

Les techniques de compression d'images répondent chacune à des contraintes particulières imposées par l'application étudiée. Le choix de la méthode de compression, la mieux adaptée nécessite donc une connaissance préalable du type d'images transmises.

Nous avons effectué une investigation pour identifier les techniques optimales de codage. La démarche scientifique prise nous a permis de faire le tour des méthodes connues et de déterminer leurs particularités propres, de façon à les exploiter de la meilleure manière possible.

Nous avons étudié en première partie les trois techniques classiques de codage :

- Codage prédictif différentiel,
- Codage par transformations mathématiques,
- Codage par quantification vectorielle.

A la suite de cette étude nous nous sommes intéressé aux méthodes adaptatives. Plusieurs approches algorithmiques concernant des adaptations particulières des techniques classiques sont citées dans la littérature. Pour décrire l'état de l'art en matière d'adaptabilité, nous avons choisi de décrire ces différentes formes en détaillant les techniques d'adaptation les plus originales.

Tenant compte des recommandations des spécialistes en ce qui concerne les extensions algorithmiques, il nous a paru avantageux d'approfondir notre connaissance des adaptations visuelles. Celles ci sont fondées sur l'extraction des structures perceptibles contenues dans l'image. Le codage visuel identifie les informations pertinentes en localisant les composantes de l'image qui agissent sur la qualité subjective après reconstitution des images décodées.

A la suite de cette recherche permettant une meilleure connaissance des algorithmes de compression d'images, nous avons imaginé un nouveau procédé de codage.

Notre méthode est une adaptation visuelle d'une technique de codage hybride exploitant une transformation mathématique et la quantification vectorielle. L'objectif initialement fixé était la proposition d'une méthode de codage qui offrirait des performances supérieures à celles offertes par les méthodes connues.

Notre proposition présente une approche originale permettant d'optimiser les différentes étapes du codage - décodage. Elle est fondée sur l'exploitation de la signification visuelle des composantes de l'image.

Il s'agit d'extraire le maximum d'informations visuelles contenues dans l'image. Les structures identifiées sont modélisées et leurs impacts subjectifs sur l'observateur déterminent les seuils de visibilité. Cette étude préliminaire a abouti à un série de constatations jamais identifiées ni utilisées dans le codage d'images. L'exploitation de ces caractéristiques visuelles a amélioré de beaucoup le taux de compression ainsi que la qualité subjective des images décodées.

Le procédé que nous avons conçu présente plusieurs caractéristiques avantageuses, la principale étant la simplicité des traitements nécessaires pour le codage - décodage. Elle agit directement sur la complexité des équipements.

Pour réaliser un système électronique exécutant notre méthode de codage, nous avons opté pour un équipement modulaire, évolutif et d'une grande puissance de calcul. Notre choix s'est fixé sur un ensemble de processeurs de traitements de signal, doublés de processeurs parallèles de type : transputer. Cette configuration ouverte nous a permis d'améliorer les performances en temps d'exécution de notre technique de codage.

Ce mémoire présente le même cheminement que celui de nos travaux de recherche. Nous y détaillons dans le chapitre II l'essentiel des techniques de codage classiques. Dans le chapitre III nous avons exposé des exemples de procédés adaptatifs de codage d'images. Les méthodes utilisant des adaptations visuelles sont rassemblées dans le chapitre IV. Dans le chapitre V nous avons décrit quelques procédés hybrides. Ces techniques exploitent les avantages de plusieurs méthodes classiques pour améliorer les performances de codage.

Pour mettre en valeur les constatations sur les correspondances entre l'espace Image et l'espace Transformé, nous les avons exposées dans le chapitre VI.

Un développement de notre procédé de codage d'image, fondé sur l'utilisation de la signification visuelle des composantes de l'image, est effectué dans le chapitre VII. Le système de codage - décodage que nous avons développé est décrit dans le chapitre VIII. Cet équipement a été conçu pour pouvoir tester plusieurs types de codage et pour exécuter la nouvelle méthode de compression que nous avons développée.

Chapitre II

TECHNIQUES CLASSIQUES DE CODAGE D'IMAGES

La bonne utilisation des différentes techniques de codage passe par une maîtrise de leurs particularités. Il nous a semblé important de traiter les méthodes classiques de compression d'images, de manière à identifier leurs avantages respectifs et leurs domaines d'application.

II - A Méthodes prédictives

Une grande partie de l'information véhiculée par l'image est dite redondante. Son absence dans l'image n'altère cette dernière que d'une façon superficielle, sans pour autant porter atteinte à son contenu informatif. La nature de l'image noir et blanc numérisée; matrice composée d'éléments picturaux ou pixels, chacun d'eux étant codé par sa luminance, et le contenu de l'image composé de plages uniformes ou de formes font que les variations entre les valeurs de luminance des pixels voisins sont faibles.

La nature matricielle de la représentation de l'image et la dépendance existant entre chacun des pixels et ses voisins traduisent une corrélation très forte sur l'image. Cette corrélation fait qu'on ne peut traiter un pixel sans se soucier des valeurs de luminance de ses voisins.

Le principe de base des techniques de codage prédictives consiste à exploiter la redondance existant entre les valeurs du signal à transmettre. Il

repose sur la constatation de la faible valeur de la différence entre les luminances de deux pixels voisins ou se succédant sur une séquence d'images.

Une prédiction du signal à coder est réalisée au niveau de l'émetteur à chaque instant. Afin d'exploiter la corrélation existante, l'écart entre la valeur réelle de la luminance et la valeur de prédiction est quantifié et transmis au récepteur sous forme de code numérique (cf. Fig. N°1).

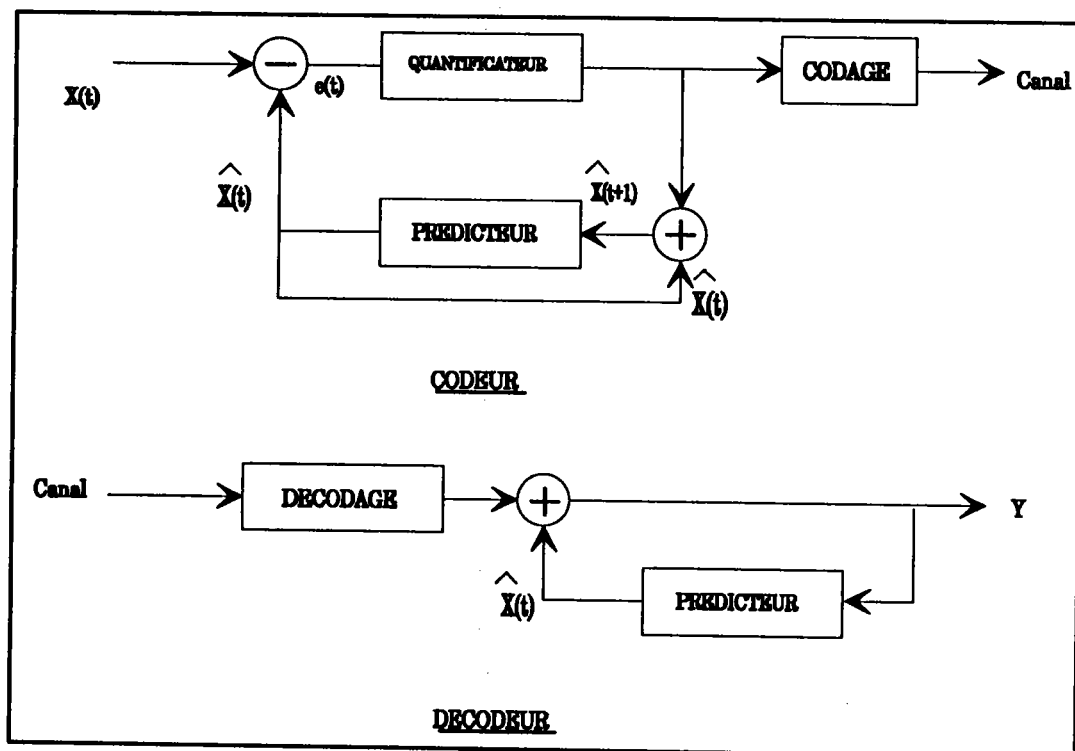


Figure N°1 Codage par prédiction différentielle

Dans le récepteur, en rajoutant la valeur de l'erreur décodée à une valeur de prédiction calculée de la même manière qu'à l'émission, c'est à dire à partir des échantillons du signal correspondant aux valeurs passées et déjà reçues au récepteur, on peut reconstituer le signal image, aux erreurs près de quantification et de transmission (cf. Fig. N°1).

Les systèmes de codage basés sur la MICD : Modulation par Impulsions Codées Différentielles, ou en Anglais DPCM : Differential Pulse Coding Modulation, sont donc essentiellement constitués par un prédicteur du signal

et un quantificateur de l'erreur de prédiction.

Nous allons maintenant détailler les différents éléments de cette chaîne de codage. En mettant en valeur les différents problèmes que pose chacun d'eux.

II - A - 1 Le prédicteur des échantillons du signal image

L'estimateur effectue une prédiction de chaque élément à transmettre (point, bloc, image...). Il n'utilise que des informations déjà obtenues et transmises, ce qui permet au décodeur de recalculer et de reconstituer le signal image initial [GUIC 86]. Le prédicteur optimal est celui qui calcule la valeur de prédiction qui s'approche le mieux de la valeur réelle du signal à coder.

La plupart des systèmes de codage d'images différentiels utilisent la valeur, plus ou moins pondérée, du point précédent de la même ligne comme valeur simple de prédiction. D'autres prédicteurs donnent une valeur en fonction des luminances des pixels voisins du point courant.

Nous pouvons répertorier les prédicteurs dans deux classes. Ceux qui exploitent les valeurs de luminance des voisins directs du pixel à coder sur l'image et ceux qui utilisent les valeurs de luminance des pixels de l'image précédente. Les deux catégories exploitent respectivement la redondance spatiale et la redondance temporelle du signal image numérisé.

II - A - 2 Le quantificateur de l'erreur de prédiction

Le gain en débit du MICD résulte principalement de la quantification du signal d'erreur de prédiction. La précision nécessaire pour le codage du signal d'erreur transmis est variable. La visibilité du bruit de quantification varie avec le contexte local : dans les zones de contraste élevé, l'erreur de prédiction est importante puisque la différence entre la valeur réelle du signal et la valeur estimée est importante; au contraire les erreurs de prédiction faibles se produisent dans les zones à faibles évolutions de

luminance [KRET 77].

Le quantificateur est l'organe du système dans lequel est réalisé la compression. En effet, il réduit la dynamique totale du signal d'écart de prédiction en zones délimitées par des seuils de décision. Le quantificateur donne à chaque intervalle de valeurs d'erreur de prédiction, présenté à son entrée, une valeur de quantification. Cette valeur sera codée et transmise au décodeur. Pour chaque quantificateur, une caractéristique détermine les zones, les seuils et les valeurs de quantification (cf. Fig. N°2).

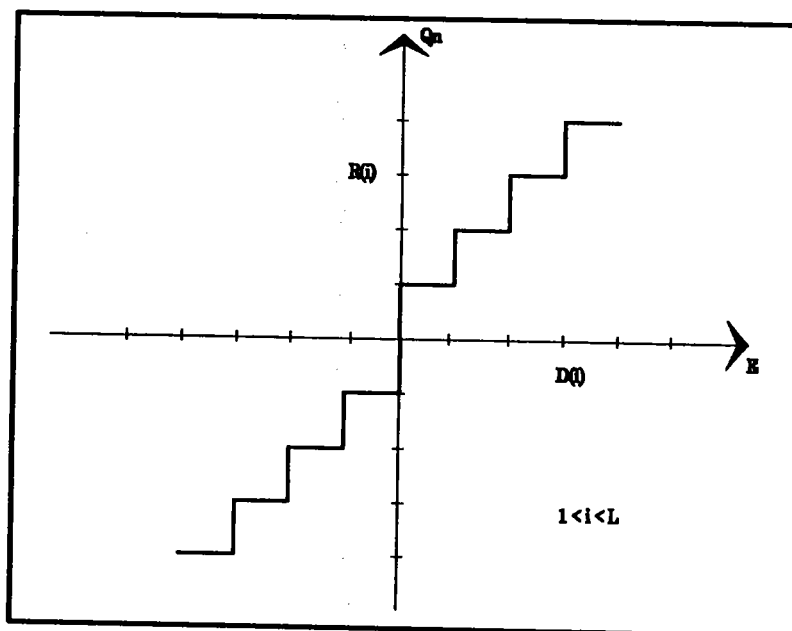


Figure N°2 Caractéristique d'un quantificateur 2L niveaux.
D domaine de décision, R niveau de quantification

La conception du prédicteur fait intervenir uniquement des propriétés statistiques du signal source. Celle du quantificateur fait appel aux propriétés de l'observateur, ce qui explique l'absence de méthode rigoureuse pour définir le quantificateur optimum.

Après la quantification, le codeur alloue un certain nombre de bits pour chaque valeur quantifiée de l'erreur de prédiction. Ce nombre est étroitement lié d'une part à l'efficacité de l'estimateur et du quantificateur et à leur capacité à minimiser les erreurs, d'autre part à la nature du signal image et aux évolutions temporelles et spatiales que présentent ses échantillons.

La compression se traduit par le codage sur un nombre limité de bits des erreurs de prédiction. Avec un estimateur optimal l'erreur reste faible.

II - A - 3 Défauts générés par la MICD

En l'absence d'erreurs de transmission, les systèmes basés sur la MICD sont susceptibles de générer un certain nombre de défauts dans l'image restituée au décodage. Du point de vue de l'observateur, ces dégradations se manifestent sous des apparences diverses. Elles ont cependant une origine commune : les erreurs de quantification.

Nous nous proposons de faire une liste non exhaustive de ces bruits de codage et de leurs causes, dans le but de mieux cerner les particularités des systèmes de codage MICD.

Le bruit granulaire

Il est visible principalement dans les zones uniformes de l'image. Lorsque l'évolution de luminance des pixels n'est pas significative, les valeurs successives de quantification de l'erreur de prédiction présentent une alternance dans leur signe. Ainsi, les valeurs de quantification oscillent autour du signal image. Sur l'image reconstituée apparaissent, dans des zones qui sont initialement uniformes, des groupes de pixels dont les luminances erronées laissent penser que nous sommes en présence de fortes variations du signal image.

Le bruit de faux contour

Ce bruit peut apparaître sur des plages de l'image où la luminosité varie de manière régulière et lente. Il est avant tout l'effet de l'erreur provoquée par la quantification, lors du réajustement de la valeur quantifiée par un passage à une valeur plus importante. Du point de vue de l'observateur, apparaît une transition de luminances inexistante dans l'image initiale.

Le bruit de traînage

Le "slope overload" est dû à une saturation du quantificateur lorsque

le signal subit des variations brutales et successives : les erreurs de prédiction dépassent les valeurs maximales de quantification. Le bruit de traînage est visible pendant le temps d'adaptation que met le quantificateur pour minimiser l'écart entre les valeurs en entrée et celles de quantification [PONC 74].

Le flottement de contour

Celui-ci se traduit par un sautillerment ou mouvement des contours. Il est surtout gênant quand les transitions sont marquées. Les causes de ce bruit sont celles du bruit granulaire combinées avec celles du faux contour.

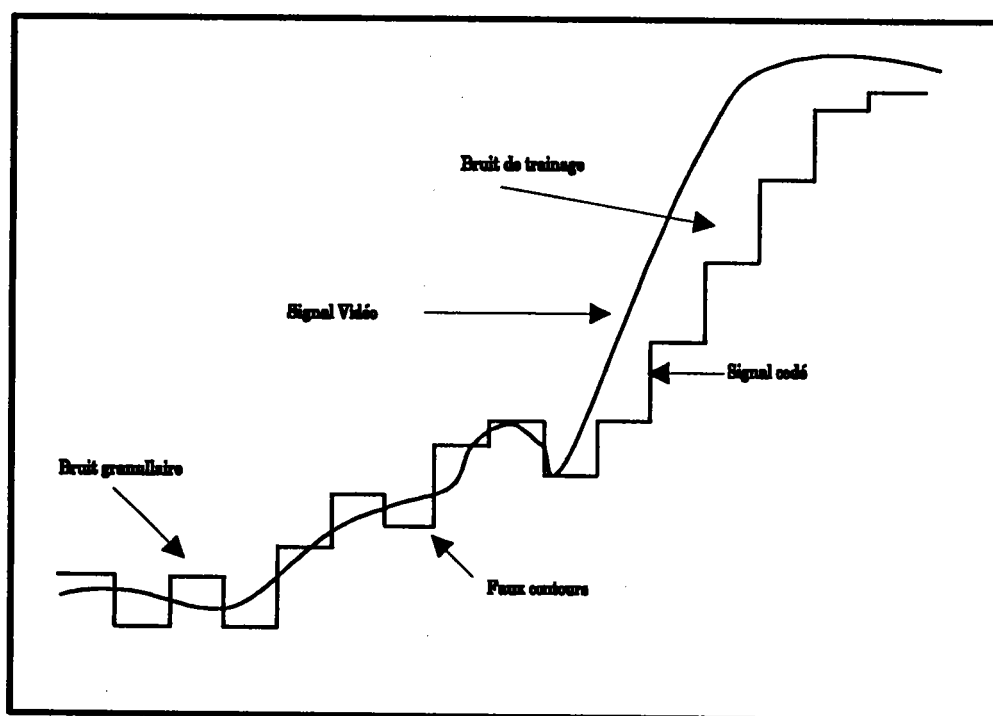


Figure N°3 Bruits de quantification identifiés dans le codage MICD

Rappelons que les techniques classiques de Modulation par Impulsions Codées Différentielles sont efficaces pour un signal image très redondant. En effet, les signaux dont les évolutions temporelles et/ou spatiales sont petites, permettent d'obtenir des prédicteurs optimaux et des erreurs d'estimation minimales. Pour cette catégorie d'images, la technique MICD permet d'obtenir une bonne qualité subjective et un bon taux de compression. Mais, pour des images contrastées et d'une grande activité temporelle et/ou spatiale, d'autres systèmes de codage sont mieux adaptés.

II - B Codage par transformations mathématiques

La représentation matricielle de l'image numérisée est source d'une importante redondance des valeurs des luminances des pixels. Elle est aussi génératrice d'une fonction d'interdépendance entre ces même valeurs. Cette représentation traduit la redondance de l'image par des zones plus ou moins étendues où la valeur de la luminance varie peu ou reste uniforme.

Plusieurs techniques de codage contournent cette représentation, en appliquant une transformation sur le plan de l'image. Cette opération est effectuée dans le but d'obtenir une meilleure décorrélation de représentants de l'image et une minimisation de la redondance (cf. Fig. N°4).

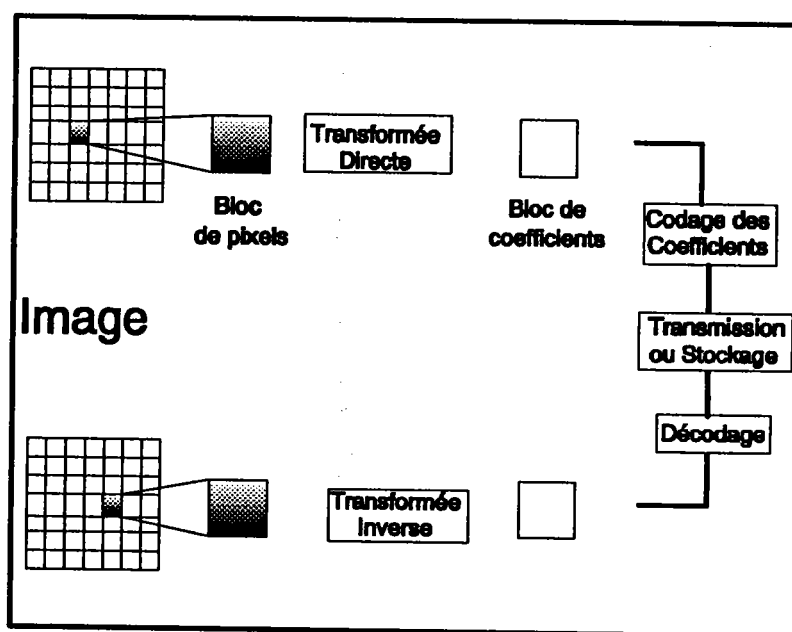


Figure N°4 Codage par transformation mathématique

Afin d'obtenir une meilleure réduction de l'information, plusieurs transformations mathématiques sont utilisées. Certaines étaient connues et répertoriées bien avant que n'existe l'image numérique. Elles ont, pour la plupart, fait leurs preuves en codage du signal mono-dimensionnel [RICH 86].

II - B - 1 Formulations mathématiques

L'image initiale est découpée en médaillons habituellement carrés de

dimension (N×N) (cf. Fig. N°4). La luminance de chaque pixel est une fonction de deux dimensions suivant ses coordonnées : f (x,y), [x et y sont comprises entre 0 et N-1].

La transformée mathématique directe et discrète est définie par l'expression linéaire:

$$F(u,v) = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x,y) a(x,y,u,v) . \quad (1 - a)$$

et la transformation inverse :

$$f^*(x,y) = \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} F(u,v) b(x,y,u,v) . \quad (1 - b)$$

Les fonctions a et b sont prises séparables suivant (x,u) et (y,v).

Il est possible d'avoir une approche matricielle des transformations; les fonctions déterminées précédemment sont matérialisées par des matrices (N×N) de dimensions deux.

La relation (1 - a) s'écrit alors:

$$[F] = [A] [f] [A]^t. \quad (2 - a)$$

$$[f^*] = [B] [F] [B]^t. \quad (2 - b)$$

si l'on prend $[B] = [A]^{-1}$ transformation inverse.

Dans le cas où la transformation [A] est orthogonale et symétrique nous avons :

$$[A]^{-1} = [A]^t = [A].$$

si l'on prend $[B] = [A]^{-1}$, l'expression (2 - b) devient

$$[f^*] = [A] [F] [A] = [f]. \quad (3)$$

On remarquera que les mêmes opérations sont appliquées pour le

calcul de la transformation directe et inverse. Dans la mesure où les calculs sont faits avec la précision nécessaire et où aucune discrimination de coefficients transformés n'est effectuée, la double transformation directe - inverse n'introduit aucune dégradation sur l'image restituée [PONC 74].

Les principales transformations qui ont été utilisées dans le domaine de codage d'images sont les transformées de Fourier, de Haar, de Hotelling ou Karhunen-Loève, Walsh, Hadamard et pour finir la transformée en Cosinus Discrète (TCD).

Les performances offertes par la TCD ont fait que les autres transformées ont été écartées pour le codage d'images.

Une autre transformation qui a un avenir prometteur - dans le domaine des traitements d'images en général et la compression en particulier - est la transformée en ondelettes [MATH 90].

II - B - 2 Transformée en cosinus discrète

Définie en 1974 par Ahmed et al [AHME 74], la transformée en cosinus discrète suscite actuellement un intérêt remarquable.

Les propriétés mathématiques que présente la TCD : orthogonalité, parité et la simplicité des calculs qu'elle nécessite, lui confère une robustesse et des performances qui manquent aux autres transformées mathématiques.

L'essor technologique qu'a déclenché la TCD traduit la compétition ouverte que se livrent les industriels pour conquérir le grand marché de la communication par l'image.

Définition

A partir d'un bloc de $(N \times N)$ pixels corrélés, par une transformée en cosinus discrète nous obtenons un bloc de $(N \times N)$ coefficients transformés. La transformée s'exprime par la définition suivante:

Pour u et v de 0 à N-1 :

$$F(u,v) = \frac{4.c(u).c(v)}{N^2} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x,y) \cos \frac{(2x+1)up}{2N} \cdot \cos \frac{(2y+1)vp}{2N} \quad .(1)$$

$$c(u) = 1/\sqrt{2} \text{ si } u = 0$$

$$c(u) = 1 \text{ si } u = 1, 2, \dots, N-1.$$

$f(x,y)$ représente la luminance du pixel au point (x,y) dans l'image.

$F(u,v)$ représente le coefficient transformé au point (u,v) .

avec u et v allant de 0 à N-1.

La transformée inverse est définie par :

$$f(x,y) = \frac{4.c(u).c(v)}{N^2} \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} F(u,v) \cos \frac{(2x+1)up}{2N} \cdot \cos \frac{(2y+1)vp}{2N} \quad .(2)$$

avec x et y allant de 0 à N-1.

Evaluation de la dynamique des composantes de la matrice transformée

La distribution moyenne de l'énergie dans le domaine transformé (TCD) est concentrée sur un ensemble d'éléments transformés réduits. En effet, la répartition statistique de l'amplitude des fréquences spatiales $F(u,v)$ dépend fortement du rang (u,v) . Elle décroît fortement quand (u,v) s'éloigne de $(0,0)$.

Mais le passage dans le domaine transformé ne provoque pas à lui seul une réduction de l'information de l'image. Si le nombre des fréquences représentatives est restreint, leur dynamique, elle, est généralement supérieure à celle des luminances des pixels.

Ainsi, si la dynamique des luminances des pixels composant l'image est $[0, 255]$. La dynamique de l'élément $F(0,0)$ de la matrice transformée est égale à : $[0, 8 \times 8 \gamma(0,0) \times 255]$. La dynamique des autres éléments de la matrice transformée $F(u,v)$ avec (u,v) différent de $(0,0)$ est égale à $[- \gamma(u,v) \times 8 \times 8 \times 255, \gamma(u,v) \times 8 \times 8 \times 255]$, où $\gamma(u,v)$ est un facteur dépendant du rang de l'élément $F(u,v)$. Par exemple $\gamma(u,v)=0,676$ si: $(u,v)=(6,4)$ [CLAR 85].

Le codage classique par transformée en cosinus discrète est essentiellement basé, pour la réduction des informations de l'image, sur la quantification scalaire des fréquences transformées. Cette opération n'est efficace que si le quantificateur élaboré est optimal. En effet, si l'allocation du nombre de bits nécessaires pour coder un coefficient n'est pas adapté à sa dynamique, alors la réduction d'informations ne sera pas maximale.

Les coefficients transformés sont, en premier lieu, comparés à un seuil prédéterminé en fonction de la qualité désirée. Cela permet d'éliminer les valeurs non nécessaires à la restitution. Les coefficients restants sont quantifiés individuellement en parcourant le bloc transformé du haut à gauche vers le bas à droite, suivant un zigzag. Les fréquences sont codées successivement des plus basses aux plus hautes.

Sachant que la densité moyenne de l'énergie est concentrée autour de la composante continue, le quantificateur se limite aux zones de coefficients transformés représentatifs.

Le codeur quantifie chaque valeur non nulle de coefficient TCD. De plus, il code chaque plage de valeurs nulles par un mot code de longueur variable donné par une table de codage entropique de Huffman. Le nombre de bits alloués à la représentation de chaque bloc est non seulement fonction du nombre de coefficients non nuls, mais aussi de la fréquence d'apparition de chaque coefficient et de l'étendue de la plage des zéros dans la matrice transformée.

Chaque bloc transformé est alors codé par une suite de valeurs matérialisant les codes des coefficients transformés en cosinus discrète, ainsi que les positions des plages de zéros. Le quantificateur des coefficients utilise souvent un mot code de taille fixe. La longueur du mot code total affecté à chaque matrice transformée est variable. Elle dépend étroitement de la distribution spectrale obtenue à la suite de la transformation appliquée au bloc image initial, donc des variations de luminance des pixels qui le composent.

Rappelons la correspondance établie entre les évolutions des niveaux de gris des pixels et la distribution de leurs coefficients transformés. Pour un bloc d'image contenant des variations de luminances limitées, les coefficients transformés les plus représentatifs sont concentrés autour de la composante continue; $F(0,0)$ matérialisant la moyenne des luminances des pixels. Pour les images très fortement contrastées, celles dont les évolutions des

luminances sont brutales et nombreuses, le spectre énergétique et la dynamique des coefficients transformée sont étendus. En effet, de nombreux coefficients situés dans les hautes fréquences sont pertinents et représentent, avec la composante continue, l'essentiel de l'information véhiculée par l'image initiale.

Après avoir étudié la dynamique des coefficients, l'importance du spectre énergétique des blocs transformés et leur signification par rapport à l'image initiale et aux évolutions de luminances qu'elle présente, nous nous proposons de donner un exemple simple de codage.

II - B - 3 Codage par seuillage

Une technique proposée dans [KHAL 88], consiste à appliquer un seuillage discriminant. Les coefficients dont la valeur absolue ne dépasse pas le seuil prédéfini sont négligés. La détermination de la valeur limite est un compromis entre la quantité d'information retenue pour représenter l'image et la qualité subjective de cette dernière.

La transformée en Cosinus Discrète est appliquée à des blocs (8x8). Une qualité satisfaisante a été obtenue pour un taux de compression moyen de 8. Elle a nécessité un seuil sélectif de 2, évalué par expérimentations successives.

Pour coder les coefficients dont la valeur absolue est supérieure à 2 sur un octet, une manipulation de la formulation mathématique est effectuée. La dynamique des coefficients se présente :

$F(0,0) : [0,510];$

$F(u,v) : [-254, 254]$ avec u,v variant de 0 à 7. pour u et v différents de zéro.

Une division par deux diminue l'étendue du domaine des valeurs. La transformation inverse réajuste les valeurs reçues par un facteur multiplicatif.

Les coefficients obtenus par la nouvelle transformée sont quantifiés et codés, sur un octet non signé pour $F(0,0)$ et sur un octet signé pour les autres coefficients transformés. La position de chaque coefficient dans la matrice TCD est codée sur un octet. Cette méthode permet de parcourir le domaine transformé à partir du coefficient $F(0,0)$ jusqu'au dernier coefficient pertinent. L'octet nul déterminant l'adresse de la composante continue

marque le début d'un nouveau bloc codé.

Cette technique, expérimentée sur plusieurs images, donne des taux de compression variant de 3 à 8. Le même procédé appliqué à des blocs d'images (16 x16) pixels offre de taux entre 6 et 10.

II - B- 4 Images adaptées au codage par transformation

Pour le codage par transformation TCD, une erreur qui modifie un coefficient se trouve répartie au niveau de tous les pixels du bloc d'image restitué. L'erreur de codage est donc "diluée" spatialement, ce qui est subjectivement favorable. Les déformations sur l'images sont plus acceptables si l'amplitude de l'écart dépasse la limite de la tolérance et si l'erreur est commise sur le coefficient représentant la composante continue, son effet sera visible sur tout le bloc image reconstituée.

Dans d'autres techniques de codage, les erreurs sont localisées et sont plus visibles à la restitution des images.

Un des principaux défauts du codage par la transformée TCD est L'effet de bloc. Il traduit le codage de chaque bloc indépendamment de ses voisins. Il est perceptible dans les plages de faible variation de luminance.

II - C Codage par quantification vectorielle

L'idée de la quantification vectorielle est de tirer parti de l'interdépendance présente entre les échantillons du signal, lors de la quantification.

D'après la théorie débit-distorsion établie par Shannon la quantification vectorielle offre de bonnes performances. En effet, le taux de compression et la qualité subjective des images obtenus par la quantification vectorielle sont meilleurs que ceux réalisés à l'aide de la quantification scalaire. Cette constatation est vérifiée même pour les sources débitant une séquence d'échantillons aléatoires et indépendants.

II - C - 1 Principe

Le codage utilisant la quantification vectorielle est une technique basée sur un découpage de l'image en médaillons. Les éléments obtenus, blocs ou matrices de pixels, sont étiquetés. L'indexation consiste à rechercher

pour chaque bloc de l'image l'élément le plus proche dans un dictionnaire pré-établi dans le codeur. L'étiquetage fait correspondre au bloc, présenté en entrée du codeur, l'adresse du vecteur voisin dans le dictionnaire.

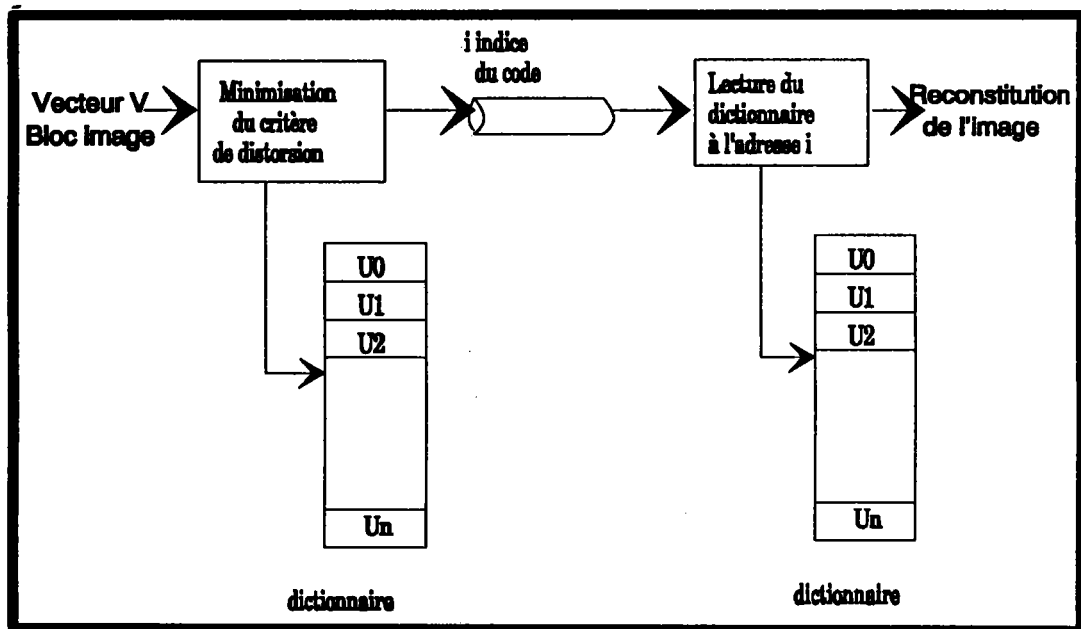


Figure N°5 Codage par quantification vectorielle

X est le bloc de pixel ou vecteur initial à coder, Y_i sont les éléments du dictionnaire; avec $i=1\dots n$, n nombre de vecteurs. Le codage par quantification vectorielle fait correspondre à X l'indice (i) de son plus proche voisin dans le dictionnaire. Cet indice est le mot code qui sera transmis.

$$QV(X) = i \quad \text{si } D(X, Y_i) \leq D(X, Y_j) \\ \text{pour tout } i \text{ différent de } j \text{ avec } j=1 \dots n$$

$D(X, Y)$ étant la distance qui sépare les vecteurs X et Y.

Nous parcourons, donc, tout le dictionnaire à la recherche du bloc de pixels qui pourra le mieux représenter le bloc incident. Une fois ce vecteur trouvé son adresse est transmise au décodeur. Au décodage, l'image restituée sera formée à partir des vecteurs indexés Y_i du dictionnaire (cf. Fig. N°5).

Le dictionnaire est mémorisé dans le codeur et aussi dans décodeur. A la réception du mot code, le décodeur fait une lecture du vecteur correspondant à l'adresse reçue et appartenant au livre de codes ou

dictionnaire. L'image finale est reconstituée par tous les vecteurs dont les composantes vont former les éléments de l'image décodée.

La notion de distance ou distorsion entre deux vecteurs est introduite pour la recherche du plus proche voisin dans le dictionnaire. Plusieurs distances ont été proposées pour optimiser la quantification vectorielle et pour s'approcher au maximum d'un modèle visuel idéal dont la fidélité serait presque sans reproche.

La distance ou distorsion appelée erreur quadratique, est parmi celles qui sont les plus utilisées pour la quantification vectorielle. L'appellation distance ici n'est pas exacte, il s'agit, en fait, du carré d'une distance au sens mathématique du terme.

$$D(U - V) = \sum_{i=0}^k (u_i - v_i)^2$$

Les vecteurs U et V sont de dimension k.

La problématique de la quantification vectorielle a été identifiée par différents auteurs: Gray et al, Goldberg et al, Nasrabadi et al etc. Nous nous proposons ici de la schématiser par la définition de ces trois composantes principales interdépendantes:

- la vectorisation ou la formation de vecteurs à partir des informations à coder,
- la formation du dictionnaire à partir d'un ensemble d'apprentissage,
- la recherche du plus proche voisin à l'aide d'une distance appropriée.

Nous allons détailler ces différentes étapes, en mettant en valeur les principaux problèmes que pose chacune d'elles, ainsi que les solutions proposées par des auteurs, ainsi que les démarches à suivre pour apporter une certaine amélioration ou évolution aux approches classiques.

II - C - 2 Formation des vecteurs

Cette étape rassemble tous les pré-traitements qui, partant de

l'information image à coder, forment des vecteurs. Les composantes de ces derniers doivent être d'une grande valeur informative. La vectorisation idéale est celle qui forme des vecteurs matérialisant exclusivement les données pertinentes du bloc image.

Les pré-traitements incluent le découpage de l'image en médaillons ainsi que l'opération d'extraction et de décorrélation des informations pertinentes à coder. Cette dernière étape est essentielle pour optimiser la quantification car elle agit directement sur la représentativité du dictionnaire ainsi que sur l'éventuelle complexité du codage et par conséquent du décodage.

L'information contenue dans l'image est traitée de façon à faciliter la quantification vectorielle. La décorrélation est basée sur des transformations simples sur les données initiales, pour aboutir à des données plus décorrélées et dont la densité d'informations par élément est plus importante.

Techniques de normalisation pour la quantification spatiale

Baker et al proposent dans [BAKE 82] une méthode qui permet la décorrélation des pixels du bloc image. En premier, la moyenne des luminances des pixels est calculée, ensuite une soustraction est effectuée entre cette moyenne et les luminances des pixels qui composent le bloc image. Le vecteur des écarts obtenus est quantifié. La moyenne, elle, est codée séparément par un quantificateur scalaire.

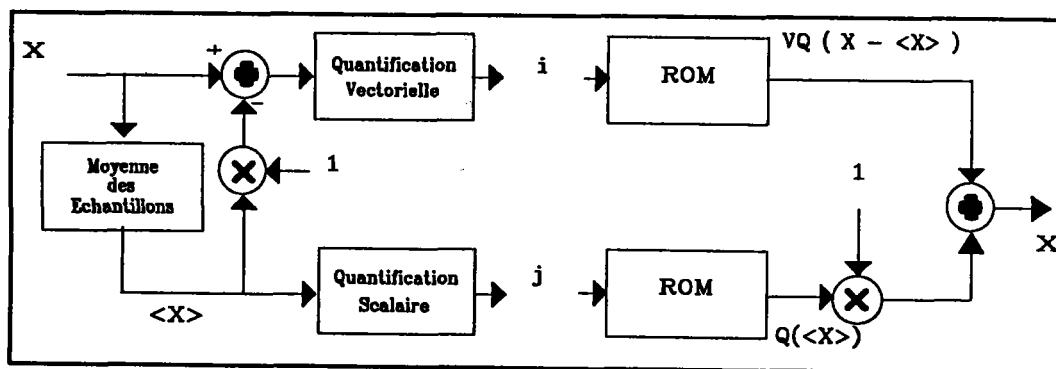


Figure N°6 Quantification vectorielle Moyenne / Gain

Une variante de cette technique consiste à prendre en compte l'erreur résiduelle de la quantification scalaire de la moyenne dans le codage vectoriel. Cela est possible en effectuant la soustraction entre les pixels et la valeur quantifiée de leur moyenne. Ainsi l'erreur de quantification est

introduite lors de la quantification vectorielle.

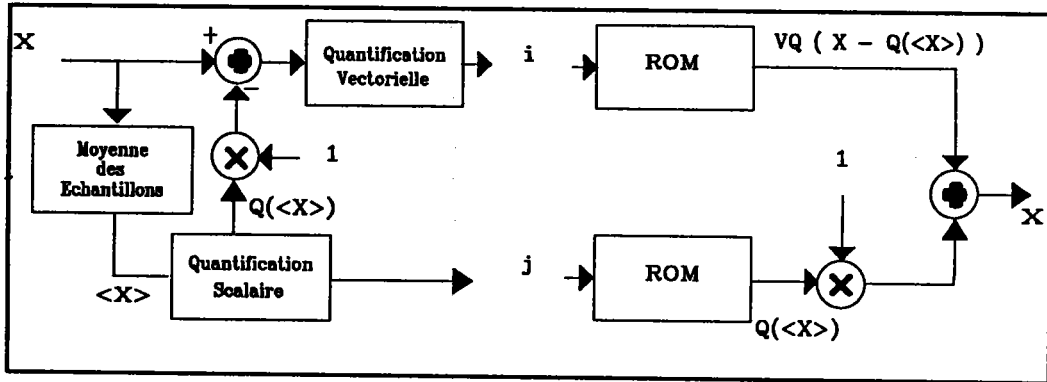


Figure N°7 Quantificateur vectorielle à moyenne résiduelle

Murakami et al proposent dans [MURA 86], une technique de codage utilisant la normalisation avant d'appliquer la quantification vectorielle (cf. Fig. N°6 et N°7). Elle offre ainsi un signal à quantifier peu dépendant de la scène à coder. Parmi les avantages de cette décorrélation l'augmentation de la représentativité du dictionnaire et donc l'optimisation du codage.

Soit un vecteur à coder $V(v_1, v_2, \dots, v_k)$, la moyenne (M) et l'écart-type (e) de chaque composante sont calculées. Le vecteur normalisé $U(u_1, u_2, \dots, u_k)$ est défini par:

$$u_i = \frac{v_i - M}{e}$$

avec

$$M = \sum_{i=0}^k \frac{v_i}{k} \quad \text{et} \quad e = \left[\sum_{i=0}^k \frac{v_i - M}{k} \right]^{1/2}$$

Les vecteurs normalisés sont quantifiés, leurs moyennes et les écarts types sont codés séparément par des quantificateurs scalaires.

Pour chaque vecteur quantifié transmis, la moyenne et l'écart-type le sont aussi, ainsi que la position du vecteur dans l'image. Le dictionnaire est construit de telle façon que la probabilité d'apparition des vecteurs qui le forment soit constante pour toute image en entrée du codeur.

Etant donné la rigidité du quantificateur vectoriel utilisé par Murakami et al. Les auteurs ont amélioré la technique de manière à la rendre adaptative.

Les valeurs des composantes vectorielles et leurs fonctions

d'interdépendances, ont une action déterminante sur l'optimisation de la quantification vectorielle. La taille des vecteurs à coder est aussi très influente sur les performances du codage. En effet, elle intervient directement sur la représentativité du livre de code, elle agit aussi sur le taux de compression et sur la qualité visuelle subjective des images restituées au décodage.

L'utilisation de vecteurs de petite dimension permet de minimiser l'erreur de quantification vectorielle et donc d'améliorer les images décodées, mais son effet se ressent négativement sur le taux de compression.

II - C - 3 Construction du dictionnaire

Partant d'un ensemble de vecteurs très large, le dictionnaire vectoriel optimal est celui qui emploie un groupe de vecteurs présentant la plus petite distorsion moyenne, parmi tous les autres groupes de vecteurs possibles. La construction du dictionnaire est basée sur la recherche des meilleurs représentants des blocs image. A l'aide de ce dictionnaire, un quantificateur vectoriel optimal pourra être élaboré.

Des algorithmes de classification sont utilisés, de façon à élaborer un dictionnaire de vecteurs représentatifs à partir d'un ensemble initial; ce dernier est appelé "training set" ou ensemble d'apprentissage.

L'ultime objectif est de créer un "codebook" ou dictionnaire universel. La principale particularité du livre de codes, est d'être parfaitement adapté à toutes les images incidentes. Ce dictionnaire permettrait de coder tous les types d'images avec un taux de compression acceptable et une bonne qualité subjective. En attendant la découverte de ce dictionnaire idéal, plusieurs auteurs, Gray et al dans [GRAY 84], Linde et al dans [LIND 80], Gersho et al dans [GERS 82] et Kohonen et al dans [KOHO 84], se sont penchés sur le problème. Plusieurs solutions ont été proposées :

Nuées dynamiques

Cette méthode exige la définition d'un espace des représentation E et un paramètre de distorsion $f(x,w)$ qui associe à tout point x de E et à tout sous ensemble w de E une valeur de R^+ :

$$\forall x \text{ de } E, \forall w \text{ de } P(E) \rightarrow f(x,w) \text{ de } R^+$$

f exprime le degré de dissemblance entre la classe w et le point x . Plus cette valeur est faible, plus x est considérée proche de w .

L'algorithme des nuées dynamique suppose connu J , le nombre des classes.

idée de base de l'algorithme

Il s'agit de dire que la meilleure partition $\Pi = \{w_1, w_2, \dots, w_j\}$ de E est celle pour laquelle la somme des distorsion entre les points et les classes auxquelles ils appartiennent est la plus faible. Cela s'écrit ainsi :

A chaque partition, nous associons la quantité :

$$g(\Pi) = \sum_{w \text{ de } \Pi} \sum_{x \text{ de } \Pi} f(x, w)$$

La meilleure partition de l'espace E en J classes est celle pour laquelle $g(\Pi)$ est minimum.

La majorité des algorithmes recherchent des minimum locaux.

L'algorithme suivant est fréquemment utilisé :

- 1) Choisir au hasard une partition Π^0 en classes $\{w_1^0, x_2^0, \dots, x_j^0\}$
- 2) $n = 0$
- 3) $n = n+1$
- 4) Pour chaque valeur de j , calculer

$$w_j^n = \{x \text{ de } E \mid \forall k \text{ différent de } j, f(x, w_j^{n-1}) \leq f(x, w_k^{n-1})\}$$

- 5) retourner en 3) tantque J, w_j^n différent de w_j^{n-1} et que $n \leq n^0$

la condition $n \leq n^0$ fixe un nombre maximum d'itérations a priori et évite que l'algorithme ne boucle indéfiniment. L'idée exprimée en 4) consiste à affecter successivement chaque point x de E à la classe dont il est le plus proche et à définir ainsi peu à peu les classes.

La convergence de l'algorithme n'est démontrable que pour certaines fonctions f .

La définition d'une fonction de distorsion se fait en deux étapes :

La première consiste à choisir une représentation de la classe appelée noyau. Le noyau peut être le centre de gravité de la classe, un ensemble de p points minimisant une fonction de ressemblance [BELA 88] .

Exemple de $f(x, w)$:

Le noyau de chaque classe est son axe principal d'inertie Δw et en définissant la distorsion f par :

$$f(x,w) = d(x, \Delta w) = \inf_y \text{ de } \Delta w \ d(x,y)$$

Algorithme de classification du dictionnaire LBG

Linde et al proposent une technique couramment appelée LBG. Elle nécessite une pré-sélection des vecteurs constituant le "training set", suivant leur nombre d'apparitions dans le codage. Rappelons que la méthode des centroïdes exploite quant à elle le maximum de vecteurs d'apprentissage [LIND 80].

Le LBG effectue itérativement des tests, de façon à former le dictionnaire le plus représentatif à partir de la série d'apprentissage. Pour un critère de distorsion donné, l'algorithme vise itérativement à optimiser la partition de l'espace des vecteurs en fonction du dictionnaire courant et vice-versa. Successivement, les éléments du dictionnaire sont réajustés, l'amélioration de représentativité des partitions de l'espace est vérifiée à chaque itération. le dictionnaire est considéré comme optimal lorsqu'une itération supplémentaire du processus n'amène plus de diminution significative de la distorsion moyenne totale.

La synthèse du dictionnaire initial nécessite une pré-sélection réalisée à la suite d'une étude statistique sur les vecteurs de l'ensemble d'apprentissage ou par la méthode des centroïdes. La création du dictionnaire initial est donc le fruit de plusieurs investigations effectuées sur les vecteurs dont le nombre d'apparitions est relativement plus important que celui des autres. Rappelons que la détermination de ce livre de codes initial est primordiale pour la suite des opérations de construction du dictionnaire.

La méthode des centroïdes, consiste à rassembler le maximum de vecteurs pour former un ensemble d'apprentissage. Les images de références sont souvent les sources qui fournissent les vecteurs. L'ensemble ainsi formé appartient à l'espace de dimension N; N étant la taille des vecteurs.

La détermination du centre de cet espace est entreprise; le résultat est un vecteur qui représente la valeur moyenne des vecteurs du "training set". Un hyperplan, passant par le centre, est ensuite formé de façon à fractionner l'espace en deux classes. La même opération est répétée plusieurs fois, décomposant l'espace en une multitude de nuages de vecteurs; chacun de ces groupes est représenté par son centre (cf. Fig. N°8 et 9).

Cette méthode, permet de répertorier toutes les sous-classes par leurs

centres respectifs. Ces centres seront une bonne représentation du "training set" et formeront ainsi un dictionnaire optimisé [GRAY 84].

Une autre particularité avantageuse de cette technique est qu'elle permet de minimiser le parcours du dictionnaire lors de la quantification vectorielle. L'arbre, formé à partir des barycentres des cellules, offre la possibilité d'être parcouru sur les seules branches voisines aboutissant au vecteur du dictionnaire. Ainsi, de proche en proche, la quantification vectorielle est réussie quand un des centres est suffisamment proche du bloc à coder, c'est à dire quand la distorsion entre les deux vecteurs est inférieure au seuil de tolérance fixé préalablement.

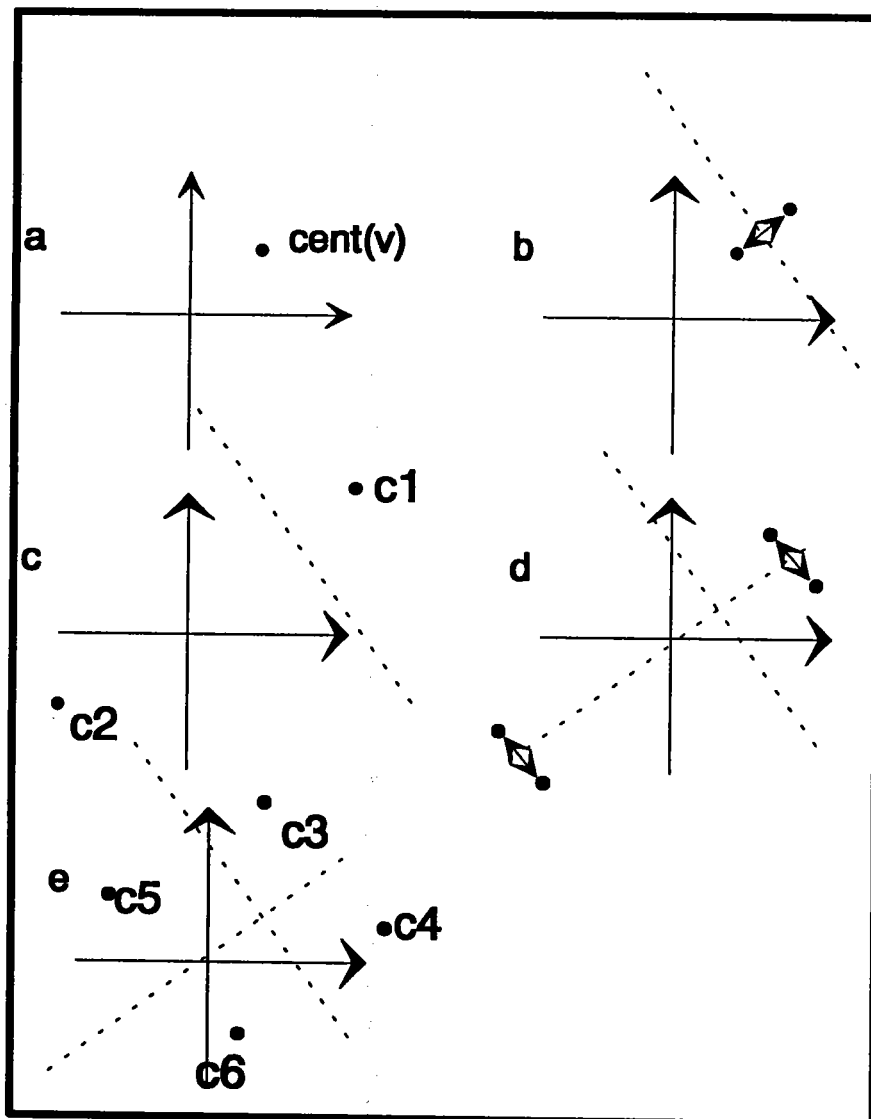


Figure N°8 Classification par la méthode des centroïdes

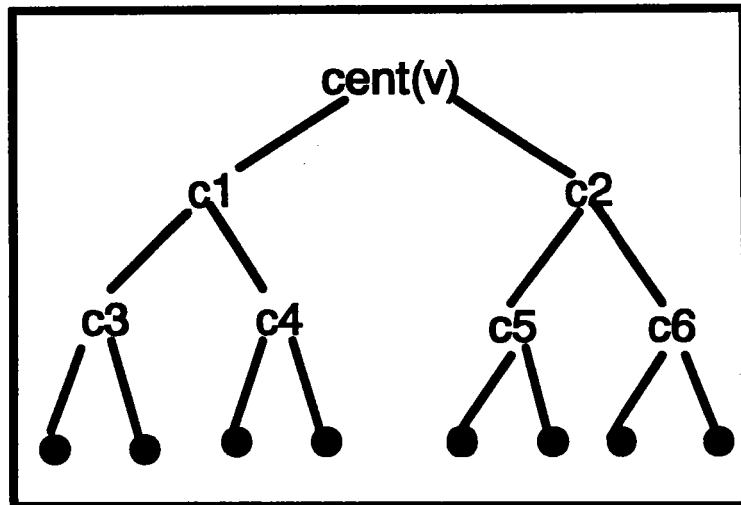


Figure N°9 Disposition des vecteurs dans le dictionnaire

Algorithme de Linde Buzo Gray

Il s'agit donc de trouver un vecteur de paramètres $y_1 \dots y_J$ à l'intérieur d'un ensemble qui minimise une mesure de distorsion.

Le problème posé est le suivant : Comment déterminer un ensemble de références ou codebook. Un dictionnaire est un ensemble fini de vecteurs de référence.

Une méthode itérative est proposée pour créer un livre de codes qui minimise une distorsion sur un grand nombre de données. Elle converge vers un minimum local et non obligatoirement vers un minimum global.

L'objectif est de constituer un ensemble de $2^b = B$ vecteurs représentant un ensemble large de vecteurs rassemblés après analyse des données.

A l'initialisation, B vecteurs sont choisis arbitrairement ou à la suite d'une étude de représentativité. La première étape détermine par la méthode du plus proche voisin les groupes de vecteurs représentés par les B vecteurs initiaux.

La distorsion totale est déterminée en sommant la distorsion de chaque vecteur de la base de données lors de la répartition des classes.

La deuxième étape se propose d'améliorer le livre de codes en choisissant un meilleur représentant pour chaque classe ; exemple le centroïde de la classe évaluée à l'étape 2.

Cet algorithme conduit à une distorsion globale monotone décroissante qui converge donc vers un optimum local.

La vitesse de convergence dépend du choix des B vecteurs initiaux.

Algorithme

1) Initialisation

Fixer un nombre N de représentants du dictionnaire

un seuil $\varepsilon > 0$

Un dictionnaire initial R_0

Une séquence d'entraînement $(X(j); j=1, \dots, J)$

$D_{-1} = \text{infini}$

$m = 0$

Phase N°1

Pour le dictionnaire courant $R_m (y_i; i=1, \dots, N)$

Trouver la partition optimale $\text{Part}(R_m = (C_i; i=1, \dots, N)$

qui minimise la distorsion moyenne sur une séquence d'entraînement

soit

$X(j)$ de C_i

si $d(X(j), Y_i) \leq d(X(j), Y_j)$ pour i différent de j

Calculer la distorsion moyenne pour la répartition obtenue et le dictionnaire R_m

$$D_m = \sum_{j=1}^J \min_i d(X(j), Y_i)$$

Phase N°2

Si $(D_{m-1} - D_m)/D_m \leq \varepsilon$, le dictionnaire optimal est R_m

sinon continuer

Phase N°3

Trouver le dictionnaire optimal R_{m+1} pour une répartition $\text{Part}(R_m)$ pour tout $i, (i=1, \dots, N)$, y = barycentre de C_i

$m=m+1$

exécuter la phase 1

Méthode de Kohonen

C'est une approche algorithmique connexioniste, basée sur les réseaux

de neurones. Cette technique consiste en un apprentissage auto-adaptatif. Le classement des vecteurs de l'ensemble d'apprentissage se fait au fur et à mesure. Les vecteurs sont présentés individuellement en entrée d'un réseau de neurones affectés respectivement par des poids de pondération. Les vecteurs appartenant à une même classe sont de plus en plus regroupés, les autres sont dispersés. La répartition des vecteurs en classes est ainsi effectuée, l'ensemble des vecteurs qui représenteront le "training set" est ainsi constitué.

L'optimisation du dictionnaire dépend principalement du nombre de vecteurs contenus dans l'ensemble d'apprentissage et de la durée nécessaire pour que le réseau de neurones effectue son apprentissage.

Cette technique est celle recueillie des avis positifs, justifiés par les possibilités qu'elle peut offrir pour le rafraîchissement dynamique des dictionnaires. Son efficacité n'égale pas encore celle de ses concurrentes; cependant, elle a beaucoup d'avenir, car elle présente une implémentation relativement simple et elle offre la possibilité d'une extension simultanée du dictionnaire. Cette dernière éventualité sera détaillée dans le chapitre suivant.

II - C - 4 Distance déterminant la distorsion entre les vecteurs

La mesure de distorsion la plus utilisée en codage d'images est l'erreur quadratique, déterminée par l'expression suivante :

$$\text{Deq}(X,Y) = \sum_{i=0}^k (x_i - y_i)^2$$

Deq (X,Y) détermine la distorsion qui sépare les vecteurs X et Y. Ce paramètre ne tient pas compte du système psycho-visuel humain; elle ne peut distinguer entre une erreur de codage visible et une autre imperceptible par l'observateur. Cependant, elle constitue un facteur qui permet de comparer l'éloignement entre les vecteurs, même si les résultats qu'elle donne sont plus ou moins grossiers.

D'autres paramètres de distorsion, qui sont des variantes de celle proposée dans le paragraphe ci-dessus, sont citées dans la littérature spécialisée; erreur quadratique moyenne normalisée, moyenne temporelle de l'erreur quadratique ou encore le critère de distorsion minimale - maximale.

Chapitre III

METHODES ADAPTATIVES DE CODAGE D'IMAGES

Dans le chapitre précédent nous avons décrit les méthodes de codage classiques MICD, TCD et la Quantification Vectorielle. Afin d'étudier les particularités respectives de chaque technique et de mieux cerner la problématique du codage, nous nous proposons de décrire dans ce chapitre quelques exemples de procédés adaptatifs de compression d'images. Le même schéma suivi au chapitre précédent sera respecté ici : méthodes prédictives, codage par transformation et enfin quantification vectorielle.

III - A Méthode prédictives

Les techniques de codage par modulation des impulsions codées différentielles exploitent les évolutions du signal image. Ces variations sont temporelles et/ou spatiales. Les techniques classiques utilisant un estimateur et un quantificateur d'erreur de prédiction fixes sont inefficaces. Elles ont montré leur incapacité à obtenir de bonnes performances de codage et à les maintenir pour différents types d'images.

Pour améliorer les performances en qualité subjective de l'image décodée et en taux de compression, plusieurs techniques font appel à l'adaptabilité. Cette dernière peut s'appliquer aux différentes étapes du codage différentiel : au niveau des prédicteurs, ou au niveau des quantificateurs de l'erreur d'estimation. Le but recherché est d'optimiser la

technique de codage en améliorant ses performances.

L'adaptabilité réside dans le fait que la technique de codage adaptative exploite un prédicteur ou un quantificateur ou les deux à la fois, s'accommodant avec les échantillons du signal à coder. Les caractéristiques des composants du codeur sont variables en fonction des évolutions et des particularités du signal image.

III - A - 1 MICDA-ER

Nous proposons ici de décrire en exemple une méthode originale de compression d'images utilisant plusieurs estimateurs et plusieurs quantificateurs scalaires.

Dans [ZHAO 89], Y. ZHAO a présenté une technique adaptative de codage d'images. Cette méthode évalue la corrélation liant des pixels d'un même bloc et celle les associant avec les pixels du bloc identique dans l'image antécédente, dans la même séquence.

En tenant compte de la fonction d'interdépendance calculée, un estimateur de la luminance est choisi. La quantification scalaire de l'écart entre la valeur réelle de luminance et celle prédite est effectuée. Le mot code transmis comporte une indication pour identifier le prédicteur ainsi que le quantificateur utilisés, le récepteur déchiffre les indications pour décoder la valeur de l'erreur d'estimation quantifiée.

L'adaptation réside dans la possibilité d'avoir le choix entre deux prédicteurs et entre deux quantificateurs. Les meilleurs sont sélectionnés en fonction des paramètres de corrélation associant les pixels se succédant sur une même image ou sur une séquence d'images.

L'estimation de la redondance est effectuée après le calcul de la fonction corrélation C_{ni} , cette dernière est liée aux trois composantes d'un vecteur test :

$$C_{ni} = F(C_{ni1} ; C_{ni2} ; C(n,n-1)_i).$$

Le bloc image (8x8) est divisé en deux sous blocs (4 x 8); les coefficients C_{ni1} et C_{ni2} sont leurs facteurs associés déterminant la corrélation intra-image. Les indices correspondent à l'instant n , à l'image i et à l'indice du sous bloc. Les coefficients permettent d'évaluer la redondance et

l'interdépendance existant entre les pixels du même sous-bloc.

$$\Delta A(j) = A_n(i,j) - A_{n-1}(i,j)$$

$A_n(i,j)$ désigne la luminance du pixel d'ordre j appartenant au bloc i dans l'image. Δ représente l'écart entre deux pixels successifs. L'indice j varie de 0 à 63.

$C(n,n-1)_i$ coefficient de corrélation inter-image; il est estimé par comparaison de tous les pixels d'ordre j du bloc i dans l'image à l'instant n et ceux du bloc correspondant dans l'image à l'instant $n-1$. $C(n,n-1)_i$ prend l'une des valeurs suivantes (0 ; 0,5 ; 1) en fonction des dépassements de seuils comptabilisés effectués par le paramètre $\Delta A(j)$.

Un calculateur commande le choix des prédicteurs et des quantificateurs. Pour l'évaluation de la corrélation temporelle, l'organe de décision utilise un compteur qui comptabilise le nombre de dépassements de seuils prédéterminés, effectués lors du calcul des $\Delta A(j)$.

Des tests comparatifs déterminent la classe d'appartenance du bloc image en fonction du taux de dépassements calculé. La valeur du taux permet de prendre une décision en fonction de l'importance de la corrélation du bloc avec son prédécesseur. Si l'écart constaté entre les deux blocs successifs est faible, alors au décodage, la visualisation du bloc précédent est maintenue. Dans le cas contraire, l'estimation de la corrélation spatiale ou intra-image est effectuée sur chaque demi-bloc :

les facteurs d'interdépendance; C_{n1i} et C_{n2i} sont déterminés à partir du taux de dépassement constaté entre les luminances des pixels de chaque demi-bloc et sa moyenne. Si ce taux est faible le demi-bloc est considéré uniforme, seule sa moyenne est donc transmise au décodeur. Sinon, la comparaison entre les facteurs de dépendance temporelle et spatiale permet le choix entre deux prédicteurs de luminance. Le premier donne pour valeur de prédiction la luminance du pixel à la même position dans l'image précédente, le deuxième propose, lui, la moyenne des luminances de deux pixels voisins.

Le prédicteur choisi est celui qui fait l'estimation la plus proche de la valeur réelle des luminance de chaque sous-bloc d'image.

En ce qui concerne la quantification de l'erreur de prédiction, deux éventualités sont possibles :

- 1 : L'écart entre la valeur réelle de la luminance et la valeur prédite, la plus fidèle des deux possibles, est limité. Dans ce cas, le quantificateur choisi délivre des erreurs de prédiction quantifiées et codées sur deux bits.
- 2 : Une variation importante s'est produite entre les deux images successives, doublée d'une évolution spatiale conséquente des luminances. Dans cette éventualité, l'écart entre les valeurs réelle et les estimation est important, le quantificateur sélectionné alloue pour les erreurs de prédiction un mot code dont la longueur est quatre bits.

La Modulation par Impulsions Codées Différentielle Adaptative avec Estimation de la Redondance (MICDA-ER) exploite plusieurs prédicteurs qui opèrent simultanément. Elle sélectionne l'estimateur optimal, celui qui va présenter l'erreur de prédiction minimale. A la suite de ce choix, un quantificateur scalaire est déterminé pour évaluer et coder l'erreur de prédiction.

La MICDA-ER améliore les performances de la technique prédictive classique, qualité subjective des images décodées et taux de compression de l'image, en exploitant des prédicteurs et des quantificateurs dont les caractéristiques sont variables en fonction de l'évolution des échantillons du signal image à coder [ZHAO 89] (cf. Fig. N°10).

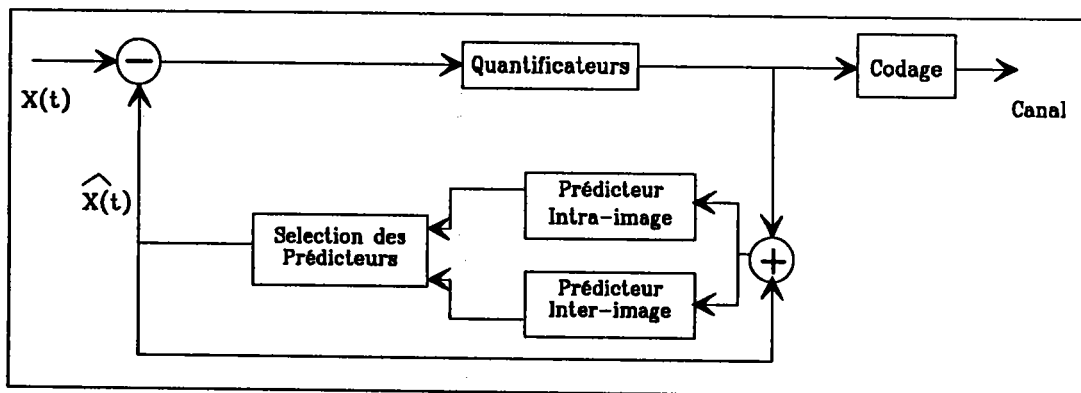


Figure N°10 Codage prédictif par estimation de la redondance

III - B Techniques adaptatives de codage par transformation TCD

L'adaptation pour les techniques de codage basées sur des transformations mathématiques peut avoir deux formes différentes. Elle peut concerner les caractéristiques de la transformation, ou la série de procédures appliquées aux coefficients transformés pour les coder.

III - B - 1 Adaptation par variation de la taille des blocs transformés

Elle est applicable à chaque fois que les images à coder se présentent sous forme de plages ayant les mêmes structures picturales. L'image est fractionnée de façon à former les plus grandes zones possibles de même structure. Ainsi, après la segmentation, des blocs de tailles variables seront formés dans les plages uniformes, ou encore dans des zones contenant des contours ou des textures particulières.

Vaissey et al proposent dans [VAIS 87] une technique de codage par transformation, avec des tailles de blocs variant de 4×4 pixels à 32×32 pixels. Pour les images dont l'évolution des luminances des pixels sont faibles sur de grandes plages, la taille 32×32 pixels leur est réservée, un taux maximum de compression est obtenu. Par contre, pour les zones très contrastées, contenant des détails indispensables à la reconstitution des images il est nécessaire d'avoir un codage adéquat. Des blocs de petite taille sont consacrés aux particularités de l'image pour respecter la fidélité et obtenir la meilleure qualité subjective de l'image décodée.

Nous constatons que la taille des médaillons à transformer est associée au contenu structurel de l'image. Cependant, cette catégorie de techniques adaptatives met en oeuvre plusieurs calculateurs exécutant la transformée ainsi qu'une puissance de traitement importante pour identifier les zones englobant des structures semblables. Sans oublier que pour identifier le type de transformation effectuée au codage, il est nécessaire de transmettre des codes supplémentaires au décodeur. Toutes ces contraintes handicapent énormément ce type de techniques adaptatives de codage. Elles posent un réel problème pour leur implémentation, ce qui compromet fatalement l'éventuelle industrialisation de codeurs - décodeurs d'images conçus à partir de procédés similaires.

III - B - 2 Stratégies adaptatives de codage de coefficients transformés

Nous avons constaté lors de l'étude préliminaire de la transformée en cosinus discrète que, compte-tenu de l'importance de la dynamique des coefficients obtenus, le passage dans le domaine transformé ne constitue pas en lui-même une compression de données. Nous avons aussi remarqué que la disposition des coefficients transformés dans la matrice finale est étroitement liée au contenu de l'image.

Toutes ces caractéristiques ont fait la faillite des méthodes classiques de codage par transformation, en contribuant à leurs mauvaises performances. De nouvelles approches algorithmiques apportent des méthodes d'accommodation aux particularités des contenus informatifs de l'image. D'autres techniques repèrent les zones de la matrice transformée contenant les coefficients pertinents.

Pour mieux cerner les opportunités offertes par les techniques adaptatives, nous présenterons celles qui nous ont semblé les plus originales, les plus robustes et les plus compétitives.

III - B - 3 Classification basée sur des paramètres énergétiques

La répartition statistique des amplitudes des coefficients $F(u,v)$ transformés s'approche d'une distribution Gaussienne; l'énergie moyenne des éléments de la matrice décroît rapidement avec le rang de leur position. De ces constatations, découle la nécessité d'octroyer un nombre variable de bits pour le codage de chaque coefficient $F(u,v)$ en fonction de sa position dans le bloc transformé. Le nombre total d'octets alloués pour chaque bloc transformé est associé à l'activité spatiale des luminances de l'image.

Une méthode de codage TCD adaptative, présentée dans [CHEN 77], exploite des paramètres évaluant la distribution d'énergie dans la matrice transformée. Un vecteur caractéristique est calculé à partir des valeurs des coefficients transformés $F(u,v)$. Une procédure de classification est accomplie suivant une règle de décision exploitant la caractéristique du bloc transformé.

La classification d'après [CHEN 77], s'effectue en calculant le paramètre matérialisant l'énergie totale:

$$\sigma_s = \left(\sum_{u=0}^{u=k} \sum_{v=0}^{v=k} F(u,v)^2 \right) - F(0,0)^2$$

La comparaison de la valeur obtenue avec trois seuils pré-établis permet d'identifier la classe d'appartenance du bloc transformé. Les limites entre les quatre familles sont fixées par apprentissage de façon à obtenir pour chaque bloc des possibilités d'appartenance équiprobables.

Quatre matrices B_k d'allocation de bits sont calculées après plusieurs expérimentations sur un ensemble d'apprentissage. La détermination des éléments des matrices B_k se fait à partir des variances des éléments transformés $F(u,v)$:

Pour chaque classe C_k , on détermine sur l'ensemble d'apprentissage les variances des éléments transformés :

$$a_{ij} : \sigma^2_{a_{ij}(k)} \text{ pour } (i,j) = 0, \dots, N-1 \text{ sauf } (0,0) \text{ et } k = 1, \dots, 4$$

on en déduit alors les matrices B_k d'allocation des bits :

$$[B_k] = [n^*_k(i,j)]$$

avec

$$n_k(i,j) = 1/2 \log_2[\sigma^2_{a_{ij}(k)}] - \log_2(D_k)$$

pour (i,j) différent de $(0,0)$

où n^* est la troncature de n à l'entier le plus proche.

D_k est le nombre total de bits alloué au codage de la classe C_k

$$n_k(0,0) = 8.$$

Le codage adaptatif et le décodage se fait d'une façon générale selon les blocs-diagrammes (cf. Fig. N°12 et 13) .

La quantification des éléments a_{ij} porte en fait sur les éléments normalisés a_{ij}^N . Les facteurs de normalisation sont de la forme de :

$$\lambda_k(i,j) = C \cdot 2^{[n^*_k(i,j)-1]} \text{ avec } (i,j) \text{ différent de } (0,0)$$

Chaque élément normalisé est donc quantifié en utilisant la loi de

quantification normalisée à $2^{[n^*k(i,j)]}$ niveaux si le bloc est affecté à la classe k . Le codage d'un bloc nécessite d'indiquer au décodeur la classe identifiée [BARB 85]. Les quatre matrices suivantes définies par Chen et al sont dans le cas de blocs (16×16) . Le débit moyen obtenu est 1 bit/point.

Les coefficients transformés $F(u,v)$ sont normalisés pour donner $NF(u,v)$, par des facteurs calculés à partir du nombre de bits associé au même rang dans le bloc.

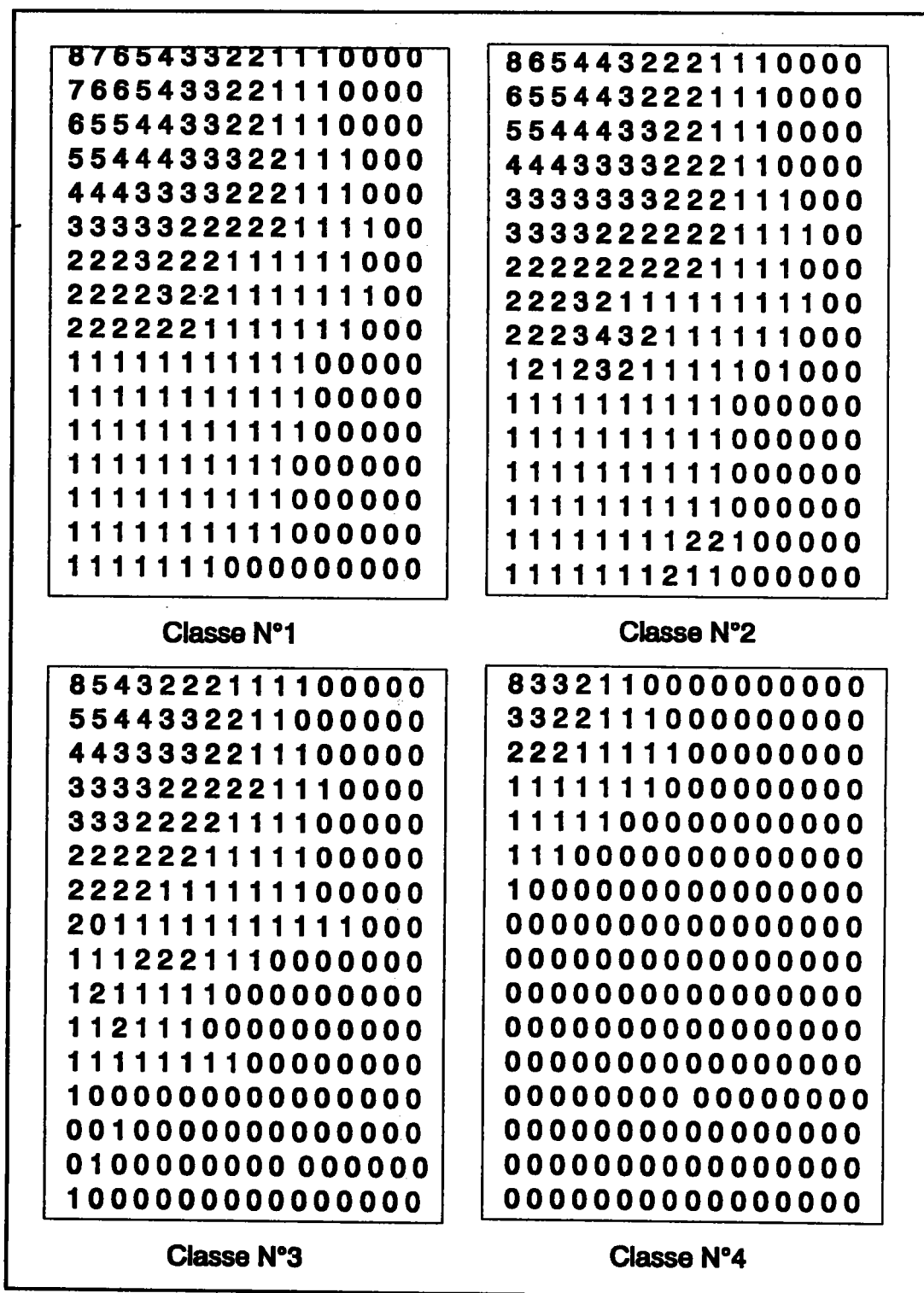


Figure N°11 Classification par matrices d'allocation de bits

A chaque classe d'énergie C_k est associé un quantificateur spécifique des valeurs normalisées $NF(u,v)$, déterminé par la loi de quantification à 2

$nk(u,v)$ niveaux. L'en-tête affectée à chaque bloc identifie sa classe énergétique C_k et par conséquent la matrice d'allocation de bits, les valeurs de normalisation ainsi que le quantificateur approprié (cf. Fig. N°11).

Les figures suivantes décrivent les différentes étapes nécessaires pour effectuer le codage adaptatif par allocation de bits. Certaines procédures sont exécutées d'une façon parallèle, cela réduit considérablement le temps total des traitements et augmente les performances du codeur-décodeur d'images.

La technique proposée par Chen et al, définit la disposition globale des coefficients TCD dans la matrice transformée. Elle ne permet pas de situer la position précise de la zone des coefficients pertinents. Définies grossièrement, les matrices d'allocation de bit ne répondent pas à tous les cas de figures éventuels dans la distribution de l'énergie dans le bloc transformé. Cette technique permet néanmoins d'obtenir un taux de compression de 8 pour une qualité de l'image subjective acceptable.

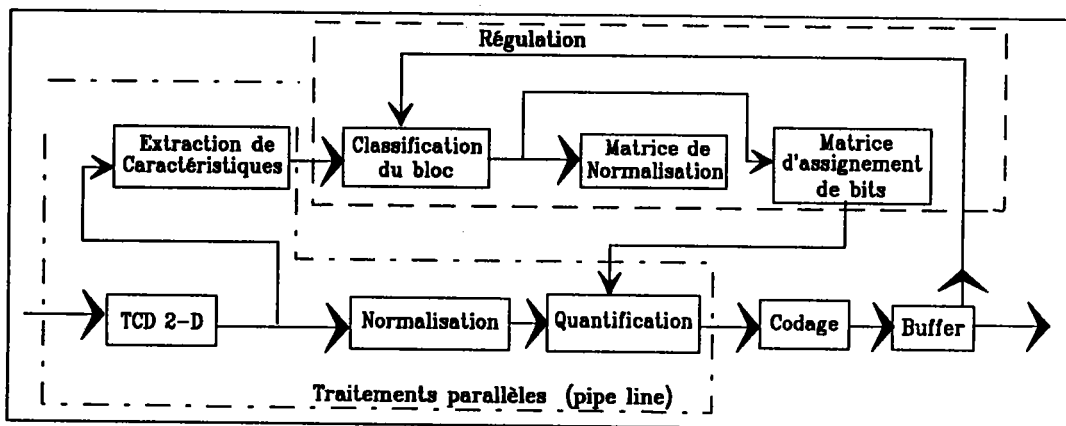


Figure N°12 Codage adaptatif intra-image

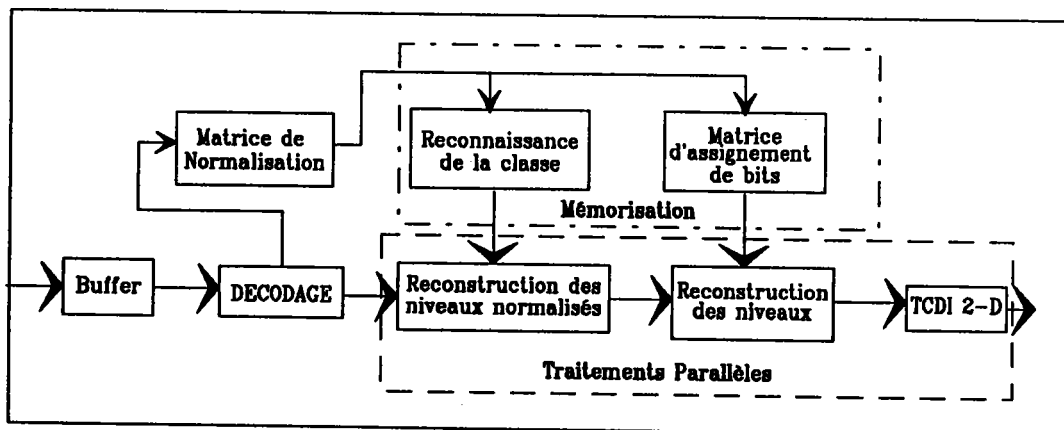


Figure N°13 Décodage par classes

Wu et al proposent dans [WU 82] une méthode basée sur une évaluation de paramètres énergétiques qui tiennent compte de la disposition directionnelle ou/et radiale des coefficients dans la matrice transformée. La classification, très sélective, est établie à la suite d'une décision qui tient compte d'un vecteur caractéristique. Cette opération provoque une répartition fine; les classes obtenues sont disjointes pour permettre d'une part d'éviter les distorsions dues au codage, d'autre part d'accommoder le nombre de classes aux situations réelles.

Un avantage de cette méthode de codage est que, non seulement elle respecte avec fidélité les particularités des blocs transformés, mais en plus le traitement mis en oeuvre pour établir le vecteur caractéristique, ainsi que les procédures de classement, restent relativement simples.

Les composantes du vecteur caractéristique de Wu et al sont ainsi définies : (*MACE*; *DIR*; *FIN*).

Mace représente l'énergie des éléments transformés dans une bande de "fréquences moyennes". L'appellation "fréquence" est souvent utilisée, bien qu'elle ne soit qu'un héritage de la transformée de Fourier.

Pour des blocs transformés de taille 16 x 16.

$$MACE = \sum_{u=2}^9 \sum_{v=2}^9 (F(u,v))^2$$

le paramètre *DIR* détermine la direction dominante de la disposition des coefficients TCD. Il est en effet sensible aux structures dans l'espace transformé dont les orientations sont horizontales ou verticales en basses et moyennes fréquences.

$$DIR = \frac{\sum_{u=0}^{12} \sum_{v=0}^{12} \text{Arctg} \left(\frac{u}{v} \right) \cdot |F(u,v)|}{\sum_{u=0}^{12} \sum_{v=0}^{12} F(u,v)}$$

avec (u,v) différent du couple (0,0)

La dernière composante du vecteur caractéristique est *FIN*. Elle évalue la diminution radiale de l'énergie sur le bloc transformé en comparant l'énergie située en basse fréquence et celle située en moyenne fréquence.

$$FIN = \frac{\frac{1}{N_1} \sum_{(u,v) \text{ de } R_1} F(u,v)^2}{\frac{1}{N_2} \sum_{(u,v) \text{ de } R_2} F(u,v)^2}$$

R_1 et R_2 sont les zones définies par les limites déterminées dans la figure suivante N°14.

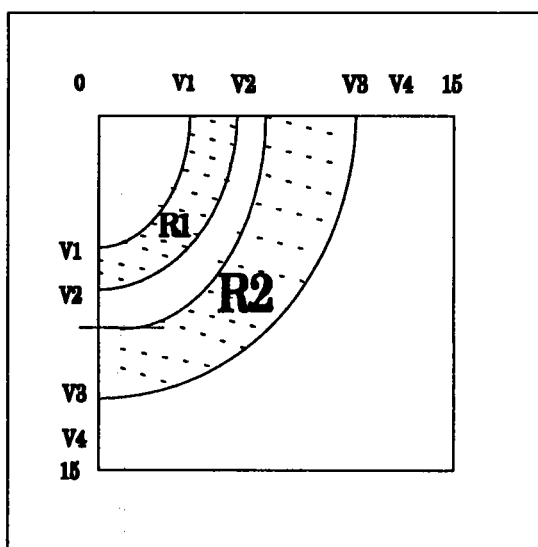


Figure N°14 Structures annulaires du spectre énergétique

Un arbre de décision est établi de façon à élaborer les différentes classes d'appartenance : Un jeu de sept seuils, quatre pour établir la sélection suivant la composante *MACE*, deux selon *DIR*, enfin le dernier seuil pour la discrimination selon la caractéristique *FIN*. Les seuils sont spécifiés à la suite de plusieurs expérimentations effectuées sur un ensemble d'apprentissage, assez représentatif pour que la classification soit fidèle à la réalité de l'espace transformé.

Dix classes sont formées. Par l'intermédiaire de *DIR* sont identifiés les groupes à structures verticales ou horizontales. Pour les classes non

directionnelles, *FIN* localise celles qui représentent des blocs images dont les textures sont fines mais contrastées. La caractéristique *MACE* et les seuils qui lui sont indexés, donnent une indication sur l'étendue des coefficients dans le domaine transformé TCD.

L'amélioration apportée par la classification décrite par Wu et al par rapport à une technique similaire à celle de Chen et al, est significative. Elle est perceptible au niveau de la qualité subjective des images restituées, puisque le codage s'adapte au contenu informatif des blocs transformés. Un autre avantage de cette méthode de classement est qu'elle améliore le taux compression de 20% à 30%.

III - C Approches adaptatives pour la quantification vectorielle

Rejoignant pour cela le codage par transformation, le codage par quantification vectorielle nécessite avant tout un découpage de l'image. Les blocs obtenus rassemblent un nombre de pixels susceptibles d'être interdépendants. La corrélation inter-blocs est ignorée en faveur de la restitution des particularités contenues dans un médaillon.

Le taux de compression ainsi que la qualité subjective des images reconstituées sont étroitement liés à la taille des blocs exploités. Pour des petites tailles de blocs, les détails, même locaux, sont pris en compte lors du codage. Cependant nous obtenons une compression importante pour des images non contrastées découpées en blocs de grande taille. Une recherche de compromis, bonne qualité visuelle et bon taux de compression, est envisageable sous forme d'adaptation de la taille des blocs pour la quantification vectorielle.

III - C - 1 Accommodation des tailles de bloc aux images codées

La diversité des motifs composant l'image est une des particularités fondamentales de cette dernière. Cette affirmation est justifiée par la richesse des caractéristiques des symboles visibles; leurs formes, leurs détails, les variations de leurs niveaux de gris ainsi que leur tolérance visuelle aux erreurs de codage. Il est donc indispensable d'accommoder la technique de compression d'images à leurs contenus.

Dans la quantification hiérarchique [NASR 85], des blocs de tailles: 2×2 , 4×4 , 8×8 et 16×16 sont utilisés, à la suite d'une segmentation spatiale de l'image en zones de structures similaires. Cette répartition différencie les plages uniformes des zones contrastées contenant des détails de taille variable primordiaux pour la compréhension de l'image décodée.

Le codeur élaboré par Nasrabadi et al, forme un arbre binaire dont les feuilles sont les zones de l'image répondant à un critère particulier. Un découpage des blocs 32×32 en quatre est réalisé. Par la suite, une identification, en fonction des structures contenues dans chaque sous-bloc 16×16 , est tentée. Les sous-blocs pour lesquels elle est réussie ont leurs adresses classées dans l'arbre, ils sont codés en fonction du résultat de l'identification. Pour les sous-blocs dont le contenu est plus complexe, un second découpage en quatre est effectué. Ainsi successivement, la reconnaissance des particularités se poursuit jusqu'à l'obtention de blocs 2×2 incluant des détails non classés.

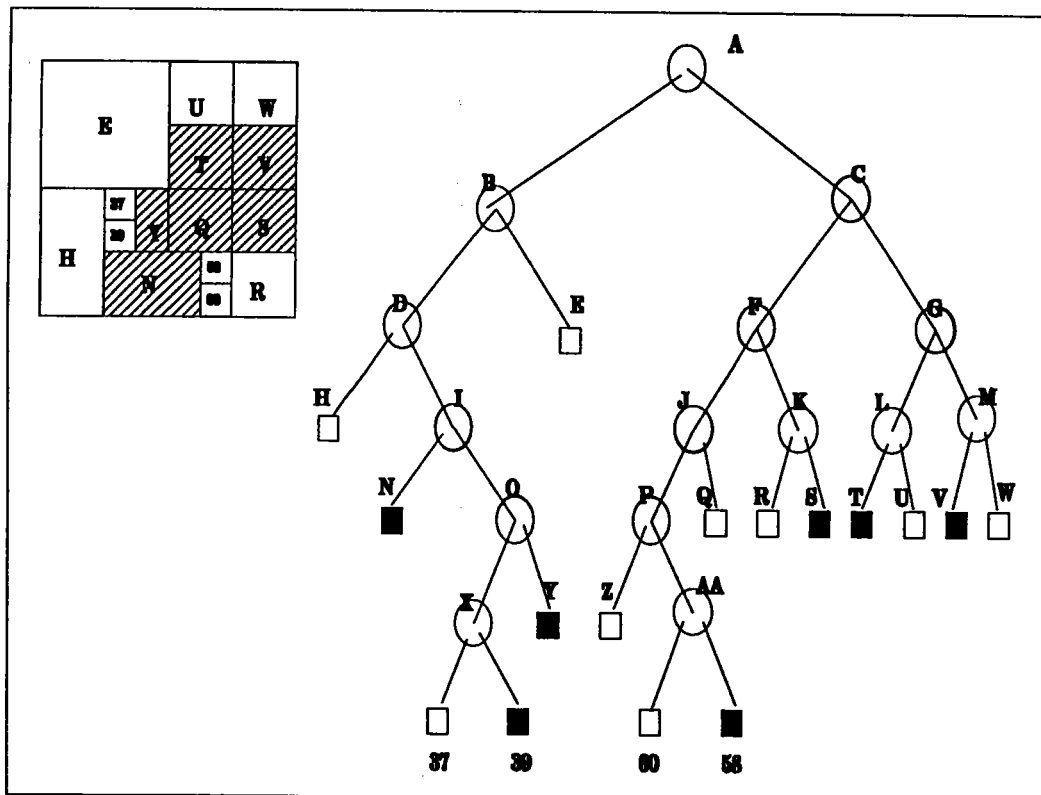


Figure N°15 Reproduction de l'image par arborescence

Les sous-blocs de pixels, dont les tailles sont égales à 2×2 et 4×4 , sont codés par un quantificateur vectoriel. Un codage par transformation est appliqué aux autres sous-blocs. L'information de position des blocs est délivrée par l'arbre binaire, elle est aussi transmise au décodeur ; la

reconstitution de l'image en dépend.

Cette méthode de codage adaptatif présente des taux de compression avantageux, tout en préservant une bonne qualité subjective de l'image restituée. En effet, les détails et les petites évolutions de luminances sont codés avec fidélité par un quantificateur vectoriel adéquat.

III - C - 2 Quantification vectorielle multi-étages

Une possibilité d'adaptation en codage d'images est d'appliquer plusieurs quantifications vectorielles successives. Cette éventualité offre au quantificateur la capacité de coder au mieux le signal image. En effet à la suite de la première quantification du vecteur incident, une seconde est réalisée par un quantificateur sélectionné d'après l'erreur de quantification. Cette même procédure peut se répéter N fois selon la précision de codage désirée.

L'étage initial de quantification utilise un dictionnaire de taille réduite formé de vecteurs pixels. Le suivant manipule, lui, des vecteurs erreurs de quantification. Ainsi d'étage en étage, l'écart entre le vecteur initial et son représentant s'amenuise. Le code final est la concaténation de toutes les adresses fournies à la suite des passages par les différents étages. Au décodage, l'image est restituée après la sommation de tous les vecteurs indexés lors des quantifications successives.

Un des avantages de la technique de quantification multi-étages est d'effectuer le codage en parcourant d'une façon sélective l'ensemble des dictionnaires stockés. Cette propriété justifie le bon rapport (qualité des images restituées / temps de calcul), meilleur que celui obtenu par un quantificateur à recherche exhaustive muni d'un dictionnaire de taille similaire. Cependant, la formation des dictionnaires reste un problème complexe; en effet si le premier dictionnaire est constitué par des vecteurs pixels, les suivants le sont par des vecteurs erreurs de quantifications. L'absence de caractéristiques connues ou d'erreurs de référence rend la conception des dictionnaires hasardeuse.

III - C - 3 Quantification vectorielle à dictionnaire rafraîchi

Pour les applications de transmission d'images animées, l'utilisation de

dictionnaires de grande taille est à bannir. En effet, le temps de calcul ainsi que la complexité des traitements sont des contraintes très sélectives pour ce type d'utilisations.

Dans [GOLD 86], est proposée une méthode de quantification vectorielle dont le dictionnaire initial est de petite taille. Conçu à la suite d'une estimation statistique des caractéristiques de l'image, ce dictionnaire est la concaténation de plusieurs sous dictionnaires locaux représentant différentes zones dans l'image. Ils sont transmis au décodeur au début de la communication. Un module de classification en entrée du codeur identifie la cellule d'appartenance de chaque vecteur. La quantification vectorielle par le sous-dictionnaire approprié est optimale si l'erreur de quantification est inférieure à une distorsion maximale prédéfinie. Si cette valeur de tolérance est dépassée, le dictionnaire initial est considéré insuffisant, son extension au vecteur incident est envisagée.

Goldberg et al proposent plusieurs techniques de restructuration du dictionnaire rafraîchi; identifier de nouvelles cellules englobant les nouveaux vecteurs ou redéfinir les représentants des cellules existantes en les remplaçant par la moyenne globale des vecteurs.

Parmi les performances annoncées, le taux moyen est de 1 à 1.5 bit/pixel pour des vecteurs formés par des blocs de 2×2 pixels.

La technique présentée ici est intéressante lorsque la scène à traiter comporte des zones de caractéristiques distinctes. Le recouvrement des sous-dictionnaires est, en effet, un problème non négligeable et lourd de conséquences; une représentativité médiocre des sous-dictionnaires initiaux nécessite un rafraîchissement important et donc une charge lourde de calcul pour le codeur.

D'autres problèmes de la quantification vectorielle à rafraîchissement du dictionnaire sont identifiées; monopolisation du canal par la transmission des sous-dictionnaires restructurés, complexité du classement simultané des sous-dictionnaires.

Chapitre IV

TECHNIQUES DE CODAGE PAR CLASSIFICATION VISUELLE

L'adaptabilité offerte par les méthodes classiques, citées dans le chapitre précédent, ne présente pas les performances espérées. Pour améliorer le taux de compression des images codées tout en gardant une qualité subjective tolérable à la restitution, une approche visuelle s'impose. Elle se traduit par une évaluation des échantillons du signal image et une quantification de leurs variations. Ces opérations permettent d'estimer l'impact visuel des différentes suites d'échantillons. Il est donc nécessaire d'inclure dans le codeur un module de traitement permettant d'identifier le signal image, de reconnaître ses composantes et leurs particularités perceptibles. De cette manière, à la transmission ne sont maintenues que les informations pertinentes indispensables à une restitution acceptable des images.

Pour chacune des trois techniques classiques, des approches visuelles sont citées dans la littérature. Mais celle qui se prête le mieux à ce genre de méthode est la quantification vectorielle. En effet, son aptitude à traiter globalement un ensemble d'échantillons et sa manière d'effectuer une recherche des plus proches voisins ou ressemblants parmi un ensemble de blocs d'éléments picturaux, la désignent logiquement à présenter le meilleur perfectionnement à la suite de l'application d'une classification visuelle.

Nous consacrerons donc le premier paragraphe de ce chapitre à la quantification vectorielle, donnant ainsi une importance moindre aux autres techniques.

IV - A Quantification vectorielle

Gresho et Ramamurthi ont été les précurseurs de l'utilisation de l'approche visuelle en quantification vectorielle [GERS 82] [RAMA 84].

Pour neutraliser les distorsions qui surviennent au niveau des contours sur les images restituées, une multitude de dictionnaires adéquats sont exploités. Ces quantificateurs vectoriels visuellement adaptatifs sont conçus pour améliorer le rendu des images codées, tout en maintenant un taux de compression élevé.

La technique décrite par Gersho et Ramamurthi identifie la ou les caractéristiques visuelles attachées au bloc d'image incident. Les principales particularités recherchées sont les contours et leur disposition dans l'image. D'autres méthodes, plus fines dans la classification, étendent les caractéristiques aux textures.

La méthode utilisée par Gersho, nécessite la formation pour chaque bloc (4 x 4) d'une table de gradients horizontaux et une autre de gradients verticaux. Ces tables sont calculées d'une façon locale; pour des pixels adjacents. Si la valeur du gradient dépasse celle de la moyenne affectée d'un coefficient pondérateur, le signe du gradient est retenu sur la table dans la direction qui lui correspond; si ce n'est pas le cas, c'est un zéro. Les tables matérialisant un contour sont reconnaissable par le nombre de signes (au moins trois) et par la disposition de ces derniers.

++++	+++ -	+++ -
++++	+++ -	++ - -
- - - -	+++ -	+ - - -
- - - -	+++ -	- - - -
++++	++ - -	++++
++++	++ - -	+++ -
++++	++ - -	++ - -
- - - -	++ - -	+ - - -

Une normalisation spécifique est appliquée à chaque type de bloc, de façon à faciliter la recherche de son plus proche vecteur voisin dans le dictionnaire qui correspond à chaque classe visuelle. Le traitement affecté au

classement entraîne une charge supplémentaire pour le codeur. Cependant, le gain qu'apporte cette répartition au niveau de la quantification vectorielle est significatif.

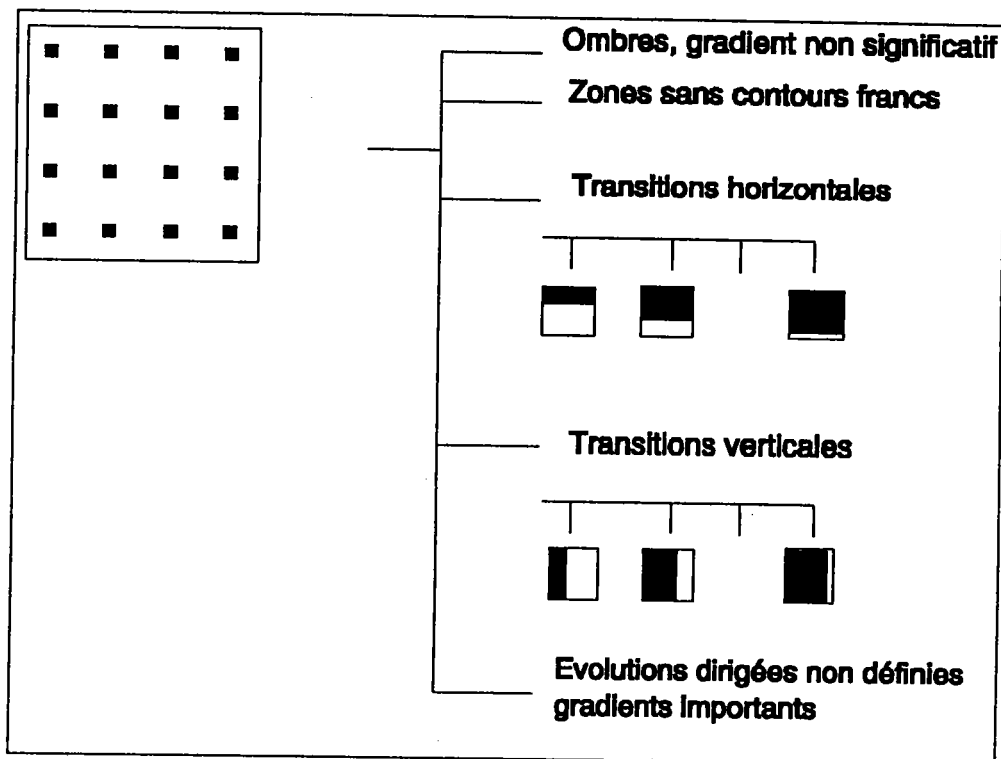


Figure N°16 classification visuelle pour la quantification vectorielle

Avec une approche algorithmique similaire à celles présentées dans [GERS 82] [RAMA 84], Y. Zhao propose un quantificateur vectoriel basé sur une identification des structures contenues dans l'image. Le procédé consiste à reconnaître plusieurs tendances de variation des luminances en se basant sur un vecteur de critères visuels.

L'évolution des luminances dans une image prend plusieurs formes : variations brutales, évolutives ou insignifiantes pour l'observateur. Ces évolutions traduisent les formes, les objets et les structures composant l'image. Leur localisation et leur exploitation pour le codage permet d'optimiser toutes les procédures de traitement, en maintenant une bonne qualité subjective pour les images restituées au décodage.

La quantification vectorielle exploite plusieurs sous-dictionnaires représentant les différentes familles visuelles; blocs uniformes, blocs contours et blocs composés de variations évolutives de niveaux de gris. Des paramètres sont identifiés pour représenter les particularités déterminant chacune des familles de blocs.

Les images sont découpées de façon contiguë, sans recouvrement, en blocs carrés de quatre lignes quatre colonnes. L'avantage offert par cette taille de bloc est la possibilité de localiser de petits détails dans l'image et de les prendre en compte au codage. Pour chacun des blocs (4 x 4), des gradients directionnels normalisés sont calculés localement. Cette dernière étape nécessite l'extraction des moyennes des luminances affectées aux couples de pixels adjacents dans le même bloc ainsi que les distances absolues qui les séparent:

$$M_v[l,c] = \frac{A[l,c] + A[l+1,c]}{2} .$$

M_v moyenne locale de deux luminances successives sur une même colonne.

$$l,c=0,1,2,3.$$

$$D_v[l,c] = \frac{A[l,c] - A[l+1,c]}{M_v[l,c]}$$

$$l = 0,1,2 \quad c = 0,1,2,3.$$

$D_v[l,c]$ gradient vertical local.

La même opération est réalisée pour les variations horizontales.

$$M_h[l,c] = \frac{A[l,c] + A[l,c+1]}{2} .$$

M_h moyenne locale de deux luminances successives sur une même ligne.

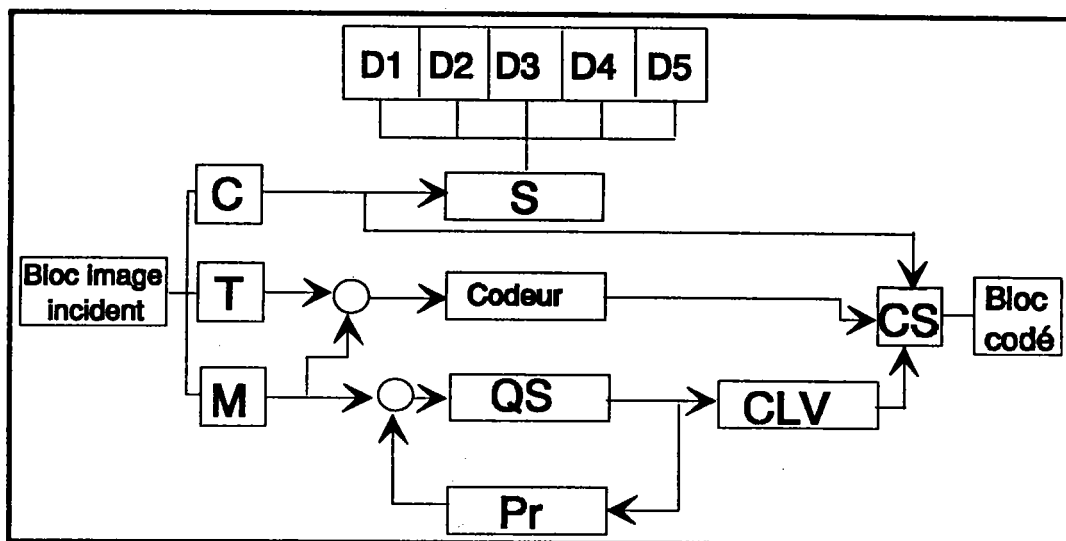
$$D_h[l,c] = \frac{A[l,c] - A[l,c+1]}{M_h[l,c]}$$

$$l = 0,1,2,3 \quad c = 0,1,2.$$

$D_h[l,c]$ gradient horizontal local.

La quantification de ces gradients aide à l'évaluation des variations de luminance. Leur comparaison avec des seuils prédéfinis contribue à

l'identification de la famille visuelle d'appartenance du bloc image. Les seuils sont déterminés à la suite de plusieurs essais subjectifs. Pour cette opération, la sensibilité des yeux pour les différentes variations de luminance est prise en compte, ainsi que l'élimination des effets provoqués par les bruits.



C : Classification visuelle de l'image.

T : Retard.

M : Calcul de la moyenne.

Pr : Prédicteur.

QS : Quantificateur scalaire 4 bits.

CLV : Codage à longueur variable.

CS : Codeur de sortie.

S : Sélecteurs de sous-dictionnaire.

Di : les différents sous-dictionnaires

Figure N°17 schéma synoptique du codage par quantification vectorielle visuelle adaptative.

Un arbre de sélection détermine la classe d'appartenance des blocs en entrée du codeur d'images. La fonction de décision est caractérisée par un vecteur de dimension deux. Elle tire parti des cumuls de dépassements des seuils affectés à chacun des deux gradients et comptabilisés par deux compteurs incrémentaux. En effet, à chaque fois que le seuil est dépassé par les valeurs caractéristiques, le compteur correspondant est incrémenté. Le total des dépassements détermine la classe visuelle.

Les classes visuelles suivantes sont identifiées :

- bloc de luminances uniformes.

- bloc à transition verticale.
- bloc à transition horizontale.
- bloc à transition simple de direction indéfinie
- bloc à transitions multiples de directions indéfinies.

L'organe de décision du codeur réside dans le module de classification visuelle, il sélectionne le sous-dictionnaire par l'intermédiaire duquel la quantification vectorielle est appliquée au bloc d'écart de luminances en entrée.

Pour chaque famille un sous-dictionnaire est conçu à partir de plusieurs images de référence. Le nombre de vecteurs contenus dans les différents sous-dictionnaires n'est pas identique; en effet, la probabilité d'appartenance aux différentes familles est étroitement liée aux structures qui composent l'image, elle dépend aussi des valeurs pré-établies des seuils de sélection. Les dictionnaires matérialisant les transitions verticales et horizontales restituent une bonne qualité d'image à partir d'un nombre limité de vecteurs : 64. Les dictionnaires représentant les transitions multiples de directions indéfinies ne donnent une qualité subjective acceptable qu'à partir de 256 vecteurs. Malgré la faible probabilité d'apparition de cette catégorie de blocs le nombre de vecteurs nécessaire est important, à cause de la diversité des évolutions multiples non orientées dans l'image.

L'action de quantification est appliquée sur des vecteurs d'écart de luminance par rapport à la moyenne du bloc. Ce réajustement des composantes permet d'étendre le domaine de représentativité des dictionnaires.

$$M = \frac{1}{N} \sum_{i=0}^7 \sum_{j=0}^7 A(i,j) \quad (1).$$

M : La moyenne du médaillon:

Le vecteur à quantifier a pour composantes:

$$E(i,j) = A(i,j) - M. \quad \text{avec } i,j = 1,2,3,4.$$

La moyenne des luminances du bloc est traitée séparément par codeur prédictif différentiel. Le prédicteur utilisé est une fonction des valeurs moyennes attachées aux deux blocs précédents, le premier se trouve sur la même ligne que le bloc à coder et le second sur la même colonne que le bloc à coder. A la sortie du quantificateur scalaire nous obtenons des codes de

longueurs variables.

Chaque bloc est donc représenté par un mot code d'identification du sous-dictionnaire - sur 3 bits -, par l'adresse du vecteur écart le plus proche - sur 6 à 8 bits - et enfin par le code de sa moyenne qui vaut en moyenne 0,7 bits.

Le taux moyen avancé par Y. Zhao et de 0,61 bit/pixel. Cette compression maintient néanmoins un taux d'erreurs visible. En effet, des distorsions significatives sont perceptibles au niveau des contours. Toutefois, l'amélioration apportée par cette méthode de codage par classification visuelle est très marquante par rapport aux techniques classiques de quantification vectorielle.

Les méthodes de classification décrites auparavant nécessitent un charge de prétraitements très coûteuse pour l'équipement de codage d'images. Cette charge réside dans le calcul de plusieurs gradients locaux pour chaque couple de pixels. Pour contourner cette étape, A. Davignon a optimisé la classification visuelle [DAVI 89]. Une répartition de l'espace image en familles de transitions primitives orientées est effectuée. Elle est fondée sur le calcul de la moyenne du bloc image et de l'écart-type présenté par les luminances du bloc par rapport à cette moyenne.

L'activité des luminances est quantifiée à l'aide de seuils de décision. Si l'écart-type est supérieur à une certaine valeur, le bloc est déclaré contenant un contour. Cette opération permet de reconnaître les blocs uniformes. Une représentation binaire est choisie pour les médaillons actifs. A la position des pixels dont la luminance dépasse la valeur moyenne, la valeur "1" est affectée; pour les autres positions la valeur "0" leur est réservée.

Cette nouvelle description des blocs permet de différencier huit classes de transitions primaires orientées. Elles sont définies pour représenter les possibilités de contours horizontaux verticaux et diagonaux. Ces derniers sont appelés par l'auteur Blocs Binaires de Contours Fondamentaux : BBCF.

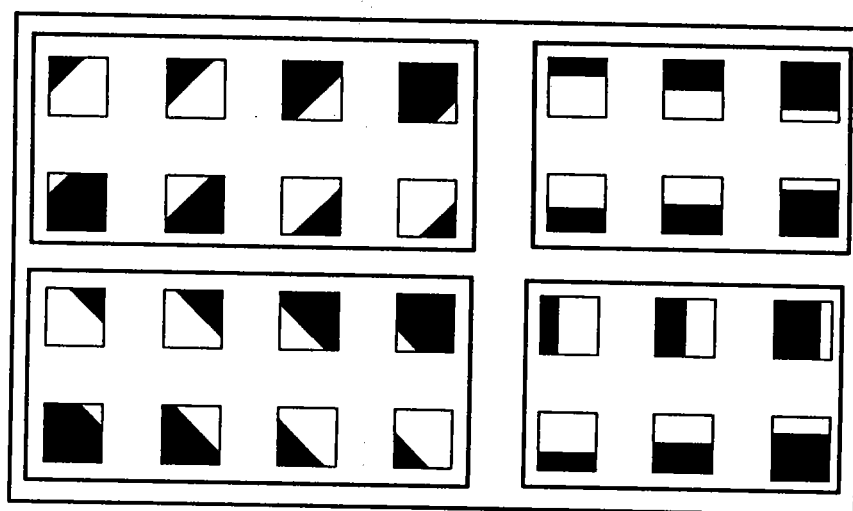


Figure N°18 Classification visuelles

Signalons que ces familles ne font pas une description exhaustive des transitions, mais elles y contribuent largement [DAVI 89]. Par une comparaison logique entre la reproduction binaire du bloc incident et les différents BBCF, le groupe d'appartenance du médaillon en entrée est déterminé. Cette opération est une quantification vectorielle manipulant une distance particulière élaborée par A. Davignon. Pour deux vecteurs a et b , la distance $d(a,b)$ les séparant est définie par la relation :

$$d(a,b) = \sum_{i=0}^7 Ci(a) \text{ EXOR } Ci(b)$$

Ci désigne la fonction qui retourne la i -ième coordonnée binaire d'un mot ou d'un vecteur. EXOR l'opérateur logique OU exclusif. Toutes les classes sont parcourues pour chaque bloc recherchant le BBCF qui lui est le plus proche. Ce dernier sera la représentation la plus convenable pour reconstituer l'image. La discrimination se fait en tenant compte d'un seuil maximum de distorsion égal à 3. Si la distance minimale séparant un bloc incident et son plus proche voisin BBCF dépasse ce seuil, le bloc en question est classé dans les "détails aléatoires", il sera codé avec une autre technique [DAVI 89].

Une autre particularité de la technique d'André Davignon est de segmenter l'image en blocs (8x8). Des tests d'identification des zones uniformes déterminent si cette caractéristique s'étend sur tout le bloc (8x8) ou si elle est seulement locale, ne concernant qu'une partie des médaillons (4 x 4) le composant. Cette nouvelle adaptation apporte un gain en débit; mais aussi une complexité supplémentaire pour l'équipement de codage-décodage.

IV - B Codage par transformé en cosinus discrète

Les approches adaptatives utilisant la TCD tirent parti de la répartition statistique de l'énergie dans le domaine transformé. Une classification détermine l'allocation des bits nécessaires pour coder le bloc transformé en fonction de l'étendu de son "spectre" énergétique. Toutefois, dans un besoin d'amélioration des performances des solutions s'appuyant sur une approche visuelle sont proposées.

Selon Lohsheller [LOHS 84] la classification visuelle est fondée sur quatre caractéristiques, grâce auxquelles des structures orientées sont détectées sur l'image. Les paramètres sont calculés à partir des luminances des pixels de l'image.

- La différence quadratique moyenne horizontale :

$$D_h = \frac{1}{N(N-1)} \sum_{m=0}^{N-1} \sum_{n=0}^{N-2} (x_{m,n} - x_{m,n+1})^2$$

avec $x_{m,n}$ est la luminance du pixel situé à la position (n,m) dans le bloc image.

- La différence quadratique moyenne verticale :

$$D_v = \frac{1}{N(N-1)} \sum_{m=0}^{N-2} \sum_{n=0}^{N-1} (x_{m,n} - x_{m+1,n})^2$$

- La différence horizontale maximale :

$$H_e = \text{MAX}_{m,n} | x_{m,n} - x_{m,n+1} |$$

- La différence verticale maximale:

$$V_e = \text{MAX}_{m,n} | x_{m,n} - x_{m+1,n} |$$

Une répartition grossière est élaborée à partir des deux premières caractéristiques D_h et D_v . A la suite de cette opération, cinq familles sont définies. Les frontières qui les délimitent sont déterminées par un vecteur de

paramètres fixes (T_1, T_2, T_3). Le paramètre T_3 permet de différencier les structures verticales, horizontales ou diagonales.

Suivant une règle discriminatoire, l'appartenance visuelle du bloc en entrée est confirmée. Cependant, cette préclassification est insuffisante, car la variation structurelle des images ne se limite pas aux classes primaires décrites. Cette constatation se traduit par le fait que les paramètres de sélection ne permettent pas de distinguer entre les blocs de contours multiples et les blocs représentant de larges plages de textures, ces deux types de blocs génèrent des vecteurs spécifiques similaires.

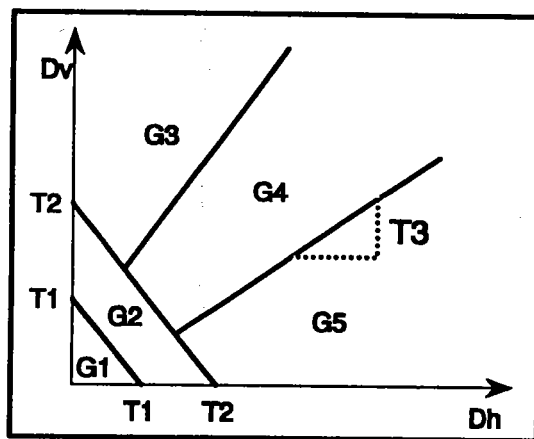


Figure N°19 Règles de pré-classification des blocs en 5 classes

Les caractéristiques H_e et H_v sont utilisées alors pour affiner la répartition. Les zones G_3 , G_4 et G_5 sont subdivisées pour former six nouvelles sous-classes. Le critère de sélection est un paramètre T_4 , appliqué à H_e et H_v . Le vecteur test final (T_1, T_2, T_3, T_4) est fixé de façon à obtenir une équirépartition des classes [LOHS 84-1].

La version améliorée du codage par classification de Lohsheller est fondée sur l'élaboration de seuils adaptatifs permettant la discrimination de certains coefficients transformés.

Par une suite de tests de visibilité des actions provoquées par les coefficients, une fonction γ_{ij} est déterminée [LOHS 84-2]. Elle identifie le seuil différentiel de visibilité pour chacune des fréquences de position (i, j) du bloc transformé. Le rapport de cette valeur avec l'écart type correspondant à la même position dans le bloc image fournit une fonction $W_{ai,j}$.

$$W_{ai,j} = 20 \log_{10} \frac{\gamma_{i,j}}{a_{i,j}}$$

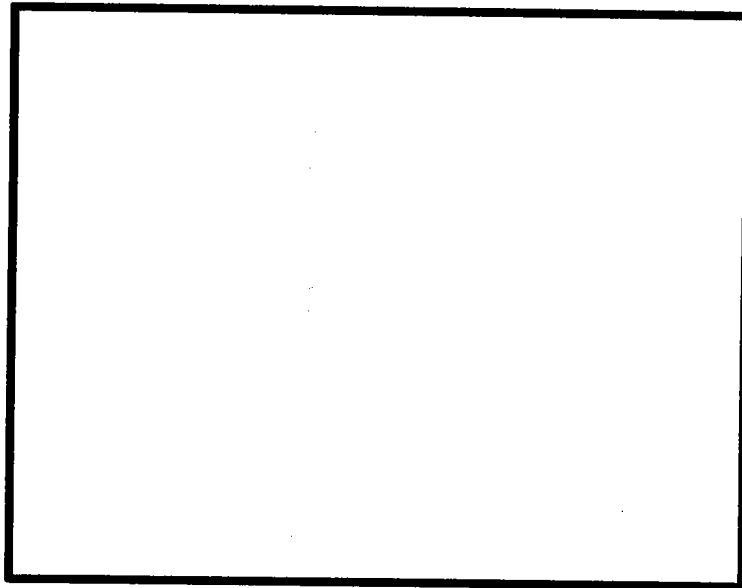


Figure N°20 Valeurs mesurées des seuils différentiels de visibilité $\gamma_{k,l}$ dans l'espace transformé (taille de bloc : 8x8)

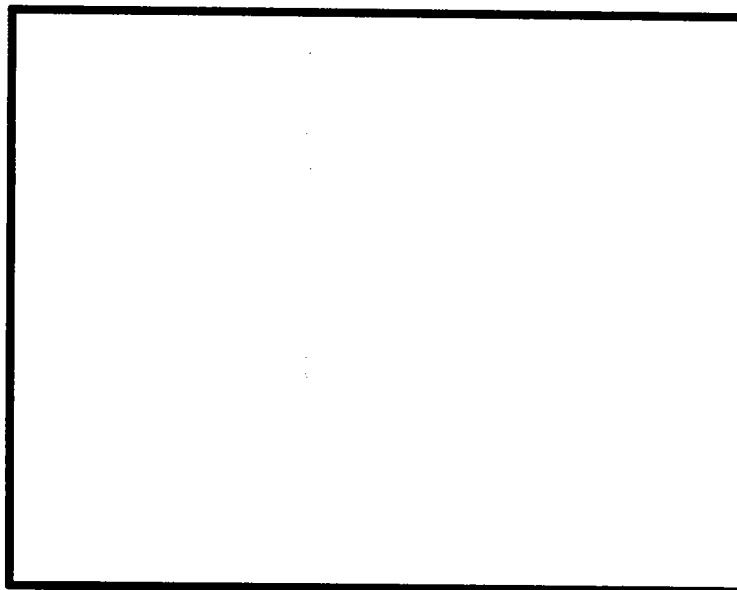


Figure N°21 Rapports normalisés des seuils différentiels de visibilité dans l'espace transformé ($W_a = \gamma / \sigma_a$ en db)

Cette fonction détermine l'impact visuel de chaque coefficient de la matrice transformée. Elle met en valeur tous les éléments dont l'amplitude $a_{i,j}$ est

inférieure à leur seuil de visibilité, nous obtenons pour ce type de coefficients un rapport $W_{ai,j}$ égal à 0 db.

Un quantificateur scalaire élabore les codes des coefficients en fonction de leur classe de visibilité et en fonction de l'importance du nombre de bits qui leurs sont alloués.

Pour l'ensemble des huit classes, le nombre de coefficients à coder varie de 1 à 16; cela nécessite un nombre total de bits variant de 10 à 67. Le coefficient matérialisant la composante continue $F(0,0)$ est traité à part. La longueur du mot code qui est réservé à $F(0,0)$ est différente de celles associées aux autres coefficients. En effet, pour ces derniers la taille du mot code varie entre 2 bits et 7 bits.

Les performances annoncées pour cette méthode de codage sont bonnes en termes de taux de compression à qualité subjective donnée : sur des images de type portrait, on obtient un débit moyen de 0.32 bit/pixel dont 0.274 bit/pixel pour le codage des coefficients.

IV - C Codage Différentiel

Cette technique a permis de développer des codeurs décodeurs avec de bonnes performances. Le fait de traiter séparément les éléments picturaux donne au codage prédictif une robustesse au niveau de la restitution des images. Mais dès que nous aspirons à obtenir un taux de compression important les erreurs de quantification prennent des proportions importantes et intolérables surtout autour des contours et des zones contrastées. Une adaptation aux structures composant l'image serait bénéfique pour les performances du codage.

L'analyse des dégradations introduites par le codage MICD et de leurs seuils de visibilité a poussé Dewitte et al [DEWI 81] à adapter la loi de quantification de l'erreur de prédiction aux particularités de l'image, et ce afin de gagner en performance de codage.

Afin de pouvoir choisir un prédicteur adéquat, une étude paramétrique définit l'appartenance de la zone image. Elle est fondée sur l'évaluation de la dynamique de l'évolution des luminances des pixels. Elle permet la distinction entre plusieurs catégories de blocs :

- Zones quasi-uniformes, où la dynamique locale (différence entre 2

points adjacents) est inférieure à 2% de la dynamique totale.

- Zones de détails peu contrastés, où la dynamique locale peut atteindre 5% de la dynamique totale. Dans ce cas comme dans le précédent, si l'erreur de prédiction est faible, la visibilité du bruit de quantification est élevée.
- Zones de détails contrastés, où la dynamique locale comme le gradient peuvent atteindre 15% de la dynamique totale. Si la visibilité de défauts y est plus réduite, elle est par contre difficile à modéliser, du fait du nombre de paramètres qui interviennent dans la description des textures.
- Zones de contours, dans ces plages le contraste des luminances est élevé (gradient supérieur à 15% de la dynamique totale) , mais la structure est beaucoup plus facilement identifiable : en effet, la largeur de transition est de 1 à 2 points d'image. Les défauts pouvant affecter l'image sont plus facilement modélisables [KRET 83].

Une étude statistique pour évaluer le nombre des transitions dans les images de télévision ainsi que leur impact sur le taux d'erreurs de prédiction a été menée. Elle montre une prédominance pour les évolutions primaires (verticales et horizontales). Mais le résultat le plus probant est la constatation suivante : les points qui présentent une erreur de prédiction dépassant 50 ou 60, dans le cas de l'utilisation d'un prédicteur fixe, correspondent pour leur majorité à des points appartenant aux contours. Cela signifie que si la prédiction de ces points particuliers pouvait être améliorée, la dynamique du quantificateur, c'est à dire la différence entre ces niveaux extrêmes, serait réduite et les performances du codeur en bénéficieraient.

Les constatations précédentes justifient tout à fait une adaptation de la méthode prédictive par un procédé basé sur l'extraction des particularités visuelles de l'image. Cette adaptation se traduit par un choix de prédicteur suivant l'évolution de luminances détectée.

Une projection spatiale décrit la direction du contour auquel appartient le pixel à coder. Cette direction est définie à la suite de la minimisation des erreurs de prédiction découlant de plusieurs prédicteurs (cf. Fig. N°24). En tenant compte des éléments déjà disponibles dans le récepteur, on affecte l'orientation du contour déterminée au codage et donc l'estimateur spatial adéquat à partir duquel la valeur de la luminance sera

reconstituée.

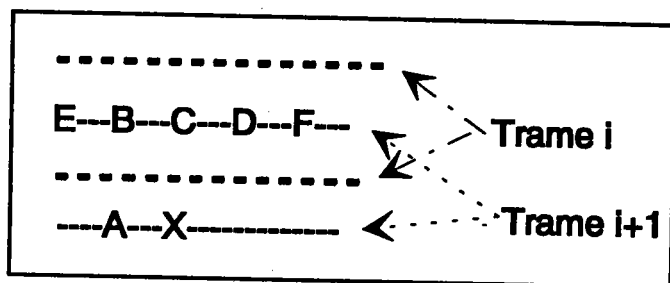


Figure N°22 Point utilisés par les prédicteurs

Orientations des contours	Prédicteurs I	prédicteurs II
153.4°	$\frac{A+B}{2}$	$\frac{2A + C}{3}$
135°	E	$\frac{2A + C}{3}$
116.6°	B	$\frac{A + 2C}{3}$
90°	C	$\frac{A + 2C}{3}$
63.4°	$\frac{C + D}{2}$	$\frac{2A + C + D}{4}$
45°	D	$\frac{A + D + F}{4}$
26.6	F	$\frac{2A + D + F}{4}$

Figure N°24 Exemples de prédicteurs adaptatifs

La diversité des cas envisageables fait que cette technique est très sensible aux erreurs de transmission. Elles produisent une désynchronisation entre l'émetteur et le récepteur. Ce genre d'erreur se propage sur le reste de l'image et affecte le calcul de prédiction, le choix du prédicteur ou celui du quantificateur.

Pour augmenter la stabilité des systèmes prédictifs adaptatifs face aux différentes erreurs auxquelles ils sont exposés, le rafraîchissement de l'image s'impose. Cette contrainte permet de limiter l'étendue des pixels affectés par les erreurs et de remplacer des valeurs erronées sur l'image. Au détriment du débit, le rafraîchissement de l'image renforce la robustesse du système de codage en comblant une de ses principales lacunes, la répercussion des erreurs sur les valeurs codées successivement.

Chapitre V

METHODES HYBRIDES DE CODAGE

L'association de plusieurs techniques de compression permet d'accumuler les performances du codage d'images. Nous nous proposons ici de traiter une méthode hybride qui nous a semblé intéressante du point de vue algorithmique est qui apporte une nouvelle approche du problème de codage sans cumuler la complexité des techniques utilisées.

V - A La quantification vectorielle appliquée dans le domaine transformé

L'utilisation de cette technique découle directement de la constatation d'après laquelle les coefficients obtenus à la suite d'une transformée mathématique ne sont pas complètement décorrélés. En effet, entre les coefficients persiste une fonction d'interdépendance. Cette corrélation résiduelle est due principalement à la matrice image initiale et aux évolutions des luminances qu'elle présente.

La distribution énergétique du spectre est variable en fonction du contenu de l'image, cela se répercute directement sur l'importance des représentativités des différents coefficients. De plus, cet état de fait est complètement quantifiable, en effet, un module de traitement classifie les blocs transformés selon des paramètres énergétiques. De cette manière sont identifiés les éléments nécessaires au codage.

Plusieurs auteurs ont développé des codeurs à base de quantification vectorielle des coefficients transformés. Dans [NASR 88], Nasrabadi et al cite plusieurs cas de figure, présentant chacun des approches particulières.

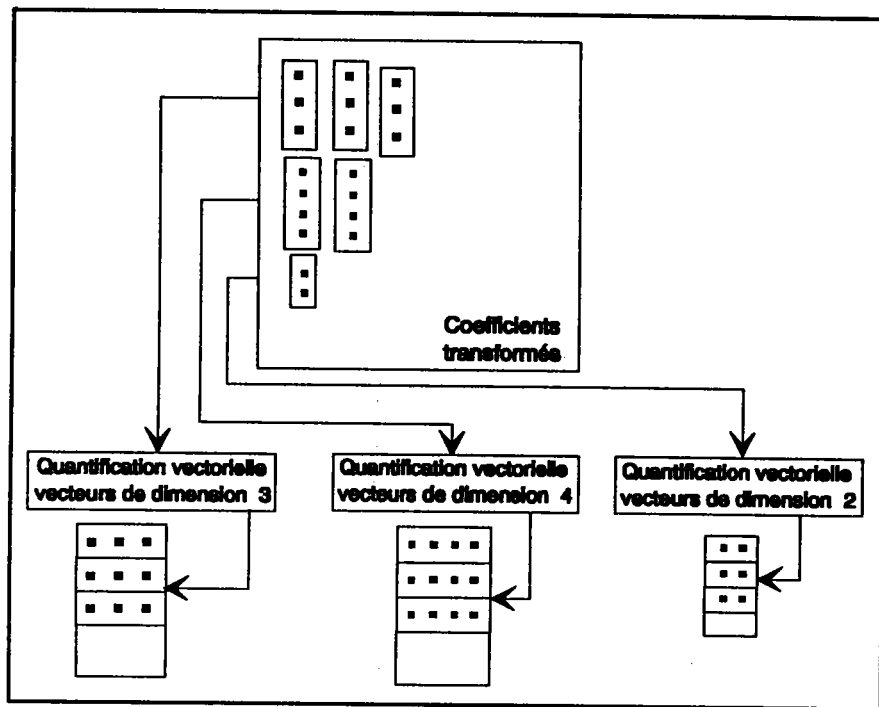


Figure N°25 Quantification vectorielle sur l'espace transformé
Multi-classes et dictionnaires

Nasrabadi [NASR 84], quant à lui, aborde le problème par des adaptations aux différentes classes énergétiques de la taille et de la position des vecteurs coefficients. Après avoir écarté les coefficients "hautes fréquences", l'adaptation est accomplie par la sélection de plusieurs vecteurs de dimensions différentes et ce, en correspondance avec les distributions statistiques dans la direction verticale de la matrice transformée. La taille des vecteurs est déterminée par un calcul préalable de fonction de corrélation entre les éléments d'un même vecteur. Un test permet de vérifier si le résultat est suffisant pour réaliser le codage du vecteur avec la dimension qui lui est affectée. La quantification vectorielle est appliquée à l'aide de dictionnaires formés de vecteurs de différentes tailles.

Le débit moyen obtenu se décompose en :

$$R = A_1 C_1 \log_2 \frac{M_1}{K_1} + \dots + A_i C_i \log_2 \frac{M_i}{K_i}$$

avec

- A_i Le nombre de fois où le dictionnaire i est utilisé.
- M_i La taille du dictionnaire.

- K_i La dimension de ses vecteurs.
- C_i Le pourcentage des coefficients retenus.

L'approche de Nasrabadi met en oeuvre un dispositif conséquent en temps de calcul pour réaliser la classification des blocs et la détermination de la dimension maximale de chaque vecteur.

V - B Codage hybride; adaptation visuelle

V - B - 1 Quantification vectorielle dans l'espace transformé

Pour des raisons d'amélioration des performances surtout de la qualité subjective, J. P. Marescq a étudié une technique de quantification vectorielle dans l'espace transformé, fondée sur l'exploitation des structures perceptibles sur l'image télévisuelle [MARE 86].

Une des particularités de cette méthode est qu'elle effectue un pré-classement suivant les propriétés visuelles des blocs d'image (4x4). Des dictionnaires adaptés à chaque famille sont constitués par la technique Linde Buzo Gray (LBG).

Une autre particularité de la méthode de Marescq est d'évaluer l'énergie totale de chaque bloc transformé. Cette entité intervient directement sur les tests de classements. Elle est calculée à partir des valeurs de tous les coefficients transformés sauf du représentant de la moyenne de la matrice (4 x 4) d'image :

$$A_{tot} = \sum_{i=1}^N (F_i^2) - F_1^2$$

F_i : Coefficients transformés.

F_1 : Coefficient de moyenne.

N : Nombre total de coefficients (dans ce cas:16).

Rappelons que pour garder une parfaite qualité visuelle des images reconstituées, le coefficient F_1 est transmis indépendamment des autres. En effet, 8 bits sont alloués pour le codage de F_1 , il présente une dynamique de 0 à 255.

L'activité A_{tot} , représente l'énergie totale véhiculée par les coefficients transformés. Ce terme donne une idée de l'activité spatiale des blocs, en quantifiant le taux de variation des luminances des pixels les composant.

Toutefois, si A_{tot} permet de répartir grossièrement les blocs dans des catégories d'activités différentes, il n'est pas suffisant pour distinguer des structures complexes : contours, textures etc.

Les éléments de la matrice obtenue après la transformée en cosinus discrète ont une signification semblable à ceux obtenus après une transformée de Hadamard. Ils sont des facteurs de pondération pour des transitions primaires ou des compositions de transitions.

La représentativité visuelle de la matrice transformée est partiellement identifiée. Quelques correspondances entre l'espace image et l'espace transformé sont exploitées. Cela se traduit principalement par une série de constatations :

- La première ligne, respectivement première colonne de la matrice transformée, en excluant l'élément qui correspond à la moyenne F_1 , reproduisent dans l'image essentiellement les contours verticaux, respectivement horizontaux.
- si la diagonale de la matrice transformée est importante, le coefficient F_1 toujours écarté, le bloc de pixel correspondant représente les médaillons d'image présentant une symétrie centrale.
- D'autres "tendances d'évolutions" sont identifiées à partir d'activités partielles.

Un vecteur de caractéristiques de perception est défini à partir d'activités partielles ($A_{diag}, A_{sup}, A_{inf}, A_{lig}, A_{col}$). Il est formé par la somme des carrés des coefficients localisés dans une zone de la matrice transformée et liés à une structure visuelle (cf. Fig.).

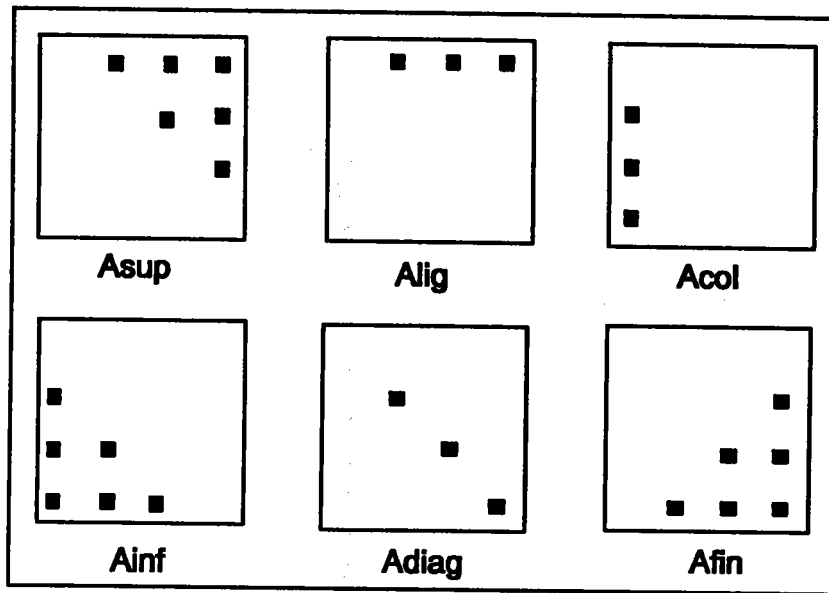


Figure N°26 Choix des coefficients transformés en fonction des classes

Un arbre de sélection, dont la fonction de décision exploite les activité spatiales partielles, est mis en oeuvre. Des paramètre caractéristiques λ sont définis pour aboutir à la répartition finale de l'ensemble des blocs de l'image par comparaison à des seuils:

$$\lambda_{diag} = \frac{A_{diag}}{A_{tot}}$$

$$\lambda_{sup} = \frac{A_{sup}}{(A_{tot} - A_{diag})}$$

$$\lambda_{inf} = \frac{A_{snf}}{(A_{tot} - A_{diag})}$$

$$\lambda_{lig} = \frac{A_{lig}}{A_{sup}}$$

$$\lambda_{col} = \frac{A_{col}}{A_{sup}}$$

Les blocs homogènes sont déterminés par une comparaison entre A_{tot} et un seuil variable défini par la loi de Weber transposée dans le domaine transformé [MARE 86]. Cette loi décrit la sensibilité visuelle par rapport aux variations de luminances. A_{tot} est aussi utilisé pour classifier les blocs image suivant l'activité spatiale totale; blocs à variations faibles, blocs à évolutions brutales etc.

Les structures visuelles, quant à elles, sont identifiées par cette méthode à la suite de la localisation des coefficients transformés les plus significatifs. Cette approche discrimine les zones non représentatives. L'importance de l'activité "fréquentielle" d'une zone donnée du bloc transformé détermine la direction dominante du gradient de luminance du bloc image. Puisqu'à gauche de la première diagonale du bloc transformé (respectivement à sa droite) sont identifiés des coefficients représentant les blocs dont le gradient de luminance est vertical (respectivement horizontal).

L'arbre de sélection est conçu après une série de tests sur des images de référence présentant des composantes visuelles variées. Le vecteur caractéristique ($A_{diag}, A_{sup}, A_{inf}, A_{lig}, A_{col}$) et l'ensemble des seuils qui lui sont associés permettent une classification dite visuelle. Cette appellation est justifiée, puisque les classes obtenues se différencient par les particularités perceptibles : blocs homogènes, contours orientés etc..

Une variante de la méthode Marescq de classification est citée par C. Labit [LABI 88]. Elle introduit la reconnaissance des blocs image présentant des textures par un nouveau critère caractéristique A_{fin} en regroupant les carrés des coefficients "hautes fréquences".

$$\lambda_{fin} = \frac{A_{fin}}{A_{tot}}$$

Une fois les différentes zones de coefficients pertinents fixées, un quantificateur adaptatif traite les vecteurs formés à partir des éléments non écartés. Par un parcours en zigzag les composantes des vecteurs respectent la croissance du rang des "fréquences spatiales" qu'elle représentent.

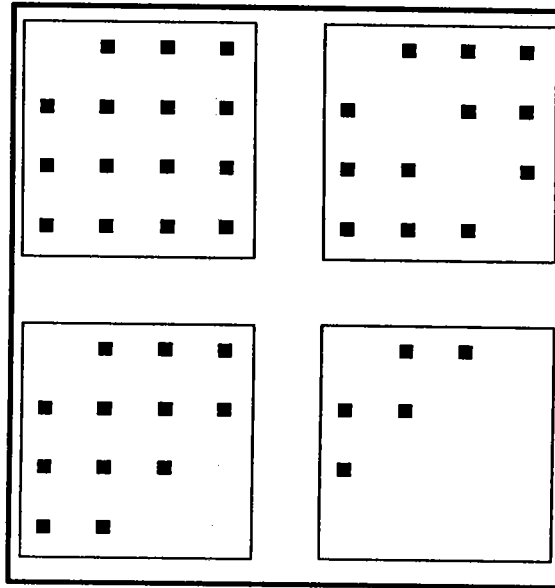


Figure N°27 Echantillonnage adaptatif des coefficients transformés selon les classes

Le quantificateur vectoriel exploite plusieurs dictionnaires correspondants aux 12 classes. Chacun d'eux est formé de vecteurs de dimension variant entre 5 pour les blocs homogènes et 15 pour les blocs images dépourvus de particularités visuelles.

La mesure de distorsion retenue est la distorsion vectorielle. Elle est appliquée aux vecteurs incidents pour trouver leurs représentants les plus proches dans les sous-dictionnaires. Ces derniers sont initialement constitués par l'algorithme LGB d'optimisation et de classement.

Les performances de cette technique de codage par classification visuelle ont été mesurées par le nombre moyen de composantes prises en compte pour le codage:

$$n_c = \sum_{i=1}^{\text{nb de classes}} p_i \dim \text{vect}_i$$

avec :

- $\dim \text{vect}_i$: dimension du vecteur de composantes de la classe i .
- p_i : probabilité d'occurrence de la classe i .

Le schéma de codage que nous avons décrit a été conçu pour les images télévisuelles. La principale contrainte requise était donc une parfaite qualité subjective des images décodées. Cette obligation fait que le nombre de

coefficients écartés reste très limité, cela agit directement sur la taille des vecteurs, donc, sur l'optimisation de la représentativité du dictionnaire.

Le quantificateur vectoriel reste assez rigide puisque son adaptation ne concerne que des caractéristiques limitées. En effet, pour répondre à la grande variété des structures visuelles composant les images, la taille du dictionnaire doit être assez conséquente. L'exemple présenté par C. Labit décrit un dictionnaire qui peut atteindre une taille de 4096 vecteurs.

Une solution est effectivement envisageable, il s'agit du rafraîchissement temporel du dictionnaire. Cela se traduit par un enrichissement simultané du dictionnaire avec de nouveaux vecteurs. Ceux dont la distorsion minimale par rapport aux vecteurs du dictionnaire dépasse le seuil de tolérance prédéfini, sont affectés au dictionnaire. Cette technique apporte une amélioration importante, mais elle apporte aussi une complexité accrue pour la restructuration des dictionnaires. En effet, la disposition des vecteurs dans le dictionnaire tient compte de leurs taux d'apparition pendant le codage. Les vecteurs les plus utilisés sont classés au début du dictionnaire de façon à éviter le parcours complet dudit dictionnaire.

V - B - 2 Quantification vectorielle dans l'espace image

Dans la famille des techniques hybrides nous nous proposons de détailler la méthode présentée par Y. Buhler dans [BUHL 86]. Les principaux avantages de cette technique résident dans l'utilisation d'un codage hiérarchique exploitant la transformée en cosinus discrète doublée d'une quantification vectorielle basée sur les particularités visuelles contenues dans l'image.

Classification visuelle

Le codage s'effectue sur plusieurs étapes. La première étant le fractionnement de l'image en blocs de tailles différentes, en se basant sur l'analyse du contenu de l'image. Ce découpage s'effectue à la suite d'une classification en fonction de l'activité de l'image. Il permet de former à partir d'images 256x256 des blocs de tailles 32x32, 16x16, 8x8 et 4x4 pixels. Le fractionnement et la restitution de l'image s'effectuent de gauche à droite et de haut en bas suivant l'algorithme "quadtree" qui détermine suivant la position et la taille d'un sous bloc donné la position et la taille du bloc suivant.

Une première répartition grossière détermine des caractéristiques

visuelles contenues dans l'image. Elle exploite les significations des histogrammes de luminance de plusieurs tailles de bloc. Une classification visuelle détermine les familles suivantes :

- Blocs de type 1; blocs uniformes ou quasi-uniformes, le contenu est composé essentiellement de plans formés de pixels dont la luminance varie peu. A la rencontre d'un bloc de ce type l'algorithme de fractionnement cesse la division pour obtenir des blocs de petites tailles. La distribution dans l'histogramme de l'ensemble des valeurs de luminance sur le nombre de pixels est restreinte. Cela se traduit dans l'histogramme par un pic de base étroite.
- Blocs de type 2; zones d'ombrages doux, elles renferment des plages de pixels dont la luminance varie spatialement d'une façon progressive. La disposition des pixels dans l'histogramme est caractéristique, elle est différente des dispositions correspondant aux autres classes visuelles. Pour le codage de ce type de bloc Buhler et al. appliquent la transformée en cosinus discrète TCD. La quantification des "fréquences basses" est suffisante pour la reconstitution du bloc image.
- Blocs de type 3 formés de textures fines. Ce sont des régions contenant beaucoup de détails. La classification par histogramme n'étant pas possible, car leurs caractéristiques sont semblables à celles des blocs uniformes ou quasi-uniformes. Une nouvelle caractéristique de répartition permet de différencier les deux familles. Elle quantifie les coefficients "hautes fréquences", obtenus après application de la TCD, à l'aide d'un paramètre *EHF* énergie hautes fréquences.
- Blocs de type 4 contenant des textures fortes. Les détails contenus dans ces zones sont pertinents et d'une grande importance visuelle. Ces zones sont codées par des blocs de taille 4x4 de façon à restituer leurs contenus.
- Bloc de type 5, ils représentent les contours nets. Ces parties de l'image sont indispensables à sa compréhension. Ces zones sont facilement reconnaissables par leurs histogrammes. Les amplitudes élevées et l'étendue spatiale importante sont autant de caractéristiques qui les différencient des autres structures visuelles. Pour leur codage, une grande fidélité est nécessaire. Un fractionnement maximal leur sera appliqué de façon à obtenir des blocs de 4x4. La quantification vectorielle adaptative permet de les coder à l'aide de dictionnaires appropriés.

Résumé

Le tableau présenté ci-dessous permet de comprendre la classification visuelle utilisée par Buhler [BUHL 86], ainsi que l'efficacité des paramètres caractéristiques qu'il exploite. En effet, les caractéristiques utilisées sont complémentaires. Dans le cas où deux familles se recouvrent partiellement et que l'étude de l'histogramme ne permet pas la détermination de l'appartenance d'un bloc image donné le recours au second paramètre *EHF* est dans ce cas indispensable.

Type de zone	Energie Hautes Fréquences	Histogramme des luminances
1	Très faible à faible	Plage étroite (pic)
2	Faible	Plage moyenne
3	Forte	Plage étroite
4	Forte	Plage très étendue
5	Assez forte à forte	Plage étendue

Tableau N°1 Identification des structures visuelles par les caractéristiques élaborées pour le codage.

Les caractéristique descriptives d'identification des familles visuelles sont les histogrammes, l'énergie des hautes fréquences et enfin les aspects visuels de l'image identifiés par l'étude des évolutions de luminance. Cette dernière caractéristique est utilisée pour les blocs 4x4 codés par quantification vectorielle avec adaptation visuelle.

L'histogramme des luminances est un outil pertinent pour l'analyse de l'image et de son traitement. Les histogrammes présentés ici sont normalisés par le nombre des pixels de la zone, de façon à permettre la comparaison de leurs signification au delà de la taille des blocs image correspondants. La valeur représentée dans les histogrammes pour une luminance donnée l sera notée $h(l)$ limitée par : $0 \leq h(l) \leq 1$.

Le critère de sélection *Act* extrait des histogrammes est déterminé par :

$$Act = - \frac{1}{M} \sum_{l=0}^{L-1} \log_2[h_n(l)].$$

$h_n(l)$ différent de 0.

$$h_n(l) = h(l) \cdot f_n(l - L) \quad .$$

Avec M désignant le nombre total des pixels de la zone. $h(l)$ désigne la valeur de l'histogramme pour la luminance l et $h_n(l)$ représente le nouvel histogramme pondéré par la fonction f_n de l'écart de la luminance l avec la luminance moyenne de la zone L . Afin que la contribution à l'activité Act augmente avec écart $(l - L)$, la fonction f_n a été choisie décroissante. Deux fonctions ont été testées par Buhler : l'une agit comme l'inverse du carré de l'écart et l'autre comme une fonction gaussienne du même écart.

Le second critère de sélection est fondé sur le calcul de l'énergie des coefficients hautes fréquences obtenus à la suite de l'application de la transformée en cosinus discrète :

$$EHF = 1000 \cdot \frac{\text{Energie des coefficients H.F}}{\text{Energie totale}}$$

L'énergie des coefficients H.F. est la somme des valeurs des coefficients du plan TCD rejetés lors du codage par transformée. L'énergie totale est la somme des coefficients transformés, ce qui inclut l'élément (0,0) qui correspond à la moyenne de la luminance des pixels du bloc.

Codage par transformée en cosinus discrète

Il est appliqué à chaque fois qu'un bloc de taille 16x16 ou 8x8 est déterminé comme moyennement actif. Sachant que pour cette catégorie de bloc image, la matrice transformée en TCD présente des coefficients pertinents dans la zone des basses fréquences; un codage scalaire de ces coefficients est effectué. Il exploite des matrices d'allocation de bits similaires à celles décrites dans le chapitre des techniques adaptatives de codage par transformation TCD.

Codage par quantification vectorielle

Le fractionnement appliqué à l'image permet l'identification des zones refermant des pixels dont l'évolution de luminance est particulière. Ce morcellement aboutit à l'identification de bloc 4x4 formés de détails pertinents; contours nets ou textures fortes. Ce type de blocs est codé d'une façon personnalisée par un quantificateur vectoriel. Ces zones renferment la plus grande partie de l'information visuelle contenue dans l'image et leur codage doit être d'une grande fidélité.

L'expérience a montré que l'utilisation d'un dictionnaire non adaptatif donne des performances médiocres pour le codage. En effet, il est nécessaire d'utiliser des dictionnaires de plusieurs milliers de vecteurs pour avoir finalement des images décodées d'une très médiocre qualité subjective.

Pour coder efficacement les images en utilisant la quantification vectorielle, il est essentiel d'exploiter l'adaptation visuelle. Introduite par Ramamurthi et Gersho [RAMA 84] cette approche algorithmique est utilisée ici par Buhler [BUHL 86].

Le quantificateur vectoriel manipule des vecteurs de dimension 16 correspondant aux blocs de pixels 4x4. Une présélection détermine l'activité du bloc image traité et permet de définir le contour qu'il renferme. Les directions des contours retenues sont les directions verticales, horizontales et diagonales. Pour les deux premières familles de contours 14 dictionnaires ou livres de codes leur sont réservés, pour la famille des transitions diagonales 17 livres de codes représentent les différentes possibilités d'évolutions de la luminance avec un gradient dirigé en diagonale. Deux autres dictionnaires sont réservés aux blocs uniformes ou d'activité moyenne.

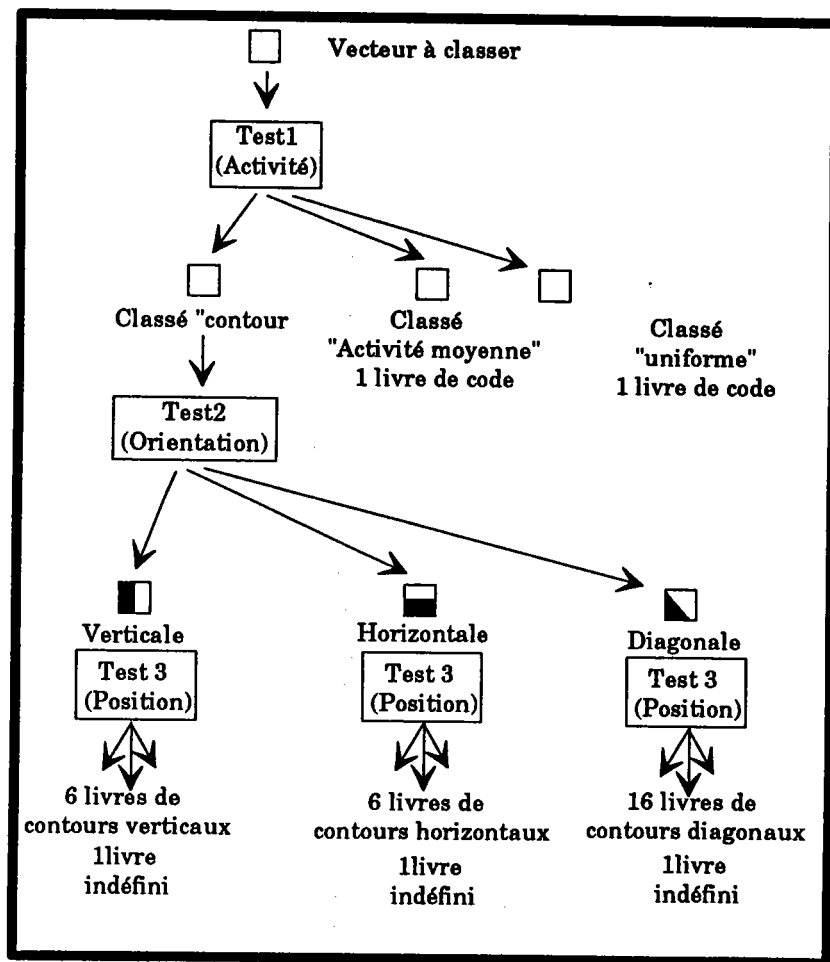


Figure N°27 arbre de décision employé pour la classification des vecteurs (33 livres de code au total).

L'algorithme de classement des blocs image est fondé sur le calcul de deux matrices de gradients $H(i,j)$ et $V(i,j)$ déterminant les directions d'évolution de la luminance horizontalement et verticalement. Les éléments de chaque vecteur ou bloc image sont notés $a(i,j)$. L'expression des deux gradients sont :

$$H(i,j) = \Phi(a(i+1,j) - a(i,j)) \quad ; i=1,..,3 \quad ; j=1,..,4.$$

$$V(i,j) = \Phi(a(i,j+1) - a(i,j)) \quad ; i=1,..,4 \quad ; j=1,..,3.$$

Avec $\Phi(x)$ fonction impaire de pondération. Sa valeur dépend de la hauteur de la transition. Elle est définie par le tableau suivant :

Hauteur de la transition x	$\Phi(x)$
$0 \leq x < 6$	0
$6 \leq x < 10$	1
$10 \leq x < 15$	2
$15 \leq x < 20$	4
$20 \leq x < 30$	6
$x > 30$	8

Tableau N°2 Valeurs de $\Phi(x)$ en fonction de la hauteur x de la transition.

Φ étant une fonction impaire, seules les valeurs positives sont données ici.

Pour déterminer la famille visuelle d'appartenance Buhler a exploité les gradients $H(i,j)$ et $V(i,j)$ pour différencier les évolutions de luminance dans chaque bloc image.

Des paramètres A_h A_v d'activité orientée sont calculés. leurs comparaison détermine la famille visuelle d'appartenance du bloc. Un arbre de décision discriminatoire réalise cette tâche.

$$A_h = \sum_{j=1}^4 \sum_{i=1}^3 |H(i,j)|$$

$$A_v = \sum_{j=1}^3 \sum_{i=1}^4 |V(i,j)|$$

Si $A_h A_v$ sont tous deux inférieurs à un seuil ϕ_1 , le bloc image est codé par le dictionnaire composé de blocs uniformes.

La somme des activités orientées détermine l'appartenance à la famille de blocs image d'activité moyenne. Si cette somme ne dépasse pas un seuil prédéfini ϕ_2 , le bloc incident est déclaré d'activité uniforme.

Si une des activités orientées $A_h A_v$ est supérieur au doublé de l'autre le bloc est considéré comme contenant un contour orienté, son codage est effectué par le livre de codes approprié.

Chaque dictionnaire réservé à des évolutions de luminance orientées renferme plusieurs sous dictionnaires. Chacun de ces livres de codes est adapté à la position de la transition contenue dans le bloc image. Dans le cas des dictionnaires de contours verticaux et horizontaux trois positions sont représentées de deux façons, une fois par des gradients positifs la seconde par des gradients négatifs. Le livre de codes correspondant aux évolution diagonales renferme des représentations de huit positions de contours.

Dans ce chapitre nous avons décrit quelques techniques exploitant les performances de la transformée en cosinus discrète et celles de la quantification vectorielle.

La méthode de codage que nous proposons est fondée sur ce même principe. Cependant, pour augmenter ses performances, nous nous sommes intéressés à la signification des coefficients transformés au point de vue visuelle.

Chapitre VI

CORRESPONDANCE IMAGE - ESPACE TRANSFORME

VI - A Introduction

Différentes méthodes de codage, citées précédemment, sont fondées sur un schéma de quantification vectorielle sur l'espace transformé. Cependant, elles se limitent à identifier des regroupements de coefficients transformés. Cette approche ne s'adapte pas aux structures visuelles contenues dans le bloc image initial. Les zones qui regroupent les coefficients les plus importants sont localisées par un ensemble de paramètres. A la suite de cette opération, un vecteur est formé suivant un parcours en zigzag des zones retenues.

L'inefficacité des méthodes de regroupement des coefficients classiques se traduit par la faible représentativité des dictionnaires utilisés pour la quantification vectorielle. Pour obtenir une bonne qualité subjective des images décodées malgré cette faiblesse, il est essentiel d'utiliser des dictionnaires de grande taille. Cela a des conséquences directes sur les performances du codage, spécialement sur le taux de compression.

la méthode de codage décrite dans ce mémoire, est essentiellement le résultat d'investigations approfondies pour identifier les correspondances entre l'espace image et l'espace transformé. A partir de ces constatations, nous avons élaboré une méthode de codage dont la classification est dite visuelle.

VI - B Correspondance pour les transformées mathématiques

Les transformations orthogonales, Karhunen-Loève, Hadamard et la transformée en cosinus discrète, ont des coefficients transformés comparables. Les correspondances entre l'espace image et l'espace transformé sont, à quelques particularités près, semblables.

Cette correspondance est illustrée par l'exemple suivant; soit le médaillon d'image (2 x 2).

$$\begin{bmatrix} 40 & 4 \\ 44 & 12 \end{bmatrix} = 25 \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} - 3 \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix} + 17 \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix} + 1 \begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix}$$

Figure N°28 Décomposition en évolutions primaires par la transformée de Hadamard

$$f = F_{00} \times B_{00} + F_{01} \times B_{01} + F_{10} \times B_{10} + F_{11} \times B_{11}$$

Soit un bloc image formé par quatre pixels, l'application de la transformée d'Hadamard décompose ce bloc en une composition de médaillons d'images, matérialisant des évolutions primaires de luminances B_{ij} , indexées respectivement par des coefficients obtenus à la suite de la transformation mathématique F_{ij} .

Nous remarquerons que le coefficient F_{00} représente la "composante continue". On remarquera aussi que le coefficient le plus important est F_{10} , le bloc image analysé est par conséquent dominé par la transition (B_{10}) [PONC 74]. Pour la transformée en cosinus discrète une représentation similaire est décrite dans la figure suivante.

VI - B Espace Transformé en Cosinus Discrète

La transformée en cosinus discrète est une approximation mathématique de la transformée de Fourier. Les coefficients transformés expriment la répartition fréquentielle dans le plan transformé de l'amplitude de la luminance des pixels.

La notion de fréquence utilisée ici n'est pas celle qui est associée à une évolution temporelle mais à une évolution spatiale. Les variations

progressives de niveaux de gris ou de luminances de pixels sont associées aux basses fréquences, alors que les variations brutales correspondent aux hautes fréquences. Autrement dit, une évolution lente et continue des luminances correspond à une variation spatiale lissée dans le domaine des basses fréquences ; à l'inverse, une texture grossière caractérisée par des variations brutales de luminances offre un spectre renforcé dans le domaine des hautes fréquences.

La représentation transformée en cosinus discrète est une matrice de coefficients. Chaque coefficient représente une transition primaire ou une composition de transitions. L'image correspondant à cette représentation est composée de toutes ces variations accentuées plus ou moins par les amplitudes des coefficients qui leur sont associés (cf. Fig. N°29).

La représentation symbolique de la figure N°29 met en évidence la signification visuelle de chaque coefficient transformé. Une matrice transformée où un coefficient $F(i,j)$ est très important en valeur absolue par rapport aux autres coefficients, sera associé à un bloc image ressemblant à l'imagette(i,j). Cette ressemblance traduit la tendance d'évolution de la luminance des pixels dans le bloc image.

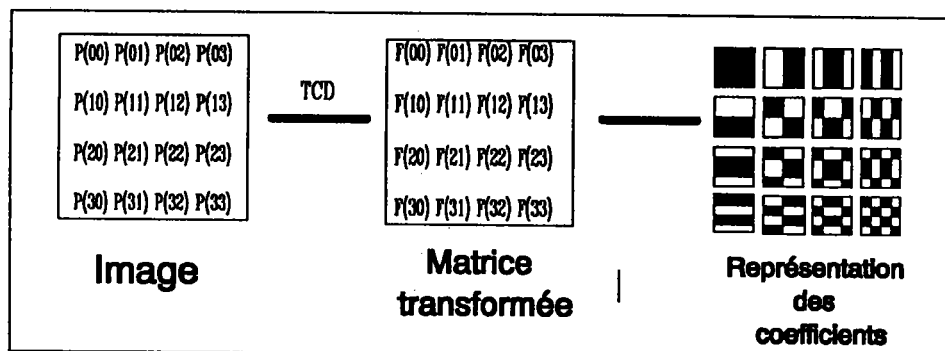


Figure N°29 Transformée en cosinus discrète

VI - B - 1 Blocs homogènes

Une des particularités de la TCD est de disposer des fréquences spatiales étroitement liées aux évolutions des luminances des pixels qui forment le bloc image analysé.

Les coefficients $F(u,v)$, avec (u,v) non nul, représentent les composantes alternatives. Les composantes basses fréquences correspondent aux coefficients de faibles indices , faible rang sur la matrice $(u+v)$. Les composantes de hautes fréquences sont identifiés par les coefficients d'indices élevés. Nous savons que les variations progressives de luminance des pixels sont associées aux basses fréquences et les variations brutales aux hautes fréquences.

Le coefficient $F(0,0)$ représente la "composante continue", il symbolise la moyenne.

$$F(0,0) = \frac{2}{N^2} \sum_{x=0}^7 \sum_{y=0}^7 P(x,y) = 2 \cdot M$$

M moyenne du bloc image. $P(x,y)$ valeur de la luminance au pixel (x,y) du bloc image.

Inversement, des expérimentations ont mis en évidence que les blocs homogènes, ceux dont les luminances des pixels varient peu autour de leur moyenne, ne sont représentés dans le domaine transformé que par le coefficient $F(0,0)$, les autres coefficients $F(u,v)$, avec (u,v) non nul, étant négligeables.

VI - B - 2 Transitions primaires

Les correspondances visuelles considérées dans cette catégorie concernent les évolutions horizontales et verticales de la luminance dans le bloc d'image. Ces transitions sont exclusivement matérialisées dans le domaine transformé par la première colonne et la première ligne de la matrice de coefficients.

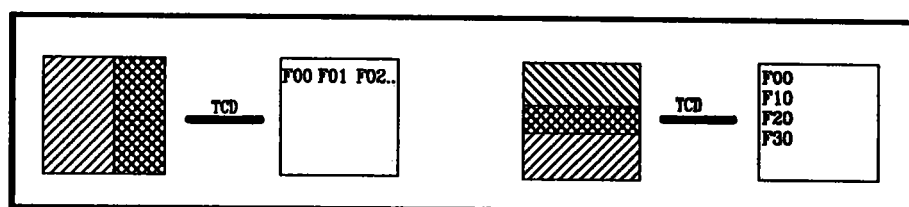


Figure N°30 Transitions primaires orientées

VI - B - 3 Manipulations du bloc image

Il s'agit ici de considérer les transformations intervenant au niveau des positions des pixels dans le bloc image, ou au niveau des valeurs de leurs luminances. Elles se représentent souvent par une inversion de signe des coefficients dans le bloc transformé correspondant.

Transformation qui donne à un pixel la valeur complémentaire à 255 de sa luminance

Cette transformation donne à chaque luminance du bloc image, la valeur complémentaire à 255:

$$P_1(x,y) = 255 - P_0(x,y)$$

$$F_1(u,v) = 510 - F_0(0,0) \text{ si } u=v=0; \\ - F_0(u,v) \text{ sinon}$$

Nous constatons que la composante continue traduit la nouvelle moyenne du bloc image. La manipulation décrite ici inverse les signes des composantes fréquentielles alternatives.

Symétrie par rapport à la médiane verticale

Cette transformation fait correspondre à un pixel dans un bloc image la valeur de luminance du pixel symétrique par rapport à la première médiane du bloc image.

$$P_1(x,y) = \text{Symétrique de } (P_0(x,y)) .$$

$$F_1(0,0) = F_0(0,0)$$

$$F_1(u,2p+1) = - F_0(u, 2p+1) \text{ colonnes impaires.}$$

$$F_1(u,2p) = F_0(u, 2p) \text{ colonnes paires.}$$

Dans ce cas de figure la composante continue $F(0,0)$ est invariable, les fréquences alternatives $F(u,2v)$ dont les indices des colonnes sont paires sont inchangées aussi, alors que les fréquences alternatives $F(u,2v +1)$ dont les indices des colonnes sont impairs ont leur signe inversé.

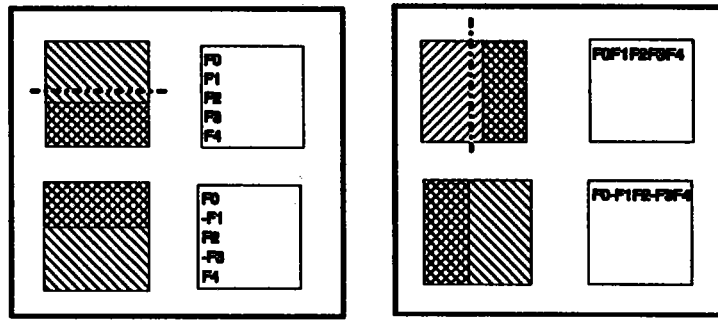


Figure N°31 Symétries par rapport aux médianes horizontale et verticale

Symétries par rapport à la médiane horizontale

Cette transformation fait correspondre à un pixel dans un bloc image, la valeur de luminance du pixel symétrique par rapport à la seconde médiane du bloc image.

$$P_1(x,y) = \text{Symétrique de } P_0(x,y).$$

$$F_1(0,0) = F_0(0,0)$$

$$F_1(2p+1,u) = -F_0(2p+1,u) \text{ lignes impaires.}$$

$$F_1(2p,u) = F_0(2p,u) \text{ lignes paires.}$$

Nous remarquons dans ce cas que la composante continue $F(0,0)$ et les fréquences alternatives $F(2u,v)$ dont les indices des lignes sont pairs restent inchangées, alors que les fréquences alternatives $F(2u+1,v)$ dont les indices des lignes sont impairs ont leur signe inversé.

Symétrie par rapport à la première diagonale du bloc image

Il s'agit ici de la manipulation qui fait correspondre à un pixel, la valeur de luminance de son symétrique par rapport à la première diagonale du bloc image. Dans le bloc transformé, cette symétrie provoque une interversion symétrique par rapport à la même diagonale.

$$P_1(x,y) = \text{Symétrique de } P_0(x,y).$$

$$F_1(u,v) = F_0(v,u)$$

Les coefficients de la transformée de $P_1(x,u)$ sont symétriques de ceux de $P_0(x,u)$ par rapport à la première diagonale.

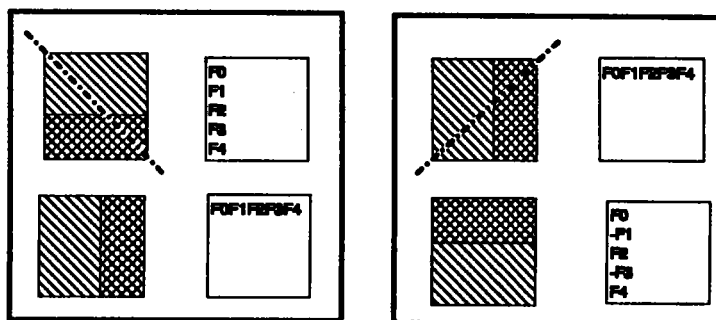


Figure N°32 Symétries par rapport aux deux diagonales

Symétrie par rapport à la seconde diagonale

Pour la transformation qui fait correspondre à un pixel dans un bloc image la valeur de luminance du pixel symétrique par rapport à la seconde diagonale du bloc image, l'effet sur le bloc transformé est différent de celui provoqué par la manipulation précédente. Elle intervertit les composantes fréquentielles alternatives $F(u,v)$ en inversant les signes pour les coefficients dont les coordonnées (u,v) sont de parités opposées. Les composantes $F(u,v)$ dont les coordonnées sont de même parité ont leurs signes maintenus, malgré l'intervention.

$$P_1(x,y) = \text{Symétrique de } P_0(x,y).$$

$$F_1(u,v) = \begin{cases} F_0(v,u) & \text{si } u \text{ et } v \text{ sont de même parité.} \\ -F_0(v,u) & \text{sinon.} \end{cases}$$

Translation des luminances d'un bloc image

Le décalage des luminances d'un bloc image d'une valeur G_0 se répercute seulement sur la composante continue. Les autres coefficients restent inchangés.

$$P_1(x,y) = P_0(x,y) + G_0$$

$$F_1(0,0) = F_0(0,0) + (2G_0).$$

$$F_1(u,v) = F_0(u,v).$$

La mise en évidence des caractéristiques dans le domaine transformé, dues à des actions que nous avons provoquées ou dont nous avons constaté les effets, nous a amené à proposer une nouvelle classification visuelle. Cette classification ne se limite pas à localiser les zones de coefficients importants, mais elle permet d'identifier la ou les formes matérialisées par le bloc image. La reconstitution du bloc après le codage sera d'autant plus fidèle que l'identification des structures dans le bloc incident est efficace.

Chapitre VII

CODAGE PAR QUANTIFICATION VECTORIELLE DANS L'ESPACE TRANSFORME AVEC CLASSIFICATION VISUELLE

Le procédé de codage d'image, que nous avons développé, est fondé sur une transformation en cosinus discrète à la suite de laquelle nous appliquons une quantification vectorielle. Le passage dans le domaine transformé permet le codage des coefficients transformés représentant le spectre énergétique de l'image. Les coefficients pertinents sont regroupés en par une classification. Cette dernière se traduit par une identification des composantes visuelles de l'image dans l'espace transformé. Les structures visuelles sont reconnues en tenant compte des constatations sur les correspondances entre l'espace image et l'espace transformé.

VII - A Classification des blocs image

La classification consiste en un ensemble de traitements appliqués au bloc image pour identifier sa famille. Nous nous proposons de décrire une méthode de reconnaissance des structures visuelles à partir d'un ensemble de caractéristiques déterminées dans l'espace transformé. L'identification des composantes du bloc image permet de le coder dans les meilleures conditions de compression et de qualité subjective après décompression de l'image.

VII - A - 1 Paramètres de classification

Activité spatiale totale

Nous avons défini un paramètre T pour définir l'activité spatiale. Ce paramètre quantifie l'importance des différentes évolutions que subissent les amplitudes des luminances dans le bloc image. Le paramètre T est donné par la relation suivante :

$$T = \left(\sum_{u=0}^7 \sum_{v=0}^7 (F(u,v))^2 \right) - (F(0,0))^2$$

Où $F(u,v)$ est un coefficient de la matrice transformée TCD, (u,v) sont ses coordonnées dans ce bloc. $F(0,0)$, le premier coefficient du bloc, ne représente pas les variations des luminances du bloc image, mais sa composante continue.

Une première répartition des blocs transformés est possible à l'aide du paramètre T . A partir de seuils déterminant les limites, l'activité spatiale différencie quatre classes grossières :

- blocs homogènes : sont reconnus uniformes tous les blocs dont la luminance des pixels varie peu ou pas autour de sa moyenne,
- blocs de faible activité : cette catégorie recouvre tous les blocs présentant de variations à peine perceptibles,
- blocs d'activité moyenne : composés de transitions de luminance peu contrastées,
- blocs d'activité forte : les blocs composés de structures dont la luminance sont contrastées appartiennent à cette famille. Ils représentent le groupe le plus sensible aux erreurs à la restitution des images. En effet, l'oeil est automatiquement attiré par les déformations sur cette catégorie de blocs image.

Activité partielle matérialisant les transitions verticales

Cette activité est identifiée par la comparaison de deux paramètres. Le premier H regroupe toutes les fréquences spatiales de la première ligne élevés au carré, sauf le coefficient représentant la moyenne. Le second N_h est le complémentaire du premier paramètre par rapport au paramètre T . Rappelons que T représente l'activité fréquentielle totale.

Le paramètre $\frac{N_h}{H}$ de cette activité partielle découle directement de la constatation suivante : les transitions verticales sur le bloc transformée sont représentées exclusivement par l'horizontale du bloc transformé.

$$H = \sum_{v=1}^7 (F(0,v))^2$$

$$N_h = \sum_{u=1}^7 \sum_{v=0}^7 (F(u,v))^2 = T - H$$

Activité partielle correspondant aux transitions horizontales

Le paramètre $\frac{N_v}{V}$ de cette activité discrimine les blocs d'image ayant des transitions horizontales de luminance. L'identification est effectuée dans l'espace transformé. Nous l'obtenons par la comparaison de la somme des coefficients au carré de la première colonne avec la somme des coefficients, n'appartenant pas à la colonne, élevés au carré.

$$V = \sum_{u=1}^7 (F(u,0))^2$$

$$N_v = \sum_{u=0}^7 \sum_{v=1}^7 (F(u,v))^2 = T - V$$

Bloc présentant une symétrie par rapport à la première diagonale

D'après les correspondances décrites précédemment si deux blocs sont symétriques par rapport à la première diagonale, leurs coefficients transformés sont aussi symétriques. Nous en concluons que si un bloc est symétrique à lui même par rapport à sa première diagonale, sa matrice transformée l'est aussi.

Pour identifier cette famille visuelle, un paramètre a été formé. En effet, le paramètre D_1 cumule les écarts entre chaque fréquence spatiale et sa symétrique par rapport à la première diagonale.

$$D_1 = \sum_{u=0}^7 \sum_{v=0}^7 [F(u,v) - F(v,u)]^2.$$

Bloc image symétrique par rapport à la deuxième diagonale

Nous avons exploité la propriété concernant la symétrie par rapport à cette diagonale. Nous remarquerons qu'un bloc auquel s'applique cette caractéristique, est identifiable à l'aide d'un paramètre D_2 .

Ce paramètre D_2 découle directement de la propriété décrite ultérieurement. D_2 tient compte de deux facteurs. Le premier correspond aux écarts entre chaque fréquence spatiale et sa symétrique par rapport à la première diagonale quand leurs coordonnées - ligne, colonne - sont de même parité. Le second représente les écarts entre les fréquences spatiales dont les coordonnées sont de parités différentes et l'inverse des fréquences qui leur sont symétriques par rapport à la première diagonale.

$$\begin{aligned} D_2 = & \sum_{p=0}^3 \sum_{q=0}^3 [F(2p,2q) - F(2q,2p)]^2 \\ & + \sum_{p=0}^3 \sum_{q=0}^3 [F(2p+1,2q+1) - F(2q+1,2p+1)]^2 \\ & + \sum_{p=0}^3 \sum_{q=0}^3 [F(2p,2q+1) + F(2q+1,2p)]^2 \end{aligned}$$

Bloc image ayant une symétrie par rapport à son centre

Cette famille de blocs est identifiée dans le domaine transformé par une grande amplitude des valeurs appartenant à la diagonale autres que celles matérialisant la moyenne $F(0,0)$. Pour identifier ces blocs, le paramètre D a été élaboré.

$$D = \sum_{u=0}^7 [F(u,u)]^2$$

VII - B Arbre de décision

L'appartenance des blocs image aux familles visuelles est déterminée dans le domaine transformé par un vecteur caractéristique. Il est donc essentiel d'appliquer la transformée en cosinus discrète à chaque bloc image (8×8).

Pour chaque bloc transformé un vecteur identification est constitué par les composantes suivantes $(T, \frac{N_h}{V}, \frac{N_v}{V}, D_1, D_2, D)$.

Sur la base de cette caractéristique jugée discriminante dans l'espace fréquentiel, la répartition en termes de classes visuelles est effectuée.

Des expérimentations ont permis une répartition de l'espace des blocs transformés. Une sélection exploitant les structures visuelles est à la base de ce regroupement. Le classement est déterminé par l'arbre de décision. Il permet de sélectionner chacune des branches de l'arbre et de déterminer l'appartenance du bloc à coder à une famille visuelle particulière. Les seuils de sélection ont été élaborés de façon à s'adapter aux différentes classes.

VII - B - 1 Classes d'activité spatiale

Il s'agit donc de coder chaque bloc d'image en fonction de la nature du bloc et de ses caractéristiques visuelles. Le paramètre T détermine la classe d'appartenance "activité spatiale" d'un bloc transformé donné. Une comparaison discriminatoire de T à un jeu de valeurs de seuil est effectuée.

Les blocs homogènes sont identifiés, puisqu'ils présentent une activité spatiale négligeable, du fait de l'absence d'évolutions remarquables des luminances. Un seuil maximal ST_0 délimite l'espace des blocs uniformes. Ce seuil a été choisi pour englober dans cette classe le maximum de blocs, sans pour autant générer *l'effet de bloc*, qui se manifeste par une mosaïque de luminances sur l'image reconstituée.

L'oeil humain est automatiquement attiré par les zones de l'image dont l'évolution spatiale est importante. En effet, notre méthode de classification partitionne l'espace des blocs transformés en quatre groupes suivant le paramètre activité spatiale: blocs uniformes, blocs faiblement actifs, moyennement actifs et blocs fortement actifs.

La sélection est déterminée en fonction de l'appartenance des blocs à chacune des classes décrites. Le seuillage est très brutal pour les blocs à faible activité, le codage est par conséquent grossier sans provoquer une dégradation visible. Cette discrimination est élaborée dans le but d'englober un maximum de blocs dans les classes de faible activité. En effet, pour ces blocs, un minimum d'information est nécessaire pour représenter le bloc image incident.

Cependant, pour les blocs à forte activité spatiale, le seuillage doit être délicat car la moindre erreur ou le moindre écart par rapport à l'image de départ est aisément visible par l'observateur. Pour cette classe, la méthode de codage décrite ici utilise des seuils fins pour rester la plus fidèle possible aux images initiales. Cette dernière catégorie nécessite pour sa représentation un maximum de fréquences spatiales.

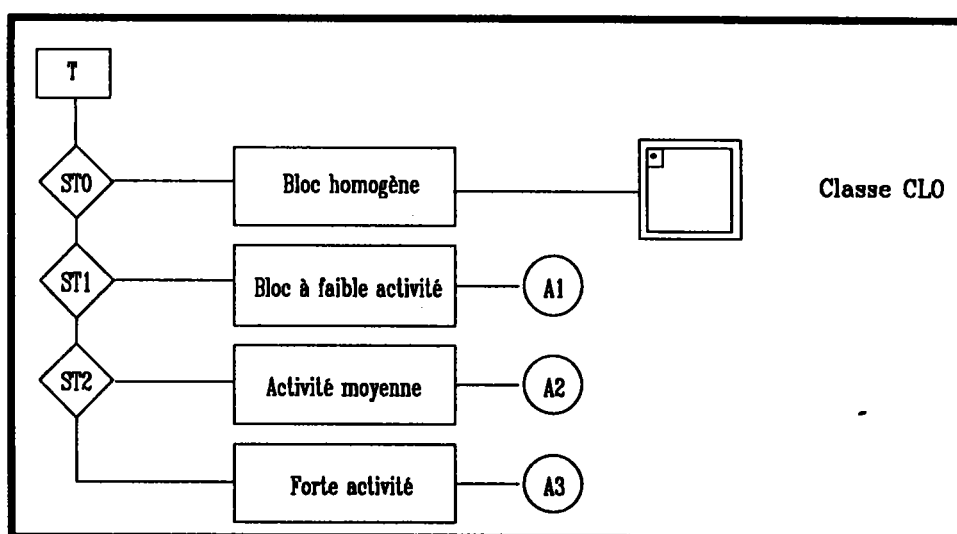


Figure N°33 Sélection des familles d'activité spatiale

Selon notre méthode, nous effectuons une pré-classification suivant le facteur activité spatiale T ainsi que les différents seuils qui lui sont indexés ST_0 , ST_1 et ST_2 . Un premier partitionnement est réalisé sur quatre classes (cf. Fig. N°33).

- CL_0 classe des blocs homogènes : $0 < T < ST_0$.
- CL_1 classe des blocs faiblement actifs :
 $ST_0 < T < ST_1$.
- CL_2 classe des blocs moyennement actifs :
 $ST_1 < T < ST_2$
- CL_3 classe des blocs très actifs : $T > ST_2$

CL_1 , CL_2 et CL_3 se décomposent respectivement, en cinq sous-classes. Ces dernières représentent les particularités visuelles contenues dans l'image.

VII - A - 2 Classes de composantes visuelles particulières

A partir de chaque classe d'activité spatiale, nous déterminons des familles de blocs vérifiant des critères visuels. Cette sélection est rendue possible de part la définition des paramètres élaborés. L'indice i identifie la classe d'activité, il prend les valeurs 1, 2 ou 3.

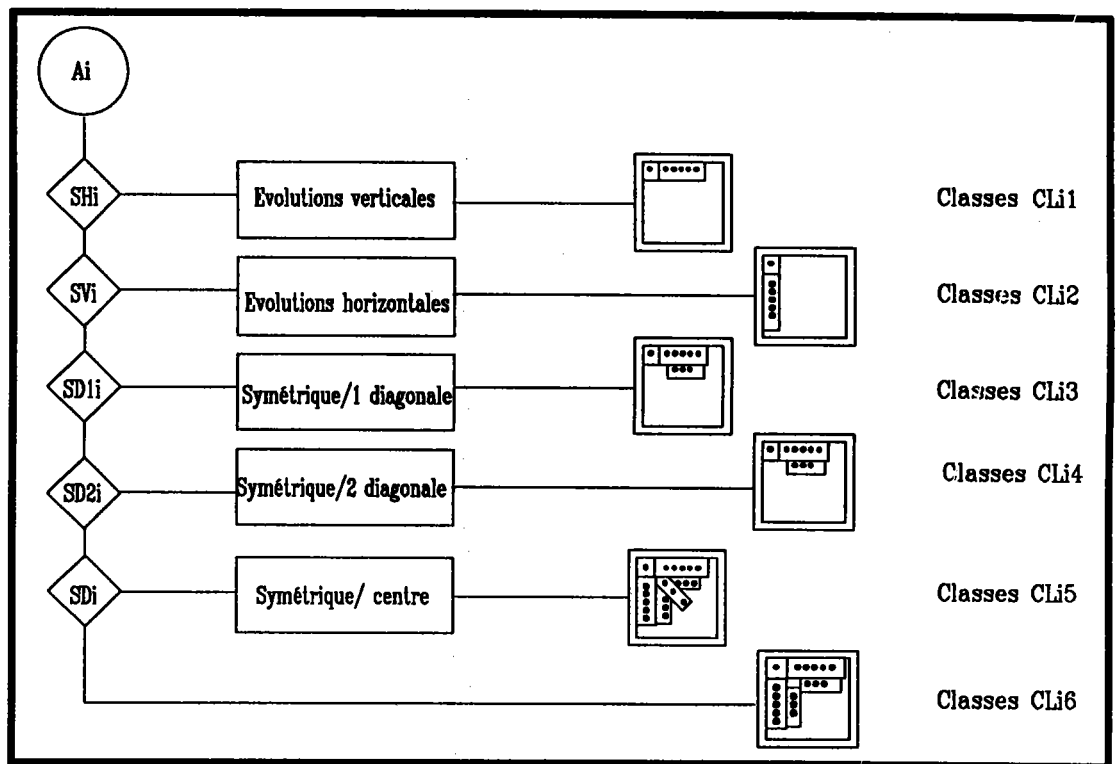


Figure N°34 Arbre de décision pour la classification visuelle

- CL_{i1} classe des blocs image visualisant des transitions verticales, déterminée par (NH/H) .
- CL_{i2} classe des blocs image visualisant des transitions horizontales, déterminée par (NV/V) .
- CL_{i3} classe des blocs image ayant une symétrie par rapport à la première diagonale, déterminée par $(D1)$.
- CL_{i4} classe des blocs image ayant une symétrie par rapport à la deuxième diagonale, déterminée par $(D2)$.
- CL_{i5} classe des blocs image ayant une symétrie par rapport au point central, déterminée par (D) .
- CL_{i6} classe des blocs image ne vérifiant aucune des propriétés visuelles précitées.

Nous avons déterminé ainsi 19 classes visuelles. En fonction de l'appartenance de chaque bloc transformé nous effectuons la vectorisation correspondante (cf. Fig. N°34).

A la suite de la classification fondée sur le paramètre T , des groupes d'activité spatiale sont établis. Le vecteur caractéristique $(\frac{N_h}{H}, \frac{N_v}{V}, D_1, D_2, D)$ est comparé à un vecteur de seuils délimitant les classes visuelles. Chaque composante de ce vecteur est liée au paramètre correspondant du vecteur caractéristique. La mise en forme de ce vecteur tient compte des correspondances *espace image - espace transformé* précitées.

A chaque bloc, classé dans une famille d'activité spatiale i , est appliqué un vecteur seuil discriminant $(SH_i, SV_i, SD1_i, SD2_i, SD_i)$. De cette manière, sa classe visuelle est révélée. Les composantes du vecteur caractéristique sont comparées successivement aux seuils prédéfinis. Ces seuils sont dit maximaux, car si le bloc transformé présente une composante inférieure à la valeur de son seuil correspondant, le bloc image est déclaré appartenant à la famille visuelle mise en valeur par le paramètre.

Rappelons que $\frac{N_h}{H}$ et $\frac{N_v}{V}$ correspondent respectivement à l'activité relative aux transitions verticales et horizontales. D_1 est utilisé pour identifier les blocs ayant une symétrie par rapport à la première diagonale, D_2 l'est pour les blocs ayant une symétrie par rapport à la seconde diagonale. D permet de reconnaître les blocs symétriques par rapport à leur centre.

Les composantes du vecteur seuil de décision ont été ajustés manuellement, de façon à avoir un minimum de dégradation sur les images codées. Elles ont été élaborées à la suite de tests sur des images de référence.

Le choix du vecteur seuil est déterminant pour les performances du codage. Il agit directement sur le taux de compression des images et sur leur qualité subjective à la réception. L'élaboration des seuils est donc un compromis entre qualité et taux de réduction de l'information. Suivant l'application, nous avons la possibilité de privilégier une des caractéristiques aux dépens de l'autre.

VII - A - 3 Formation de vecteurs de coefficients transformés

L'image numérisée est partitionnée en blocs (8x8). Sur un bloc pixels $P(x,y)$, nous appliquons une transformée en cosinus discrète pour obtenir un bloc de (8x8) coefficients $F(u,v)$. La représentativité des blocs transformés est concentrée autour du coefficient $F(0,0)$.

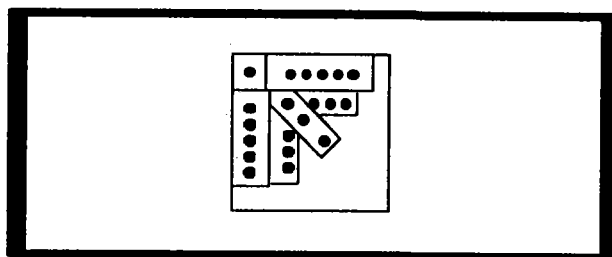


Figure N°35 Filtrage appliqué au bloc transformé avant la classification

Nous effectuons un filtrage passe bas de manière à ne considérer que les coefficients pertinents de basses fréquences. Le fait de négliger les autres fréquences n'altère pas l'image, car seules les textures étendues sur des plages importantes sont matérialisées par les hautes fréquences (cf. série d'images N°11 : annexes).

La vectorisation des coefficients transformés a été conçue en tenant compte de leur importance et de leur signification visuelle. A la suite de la classification, nous effectuons l'identification des coefficients pertinents au codage.

Nous nous sommes limités à un nombre réduit de coefficients car l'oeil de l'observateur se comporte comme un filtre passe bas. L'observateur a une vision globale et lissée de l'image.

Une des particularités de notre technique de codage est de retenir un ensemble de coefficients d'un côté et de l'autre de la diagonale du bloc transformé. Une autre est de former des vecteurs à partir de ces coefficients. Ce choix a été effectué, sachant que les coefficients à gauche de la diagonale jouent pour les structures horizontales le même rôle que celui joué par les autres coefficients pour les structures verticales.

Un des avantages de cette approche algorithmique est d'augmenter la représentativité du dictionnaire. Cet avantage se traduit par la diminution de la taille du dictionnaire exploité, tout en maintenant des performances de qualité subjective et de taux de compression.

En tenant compte des caractéristiques visuelles des blocs, nous identifions les coefficients les plus significatifs et nous effectuons le codage ad-hoc. La comparaison du vecteur caractéristique aux seuils pré-sélectionnés détermine l'ensemble des informations pertinentes par rapport au codage du bloc image.

- Classe CL_0 des blocs homogènes. Seul le coefficient $F(0,0)$ représentant la composante fréquentielle continue est pris en compte. Ce coefficient particulier est codé par quantification scalaire comme décrit ci-après.
- Pour les classes visuelles CL_{i1} des blocs d'image à transition verticale : la composante fréquentielle continue $F(0,0)$ est retenue ainsi que le vecteur composé des coefficients $(F(0,1); F(0,2); F(0,3); F(0,4); F(0,5))$ qui sont les composantes fréquentielles alternatives de la première ligne du bloc transformé à droite de sa première diagonale.
- Pour les classes visuelles CL_{i2} des blocs d'image à transitions horizontales nous retenons la composante continue $F(0,0)$ ainsi que le vecteur composé des coefficients $(F(1,0); F(2,0); F(3,0); F(4,0); F(5,0))$. Ils représentent respectivement les composantes disposées sur la première colonne du bloc transformé à gauche de sa diagonale.
- Pour les classes visuelles CL_{i3}, CL_{i4} des blocs d'image à symétrie par rapport à la première diagonale ou à la seconde, la composante fréquentielle continue $F(0,0)$ est retenue ainsi que les vecteurs composés des coefficients transformés : $(F(0,1); F(0,2); F(0,3); F(0,4); F(0,5))$ et $(F(1,2); F(1,3); F(1,4))$ qui sont respectivement les composantes fréquentielles alternatives de la première et la seconde ligne du bloc transformé et à la droite de sa première diagonale.

- Pour les classes visuelles CL_{i5} des blocs d'image à symétrie par rapport à un point central : nous retenons la composante continue $F(0,0)$ qui représente la moyenne, ainsi que les vecteurs: $(F(0,1); F(0,2); F(0,3); F(0,4); F(0,5)); (F(1,2); F(1,3); F(1,4)); (F(1,0); F(2,0); F(3,0); F(4,0); F(5,0)); (F(2,1); F(3,1); F(4,1))$ et $(F(1,1); F(2,2); F(3,3))$ qui sont respectivement les composantes fréquentielles alternatives de la première et la seconde lignes, première et seconde colonnes à droite et à gauche respectivement de la première diagonale du bloc transformé. Le dernier vecteur représente les coefficients transformés de la diagonale du bloc.
- Pour les classes CL_{i6} des blocs n'ayant aucune particularité visuelle particulière est retenue la composante continue $F(0,0)$ ainsi que les vecteurs composé des coefficients transformés : $(F(0,1); F(0,2); F(0,3); F(0,4); F(0,5)); (F(1,2); F(1,3); F(1,4)); (F(1,0); F(2,0); F(3,0); F(4,0); F(5,0))$ et $(F(2,1); F(3,1); F(4,1))$ qui sont respectivement les composantes fréquentielles alternatives du bloc transformé n'appartenant pas à sa diagonale et retenus après le filtrage passe bas appliqué à l'espace transformé.

Les dimensions de ces vecteurs ne sont pas définitivement fixées. Elles sont liées au gabarit du filtre appliqué avant la formation des vecteurs. La taille peut être augmentée si une bonne qualité subjective des images est requise.

La taille des vecteurs est adaptable dynamiquement dans le codage, suivant les besoins en performances instantanées. Ces cas de figure seront traités ultérieurement.

VII - C Quantification vectorielle des coefficients transformés

La technique de codage que nous proposons est une nouvelle approche de la quantification vectorielle dans le domaine transformé. Elle exploite les significations visuelles des coefficients par une classification adaptée des blocs transformés. Elle met aussi en pratique la décorrélation provoquée par la séparation de la composante continue des autres composantes de la représentation spectrale, obtenue par la transformée en cosinus discrète.

Le premier coefficient $F(0,0)$ de la matrice transformée est un multiple de la moyenne des luminances des pixels, ce coefficient est codé séparément par un codeur prédictif différentielle. Le prédicteur étant la moyenne entre les deux plus proches blocs voisins. Les autres coefficients représentent les différentes transitions primaires ou complexes composant l'image.

Pour appliquer la quantification vectorielle, nous avons effectué une classification. Cette opération a permis de localiser les coefficients transformés pertinents et d'identifier leurs significations visuelles. Par la suite, nous avons retenu les vecteurs de fréquences spatiales pertinentes.

La quantification vectorielle est fondée sur une recherche du vecteur le plus proche du vecteur à coder à l'aide d'un dictionnaire pré-établi. L'opération de codage consiste en un étiquetage du vecteur incident. En effet, nous remplaçons le vecteur en entrée du codeur par l'adresse de son représentant le plus proche appartenant au dictionnaire. A la réception, les adresses sont déchiffrées et les vecteurs correspondants sont restitués pour former l'image initiale.

Le codage d'un bloc transformé par quantification vectorielle, consiste à affecter aux vecteurs de coefficients retenus qui le composent les adresses des vecteurs qui leur sont les plus proches dans le dictionnaire. Le décodage réside dans la restitution et la disposition des vecteurs de coefficients dans une matrice transformée. Ces vecteurs regroupés vont générer, après une transformée en cosinus discrète inverse, un bloc étroitement dépendant des amplitudes et des signes des coefficients transformés codés.

La méthode de compression d'images que nous avons proposée est une opération fondée sur la reconnaissance et la reconstitution des évolutions des luminances dans une image. Nous proposons pour le codage de mimer les structures composant l'image, en quantifiant vectoriellement des regroupements de coefficients transformés représentant les variations de luminance.

La technique présentée ici tire ses principaux avantages essentiellement des propriétés visuelles constatées. Cela est valable aussi pour l'élaboration de la distance vectorielle.

VII - C - 1 Distorsion entre deux vecteurs

Dans leurs travaux, Marescq et al [MARE 87] et Nasrabadi et al

[NASR 84] reprennent la méthode classique de calcul d'écart entre deux vecteurs entre deux vecteurs $U(u_i)$ et $V(v_i)$. Cette distorsion selon les auteurs, est celle qui s'approche le mieux des décisions subjectives d'un observateur.

Les distorsions les plus couramment utilisées sont des variantes de la suivante :

$$\text{Dis} = \sum_{i=0}^{N-1} [u_i - v_i]^2$$

ou distorsion en valeur absolue.

$$\text{Dis} = \sum_{i=0}^{N-1} |u_i - v_i|$$

Soit un dictionnaire composé par M vecteurs V_j de dimension N . Soit le vecteur U qui a pour composantes u_i , U étant aussi de dimension N .

La quantification vectorielle du vecteur $U(u_i)$ est désignée par le vecteur $V_j(v_i)$ qui, de tous les M vecteurs $V_j(v_i)$ du dictionnaire, présente la distorsion minimale par rapport à $U(u_i)$:

$$QV(U) = j.$$

La distorsion entre deux vecteurs, selon la méthode proposée ici, est une combinaison de plusieurs distorsions. Elle découle de trois propriétés visuelles sur les blocs d'image et de leurs conséquences sur les fréquences spatiales dans le domaine transformé.

- Les symétries par rapport à un axe central ou par rapport à la médiane (verticale ou horizontale) d'un bloc image provoquent l'apparition d'un signe moins sur les composantes des lignes respectivement des colonnes paires de la matrice transformée.
- Le complémentaire à 255 d'un bloc image a comme composantes fréquentielles alternatives des coefficients qui sont identiques à ceux du bloc image initial, affectés d'un signe moins.
- Une inversion du signe des composantes des lignes ou colonnes impaires est provoquée par une symétrie du complémentaire du bloc image par

rapport à un des axes centraux principaux.

La disposition des vecteurs des deux côtés de la diagonale du bloc transformé nous permet d'exploiter toutes les propriétés visuelles. Un vecteur peut, s'il est disposé sur la ligne à droite de la diagonale, représenter une transition verticale. Il peut, s'il est disposé sur la colonne à gauche de la diagonale, correspondre à une évolution horizontale.

Ce n'est cependant pas le cas pour le vecteur matérialisant la diagonale ; ses composantes ne sont pas affectées par les symétries par rapport aux axes principaux.

L'écart entre deux vecteurs utilisé ici, détermine dans le dictionnaire le vecteur $V_k(v_i)$ le plus proche du vecteur $U(u_i)$ composé de fréquences spatiales à coder.

Dis = Minimum[Dis₁(U(u_i),V_k(v_i)); Dis₂(U(u_i),V_k(v_i)); Dis₃(U(u_i),V_k(v_i));
Dis₄(U(u_i),V_k(v_i))]
avec k variant de 0 à M

$$\text{Dis} = \text{Minimum}_{k=0 \text{ à } M} \quad \left(\text{Dis}_1 = \sum_{i=0}^{N-1} [u_i - v_i]^2 \quad ; \quad \text{Dis}_2 = \sum_{i=0}^{N-1} [u_i + v_i]^2 ; \right.$$

$$\text{Dis}_3 = \sum_{p=0}^{\frac{N-1}{2}} [u_{2p+1} - v_{2p+1}]^2 + [u_{2p} + v_{2p}]^2 \quad ;$$

$$\text{Dis}_4 = \sum_{p=0}^{\frac{N-1}{2}} [u_{2p+1} + v_{2p+1}]^2 + [u_{2p} - v_{2p}]^2 \quad)$$

La distance choisie est la valeur minimale détectée parmi ces quatre sommes pour un vecteur du dictionnaire. Une fois ce vecteur identifié, nous retenons son adresse dans le dictionnaire, ainsi que la "transformation" qu'il a subie pour s'approcher le plus possible du vecteur incident à coder. Ces informations feront partie du codage. Un mot code d'identification est affecté à chaque bloc transformé pour indiquer sa classe, d'activité spatiale, son appartenance ainsi que la disposition des vecteurs fréquences spatiales et les "transformations" appliquées aux vecteurs répertoriés dans le dictionnaire.

Les transformations appliquées à la disposition des pixels dans le bloc image, ou celles appliquées à leurs luminances ont des répercussions directes sur les signes et les dispositions des coefficients transformés TCD. Ces propriétés découlent de la linéarité de la TCD, elles découlent aussi de la propriété de parité de la fonction cosinus.

Nous parlons ici de "transformations subies par les composantes d'un vecteur"; il s'agit en effet, de trois manipulations répertoriées:

- Inversion du signe de toutes les composantes du vecteur. Cela

correspond, au point de vue bloc image, à une inversion autour de 255 des valeurs de luminance des pixels.

- Inversion du signe des composantes paires du vecteur. La signification visuelle de cette opération est une symétrie du bloc image correspondant, par rapport à une de ses médianes.
- Inversion du signe des composantes impaires du vecteur en question. Cette manipulation découle, dans l'espace image, de la symétrie par rapport à une des deux médianes et à l'inverse autour de 255 des luminances du bloc correspondant au vecteur initial.

Quatre cas de figures sont envisageables pour chaque vecteur du dictionnaire. Ces possibilités sont indiquées sur deux bits précédant chaque adresse de vecteur transmise au décodeur.

La quantification vectorielle appliquée à l'espace transformé, dans sa version décrite précédemment, compare les quatre distances, chacune réalisée en tenant compte des signes de chaque composante du vecteur en question. Pour optimiser la taille du dictionnaire, une représentativité maximale est recherchée. la distance choisie permet d'utiliser un vecteur de coefficients transformés dans plusieurs blocs images, liés ensemble par des symétries ou des transformations de luminance des pixels.

Le dictionnaire est formé de sous-dictionnaires regroupant les vecteurs des différentes classes d'activité spatiale. Ainsi nous faisons correspondre à chaque classe d'activité spatiale, faible moyenne ou forte, une table d'indexation contenant des vecteurs de dimension cinq et deux tables dont la dimension des vecteurs est égale à trois. La taille des vecteurs n'est pas définitive. Il est possible , en effet, de tenir compte de vecteurs plus étendus si la qualité subjective des images décodées l'exige.

L'efficacité de la pré-classification de la vectorisation particulière des coefficients transformés et de l'utilisation d'une distance d'approche originale sont autant d'atouts pour notre procédé de codage nous détaillerons dans un prochain chapitre les performances obtenues. Nous avons utilisé des dictionnaires de très petite taille. Une "qualité subjective" suffisante pour les applications visées par le codage est atteinte avec quelques dizaines de vecteurs par sous-dictionnaire. En comparaison avec les 5000 vecteurs de dimension entre 10 et 15, l'expression "très petite taille" est tout à fait justifiée.

VII - C - 2 Création du dictionnaire

Une des particularités de la quantification vectorielle est de nécessiter une première sélection des vecteurs les plus représentatifs d'un ensemble d'apprentissage à partir duquel le dictionnaire final est construit. La technique classique utilisée en quantification vectorielle; est la méthode de classement des vecteurs appelée LBG [LIND 80].

Nous proposons d'optimiser l'utilisation du dictionnaire obtenu par le LBG, en exploitant les propriétés visuelles des spatio-fréquences, de façon à avoir le dictionnaire le plus représentatif et le plus condensé possible. Sachant que le débit final de la liaison de transmission d'images et la qualité des images décodées sont étroitement liés à la taille du dictionnaire utilisé par la technique de codage.

Une autre technique de classification des dictionnaire est en cours d'expérimentation au CERLOR. Elle est fondée sur une optimisation des techniques classiques et sur un parcours rapide du dictionnaire. Elle va contribuer à diminuer le temps total d'exécution du codage par quantification vectorielle.

VII - D Description du codage temporel

Le codage selon le procédé présenté ici est un codage hybride par transformation mathématique, quantification vectorielle et quantification scalaire. La composante fréquentielle continue est quantifiée par un quantificateur scalaire après une étape d'estimation du mouvement sur les blocs transformés fournis en séquence. Nous savons que pour une application de transmission de séquences d'images animées, la redondance temporelle du signal d'image vidéo est importante. Les blocs d'image dont les luminances évoluent fréquemment et de façon importante sont très peu nombreux. Il est avantageux d'utiliser un codage de prédiction et estimation de mouvement en parallèle avec le codage spatial par quantification vectorielle.

Cette approche est d'autant plus nécessaire et bénéfique que les applications de transmissions d'images auxquelles était initialement destinée cette technique de codage présentent une faible activité temporelle des images à coder.

VII - D - 1 Estimation de mouvements

Le procédé d'estimation de mouvements consiste en une identification des échantillons connus du signal d'image vidéo, dans notre cas des blocs de pixels, pour coder les échantillons présentés à l'émetteur.

Une comparaison est effectuée entre le bloc transformé à coder et ses neuf prédécesseurs directs. Nous identifions le bloc de l'image précédente qui est à la position actuelle du bloc transformé, par un déplacement simple. Une indication de son adresse permet de restituer le bloc en mouvement au décodage.

La recherche de déplacement se fait par l'intermédiaire d'un vecteur caractéristique dont les composantes sont : $(F(0,0);T)$.

- $F(0,0)$ est la composante continue elle représente la moyenne des luminances du bloc image,
- T représente l'activité spatiale du bloc transformé. Il représente la somme des carrés des composantes fréquentielles alternatives correspondant au bloc image.

Est déclaré en mouvement tout bloc de l'image précédente qui présente un faible écart de vecteur caractéristique par rapport au bloc actuel. Cette quantification vectorielle particulière traduit la ressemblance produite par un mouvement entre deux blocs transformés en comparant leurs vecteurs caractéristiques.

Dans le cas où aucun déplacement n'est identifié, la procédure de codage du bloc transformé est lancée. Elle se décompose en deux étapes :

- le codage de la composante continue codage prédictif différentiel exploitant comme prédicteur la composante continue du bloc transformé précédent.
- codage des composantes alternatives, une quantification vectorielle appliquée au domaine transformé pour les autres composantes.

VII - D - 2 Codage de la composante continue par prédiction temporelle

Selon la D.P.C.M. (Differential Pulse Coding Modulation), il ne s'agit pas de coder le signal mais de coder la différence entre ce signal et une valeur de prédiction. Cette différence est codée par un quantificateur scalaire. Le code final est ainsi transmis.

Une des particularités de notre procédé de codage d'image est de coder en D.P.C.M. la composante continue $F(0,0)$, identifiant la moyenne des luminances de chaque bloc image.

Le prédicteur utilisé est matérialisé par la valeur antérieur à $F(0,0)$ dans le temps.

VII - E Tests d'évaluation

Une approche sélective détermine les différentes possibilités d'évolutions pour un bloc image. Cette discrimination s'effectue par l'intermédiaire du vecteur $(F(0,0);T)$. Ce dernier est utilisé pour l'estimation de mouvements, il est exploité ici pour comparer les couples de blocs qui se succèdent dans l'axe des temps. Nous avons mis en valeur plusieurs cas de figure :

- Si les deux composantes $F(0,0)$ et T , n'ont subi aucune modification; cette possibilité rejoint l'estimation de mouvement. Le mot code indique au décodeur de restituer, dans l'image décodée, le bloc précédent dans le temps.
- Si T subit une évolution, mais pas $F(0,0)$. Le codage spatial de quantification vectorielle dans le domaine transformé est appliqué au bloc. La classification visuelle détermine les coefficients transformés pertinents, Les vecteurs sont formés et la quantification vectorielle leur est appliquée.
- $F(0,0)$ subit une évolution, mais pas T . Ce cas correspond à une variation globale de l'éclairage de la scène représentée par l'image. La différence entre les deux valeurs successives de $F(0,0)$ est effectuée, Cette valeur est ensuite quantifiée par un quantificateur scalaire. Les composantes fréquentielles spatiales autres que $F(0,0)$ sont

maintenues.

- Les deux paramètres subissent des variations. Le quantificateur scalaire code la différence entre les fréquences $F(0,0)$. Le quantificateur vectoriel réalise le codage des autres fréquences spatiales par quantification vectorielle.

La mise en oeuvre des éventualités décrites implique une adaptation supplémentaire du codage aux différentes évolutions du signal vidéo.

VII - F Décodage rapide

Nous avons élaboré une nouvelle technique de décompression rapide. Elle évite l'emploi de la transformée en cosinus inverse. Ce procédé exploite, de la même manière que le codage, les structures visuelles formant les blocs image. Il est applicable à chaque fois qu'il est nécessaire de décoder les images le plus rapidement possible et sans, pour cela, avoir besoin de système matérielle spécifique et particulièrement sans avoir besoin de processeur de traitement du signal pour effectuer la TCD.

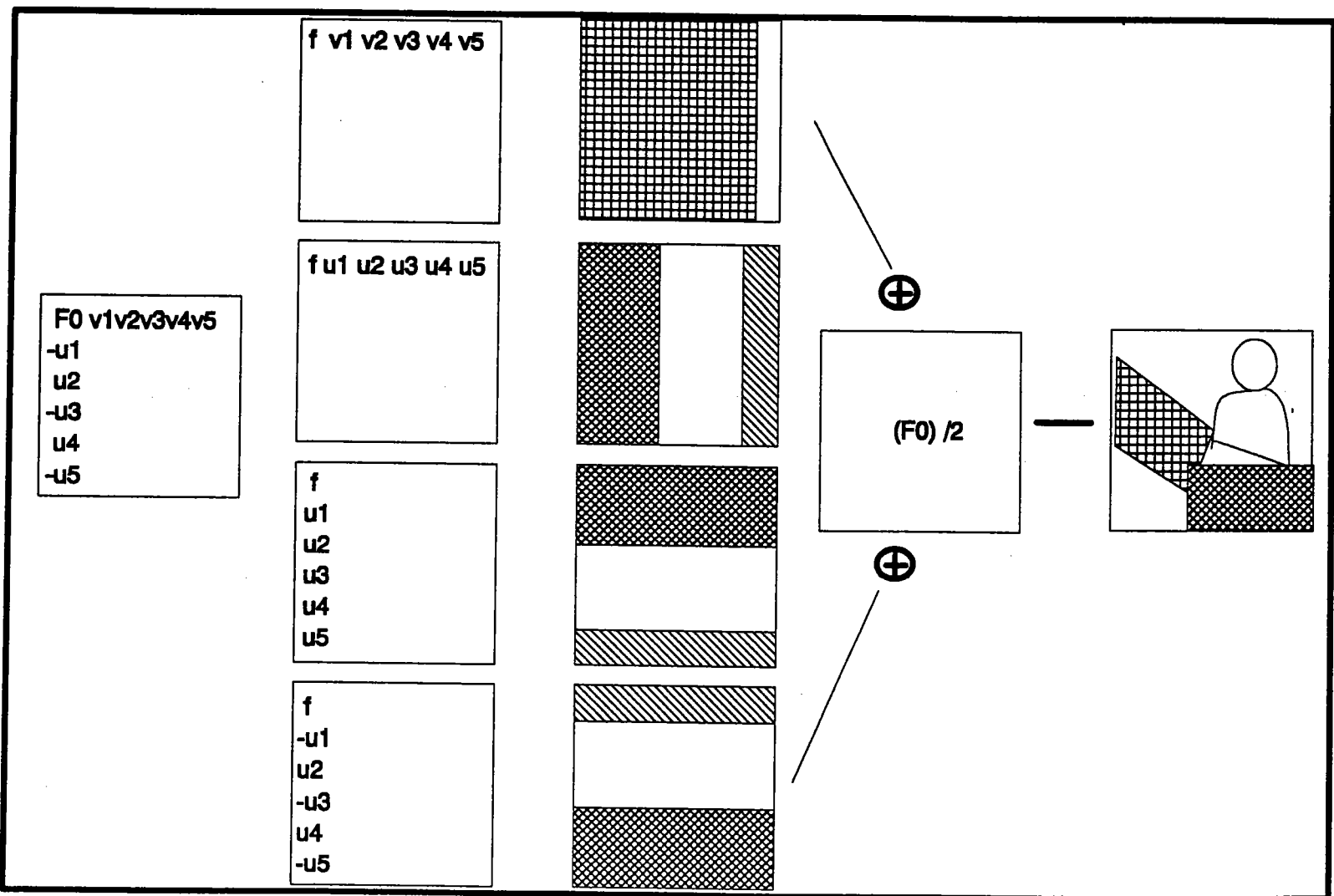
La linéarité de la transformation en cosinus discrète est la principale propriété mathématique qui justifie notre démarche algorithmique. En effet, pour éviter, la transformation inverse obligatoire pour reconstruire tout bloc image d'après ses composantes vectorielles fréquentielles, nous avons opté pour une opération de transformée inverse sur tous les vecteurs des dictionnaires, effectuée initialement avant le début du codage.

Les procédures de codage restent inchangées. Chaque bloc image est codé en affectant aux vecteurs d'éléments fréquentiels qui composent son bloc transformé l'adresse de leurs plus proches voisins dans le dictionnaire.

La nouvelle méthode de décodage rapide utilise comme dictionnaire une table de blocs de pixels correspondant à la transformée inverse des vecteurs initiaux composés d'éléments transformés TCD. Pour le codage, nous utilisons des manipulations des vecteurs transformés affectant les signes de leurs coordonnées. Ces opérations sont transposables dans le domaine des pixels, car elles découlent directement des propriétés visuelles de l'image et de leurs correspondances dans le domaine transformé. L'identification de ces caractéristiques visuelles a permis leur utilisation dans le décodage rapide.

Les affectations des signes et la disposition des composantes des vecteurs du dictionnaire dans l'espace transformé se traduisent par des symétries par rapport aux deux médianes ou par rapport aux deux diagonales du bloc image.

Figure N°36 décompression rapide des images



Le décodeur applique au bloc image, correspondant au vecteur codé, des manipulations géométriques indiquées par le codeur. Le bloc image représentant la composante continue étant facilement représentable, il est superposé aux autres blocs images obtenus après manipulations et affectés à chaque vecteur codé.

L'opération de décodage se trouve, ici, avantageusement simplifiée. Elle se traduit par une succession d'opérations simples :

- symétries par rapport aux axes principaux verticaux, horizontaux et diagonaux
- inversions, par rapport à 255, des luminances, effectués par une table d'indexation ou par une soustraction.
- sommes des blocs images obtenus à partir des vecteurs pour reconstruire le bloc image incident.

Le temps d'exécution du codage est ainsi très diminué nous avons obtenu donc des images décodées en ayant effectué la transformée en cosinus discrète une seule fois avant le début des opération de décodage. Elle a été appliquée seulement aux vecteurs du livre de code.

Lors du décodage d'un bloc, les opérations effectuées sont des soustraction, des additions et des manipulations géométriques sur le bloc image image mémorisé. Aucune multiplication n'est utilisée se qui explique la rapidité de cet algorithme de décompression.

Chapitre VIII

REALISATION MATERIELLE

Le CERLOR a initialement lancé l'étude d'un système de transmission d'image à bas débit dans le but d'effectuer un transfert de savoir-faire des produits de recherche. L'industrialisation de cet équipement nécessite la réalisation d'un prototype ouvert dans le sens où il permet de valider plusieurs algorithmes de codage en fonction de l'application visée. Rappelons qu'entre le taux de compression et la qualité subjective souhaitée un compromis permet de définir les différentes adaptations de codage retenues. Pour les mêmes raisons décrites ici, une autre qualité que doit vérifier le prototype est d'être évolutif. Cette propriété est proposée en réponse aux différentes variantes que peut présenter la technique de compression.

La complexité de calculs nécessaires pour les traitements rapides des images, ainsi que les objectifs fixés initialement nous ont amené à choisir, pour notre système de codage-décodage une configuration multiprocesseurs parallèles.

VIII - A Modules de traitements parallèles

Les unités de traitement ont des architectures comparables à celle des transputers, mais à une échelle plus grande. Elles sont composées de modules de calcul situés dans le transputer, une unité de calcul à virgule fixe une autre à virgule flottante. Elles sont aussi dotées d'une mémoire importante et d'une unité de communications équipée de quatre liens bidirectionnels de transmission.

Ces cellules représentent les éléments de base pour la construction d'un système réparti. Elles renferment leurs propres ressources indépendantes de traitement, de communication et de mémorisation.

La partie centrale de chaque cellule est matérialisée par un Transputer (mot composé à partir de : Transistor et de computer). C'est un composant VLSI programmable conduisant de manière simple à des performances élevées et à une adaptabilité importante aux besoins de l'utilisateur. Son architecture interne autorise l'implémentation d'algorithmes complexes et présentant des tâches exécutables de manière concurrentielle.

L'architecture du transputer a été optimisée de façon à fournir un maximum de fonctions. De plus, il offre des moyens de communication qui permettent, dans un système donné, la multiplication d'unités élémentaires sans accroître la complexité de mise en oeuvre [HIRS 90].

Le transputer est un microprocesseur disposant d'une mémoire locale et de liens pour connecter un circuit à un autre. Le transputer est un circuit VLSI programmable contenant un processeur, de la mémoire et des liens de communication pour les liaisons avec ses semblables ou avec d'autres circuits. Il est pourvu aussi d'une unité arithmétique travaillant en virgule flottante. Un transputer peut être utilisé seul ou en réseau pour construire des systèmes de traitements de haut niveau. Le réseau est alors obtenu en utilisant les liens de communications point à point.

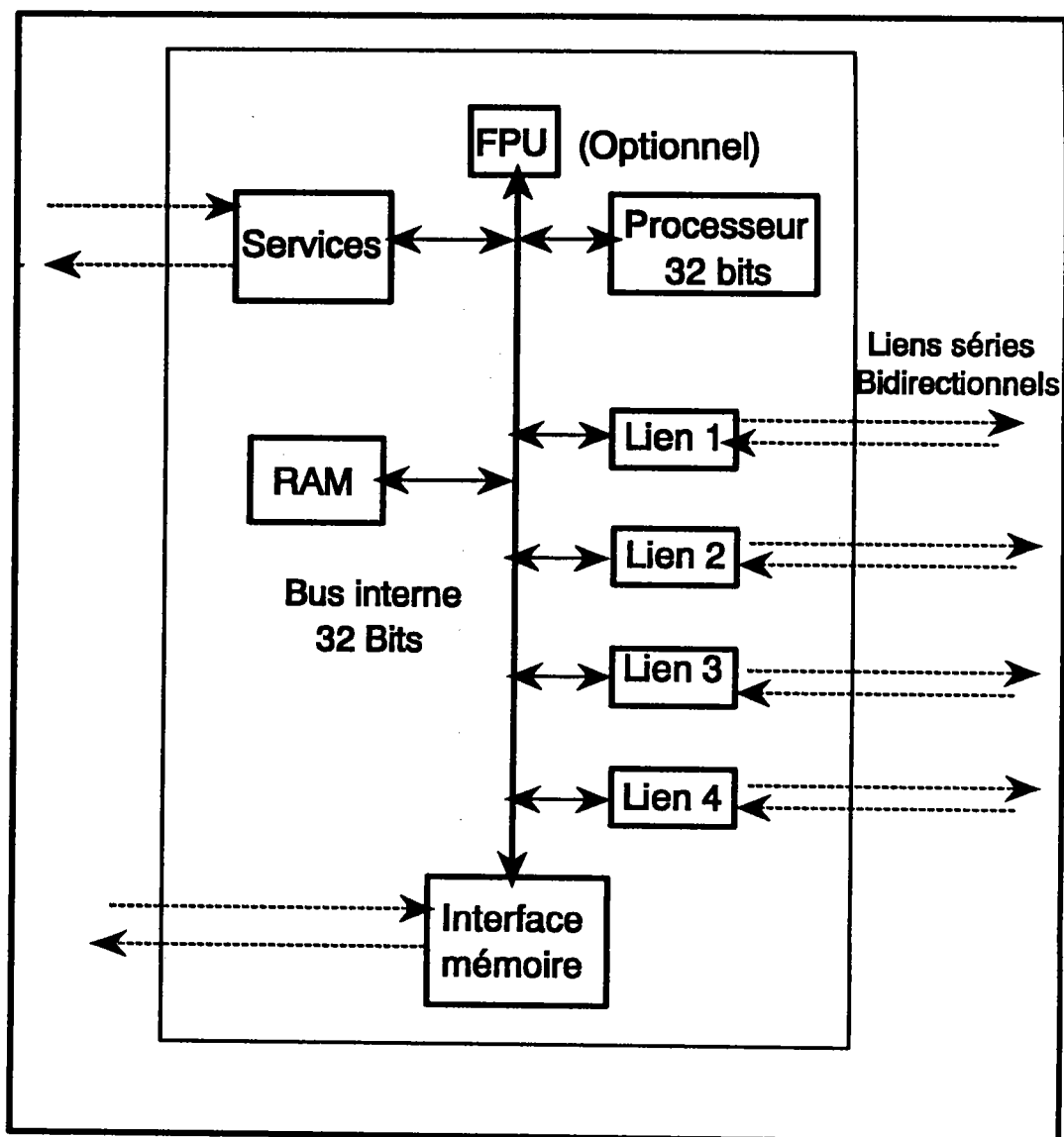


Figure N° 37 Architecture d'un transputer

Les liens de communication sont des liaisons séries à 10 Mb/s ou 20 Mb/s selon le composant. Chaque lien bidirectionnel est en fait composé de deux liaisons électriques. Le protocole de communication utilise un message d'autorisation de transmission, envoyé du récepteur à l'émetteur pour valider la communication.

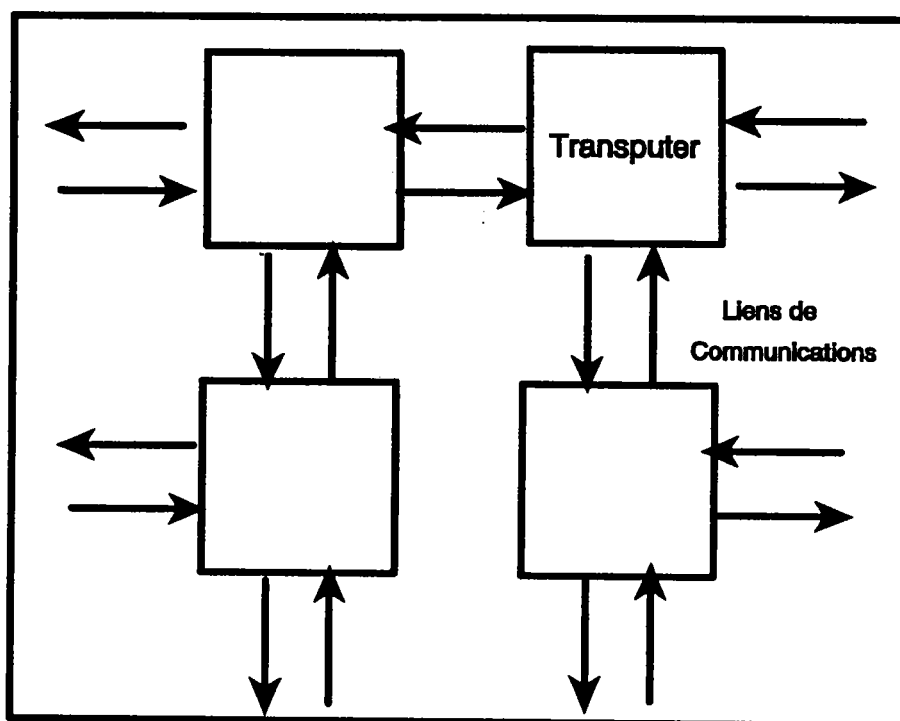


Figure N°38 Réseau de transputers

VIII - B Acquisition

L'acquisition est effectuée par un module élaboré autour d'un transputer, un convertisseur analogique / numérique et un contrôleur de signaux de synchronisation.

Le Transputer Frame Grabber (TFG de chez High-tech Services Partners) est une carte d'acquisition d'images, composée d'un processeur 32 bits le transputer T800, de 512 Koctet de RAM vidéo, de 1 Moct de RAM dynamique pour les utilisations propres du processeur, sans oublier le convertisseur analogique / numérique et le contrôleur de signaux de synchronisation TS68483.

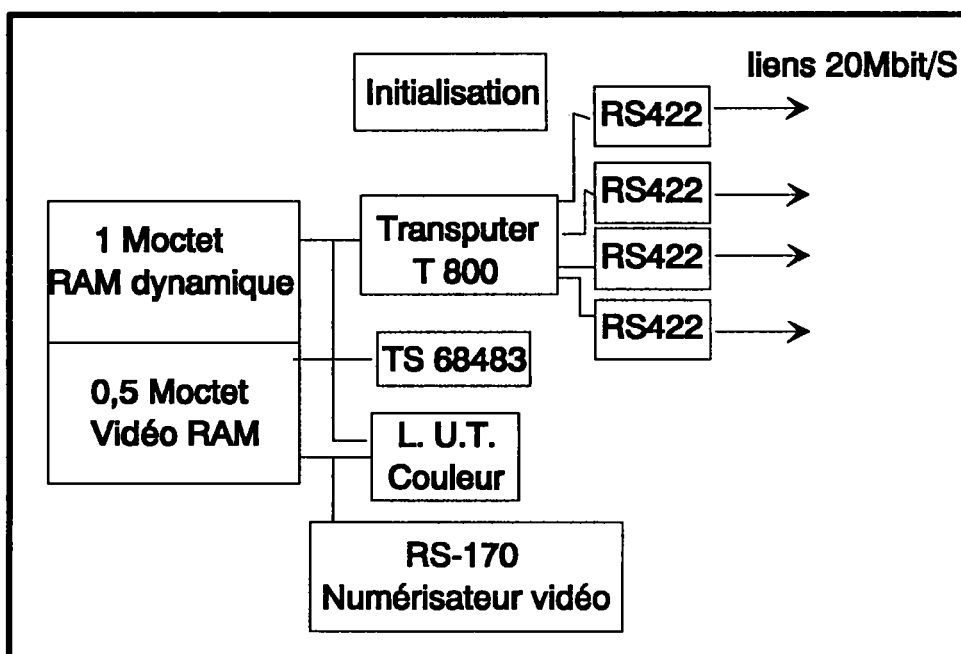


Figure N°39 schéma descriptif de la TFG

La RAM vidéo est organisée en 1024 x 512 pixels, un octet étant affecté à chaque pixel. le contrôleur graphique TS6883 est un générateur de synchronisation des signaux vidéo. Il est programmable pour reconnaître des signaux vidéo standards de type CCIR ou RS 170-A.

Le convertisseur analogique / numérique AD9502 est utilisé pour échantillonner les signaux vidéo standards à 12,5 MHz et délivre des informations numériques avec une résolution de 8 bits.

Le G170 "table d'indexation couleur" (Look-Up Table : LUT) sélectionne 256 couleurs à partir de 262144 possibles. Quand la TFG est utilisée pour la visualisation des images, la LUT génère trois signaux séparés analogiques : Rouge, Vert et Bleu, suivant le standard RS 170-A.

L'image numérisée, issue du module de conversion analogique / numérique, est décomposée en blocs de (8x8) pixels. Le transputer du module de conversion effectue cette tâche à la suite de laquelle les médaillons (8 x 8) sont transmis au module de traitements, exécutant la transformée en cosinus discrète.

VIII - C Le calculateur TCD

Il est formé d'un processeur de traitement du signal DSP dédié. Il

intègre trois fonctions à savoir la transformée TCD directe et inverse ainsi qu'un algorithme de filtrage.

Les algorithmes qui sont implémentés dans le processeur ont subi des opérations de simplification et d'optimisation de façon à augmenter la rapidité d'exécution en diminuant le nombre d'opérations nécessaires pour chaque calcul. Cela se traduit par diminution de la surface de silicium dont est formé le circuit.

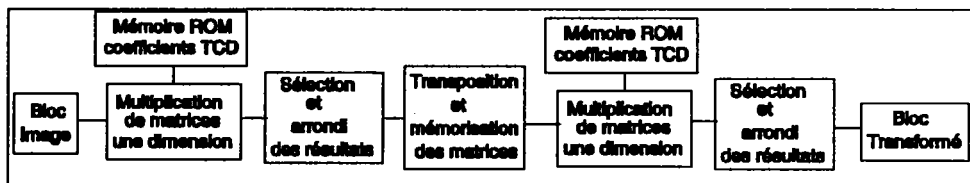


Figure N°40 Schéma fonctionnel de l'unité de calcul de la transformée en cosinus discrète

L'utilisation du composant consiste à choisir un des trois mode de fonctionnement : TCD directe, TCD inverse, filtrage. L'ensemble des calculs internes se font en virgule fixe.

L'algorithme de transformée en cosinus discrète intégré dans le DSP est essentiellement constitué de multiplication de matrices. D'une part la matrice de pixels d'autre part la matrice des valeurs de cosinus mémorisées dans des tables intégrées dans le composant. A la sortie de ce dernier nous obtenons un bloc transformé qui a subi une rotation et un retard de deux positions par rapport à sa position initiale dans le bloc image. Ce retard est provoqué par le temps de calcul nécessaire pour effectuer une TCD.

Le DSP fonctionne effectivement, en "pipe line". Il manipule trois blocs à la fois. Le premier est présenté en sortie sous forme de coefficients transformés, le second est en phase de traitement et enfin le dernier est récupéré en entrée sous forme de pixels. Ces trois opérations sont exécutées simultanément par le DSP.

Pour le mode TCD directe, le principe du circuit IMS A121 est de réaliser une doubles transformation monodimensionnelle pour effectuer une TCD à deux dimensions. Pour ce faire, on fait une TCD à une dimension sur les pixels horizontaux arrivant en entrée, les résultats sont mémorisés le temps d'exécuter la même TCD sur les colonnes.

Pour faciliter les communications avec les autres entités de calcul, nous avons muni cette unité de traitement d'une interface série/parallèle.

Elle adapte les liens série bidirectionnels des transputers aux bus entrée sortie du DSP.

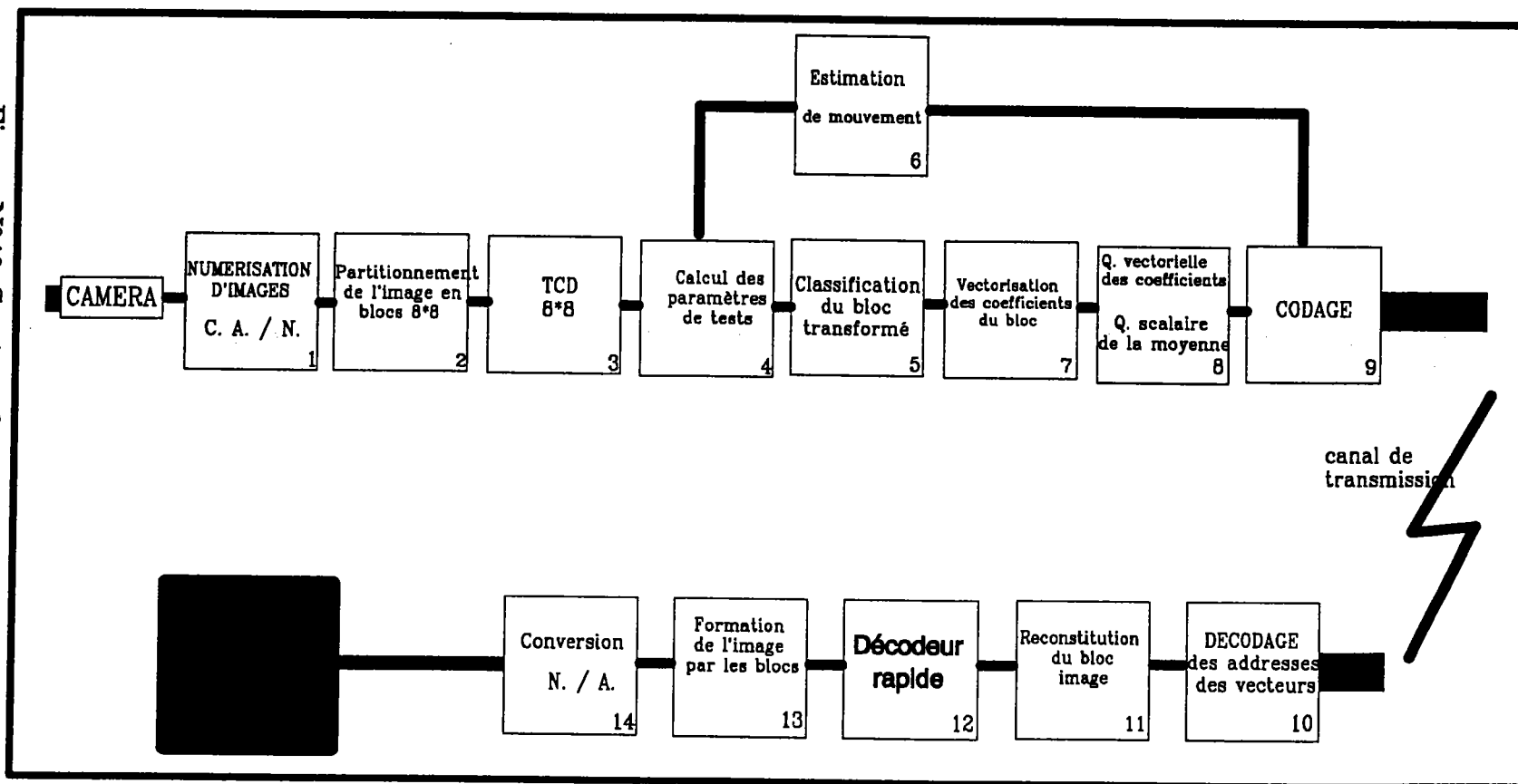
Le module calcul de la transformée en cosinus discrète est tout à fait transparent pour les autres modules. En effet, pour maintenir le standard de communication "lien de transputer" et pour minimiser la complexité de l'architecture nous avons tenu à ne pas introduire un bus rapide supplémentaire pour les données image.

Le module TCD est placé entre deux unités de traitement. La première transmet des blocs de pixels, la seconde reçoit des blocs de coefficients transformés. Rappelons que pour forcer le passage dans un sens précis et seulement dans ce sens de manière à ne pas perturber le fonctionnement du DSP TCD, nous avons configuré le réseau d'unités de traitements d'une façon adéquate, excluant ainsi toute tentative de communications à "contre sens".

VIII - D Le codeur

Le dispositif de codage comprend, en plus du module d'acquisition et celui exécutant la TCD, un réseau de modules programmables effectuant toutes les procédures de codage. Ce système englobe l'estimation de mouvements, le codage différentiel de la fréquence spatiale matérialisant la moyenne, la pré-classification des blocs transformés et enfin la quantification vectorielle des vecteurs formés par les fréquences spatiales obtenues à la suite de la transformée en cosinus discrète.

Figure N°40 Système de codage - décodage



VIII - D - 1 Description du schéma synoptique du système de transmission d'images animées

1 : Numérisation d'images obtenues à la sortie de la caméra vidéo, conversion analogique / numérique.

2 : Découpage de l'image en médaillon 8 fois 8 pixels.

3 : Transforme le bloc (8x8) pixels en bloc (8x8) coefficients transformés.

4 : Calcul du vecteur d'identification (T,NH/H,NV/V,D1,D2,D) à partir des coefficients transformés.

5 : Classification visuelle exploitant l'arbre de décision. Extraction des composantes de l'image.

6 : Nous recherchons le bloc le plus proche au bloc à coder dans l'image précédente. Le vecteur de déplacement est déterminé et transmis au codeur.

7 : Identification des coefficients pertinents. Regroupements des spatio-fréquences représentatives sous forme de vecteurs.

8 : Une quantification scalaire de la moyenne est effectuée. Nous recherchons le plus proche voisin dans un dictionnaire prédéterminé du vecteur à coder: Quantification vectorielle.

9 : Pour chaque bloc sont transmises la classe visuelle d'appartenance, la quantification de sa moyenne - matérialisée par la fréquence $F(0,0)$ - ainsi que les adresses, dans le dictionnaire, des vecteurs qui forment le bloc de coefficients transformés TCD. Si un déplacement d'un des blocs de l'image précédente a été détecté dans 6, seul le vecteur de déplacement est transmis.

10 : Le décodage est effectué après la réception. Aux adresses des vecteurs sont associés des vecteurs stockés dans le dictionnaire du décodeur.

11 : A chaque vecteur nous affectons un bloc image correspondant aux fréquences des coefficients transformés. Les manipulations nécessaires pour retrouver le vecteur codé sont effectuées; elles rassemblent les symétries rotations et autres transformations des positions et des luminances des pixels.

12 : Le décodage rapide est appliqué en sommant tous les blocs image qui

composent chaque bloc image (8 x 8).

13 : L'image est reformée à partir des blocs obtenus à la sortie du décodeur rapide.

14 : L'image est visualisée sur un moniteur vidéo après une conversion numérique / analogique.

VIII - D - 2 Modules de codage

Le module d'acquisition transforme le signal vidéo en données numériques représentant les luminances des pixels constituant l'image. L'image ainsi formée est découpée en médaillons (8x8) pixels. Chaque médaillon est par la suite transmis en entrée du calculateur Transformée en cosinus discrète.

Le calculateur effectuant les transformées en cosinus discrètes directe et inverse a été conçu autour d'un processeur de traitement du signal: IMS A121 2-D discrete cosine transform image processor de INMOS SGS - THOMSON Microelectronics.

A la sortie de ce dernier nous obtenons un bloc transformé composé des fréquences spatiales représentant le bloc image. Plusieurs opérations sont effectuées sur ce bloc transformé par le système de codage:

- Estimation de mouvements; Une comparaison, des vecteurs composés de $F(0,0)$: (moyenne des luminances des pixels du bloc image); T : (activité spatiale de ce même bloc calculée à partir de fréquences spatiales). L'estimation de mouvement est effectuée entre le bloc à coder et ses antécédents sur l'image précédente. Si un déplacement de bloc est détecté une demande est formulée pour restituer ce bloc au décodage.

- Codage différentiel de la moyenne; La fréquence $F(0,0)$ matérialisant la moyenne est codée par un codeur scalaire. Nous calculons la différence entre cette même valeur et celle correspondante sur le bloc temporellement précédent. Cette différence est affectée par une des valeurs d'un tableau qui représentent les différences les plus probables. Les valeurs d'indexations sont codées sur quatre bits. Au décodage chaque valeur sera remplacée par un représentant de l'intervalle des différences.

- Pré-classification; Le système de codage met au point pour chaque

bloc transformé le vecteur composé des paramètres de classification. En tenant compte de ces composantes une série de tests est appliquée pour définir l'appartenance du bloc à une des classes visuelles. A partir de cette classification un choix des fréquences spatiales pertinentes est déterminé. La formation des vecteurs s'en suit, les vecteurs sont composés d'éléments fréquentiels selon la décision prise après les tests de classification. Le système de codage lance une recherche dans le dictionnaire pré-défini pour déterminer les vecteurs les plus proches de ceux qui forment le bloc transformé. Une fois cette tâche terminée, le système fait correspondre une adresse à chaque vecteur pertinent du bloc transformé.

Les différentes procédures rappelées ici sont exécutées séparément par les modules de traitement. Pour optimiser le système nous avons minimisé les communications entre cellules de traitement, limitant celles-ci au strict nécessaire : la transmission de résultats intermédiaires et définitifs.

Le codeur fait correspondre à chaque bloc transformé une suite de mots de code. Le premier mot correspond au déplacement subi par un des blocs précédents ou à la classe du bloc et aux manipulations subies par les vecteurs du dictionnaire pour s'approcher le plus possible des vecteurs codés. Le second correspond au codage différentiel de la moyenne. Le nombre et la longueur des mots suivants sont variables en fonction de la famille d'appartenance du médaillon transformé. Le code est composé de plusieurs adresses de vecteurs et des transformations qu'ils ont subies.

VIII - E Le décodeur

Le décodeur est constitué par un module d'identification des codes et un système d'identification vectorielle. Pour chaque code reçu, l'identification permet de reconnaître la classe d'appartenance du bloc transformé codé, ainsi que la disposition de chaque vecteur transformé et les manipulations qu'il a dû subir pour répondre au mieux à la quantification.

Nous reconstituons le bloc image à partir de la moyenne et des blocs image élémentaires affectés aux vecteurs coefficients. Nous reformons la moyenne en sommant la différence reçue du codeur prédictif et la valeur de cette moyenne sur l'image précédente.

Les vecteurs du dictionnaire du décodeur sont les transformés par la TCD inverse des vecteurs de fréquences du dictionnaire du codeur. Leur

adresse sont indiquées par le mot code du bloc, nous appliquons sur ces vecteurs les manipulations adéquates signalées par le codeur.

A la suite de cette opération sont restitués tous les blocs image correspondants à tous les vecteurs composant le bloc transformé incident. Les manipulations géométriques leur sont appliquées, chacun selon le mot code que lui a affecté le codeur. Le bloc matérialisant la composante continue et donc la moyenne est formé, il s'agit d'un bloc image (8 x 8) dont la luminance des pixels est uniforme égale à $\frac{F(0,0)}{2}$.

Une fois terminées toutes les transformations affectant les positions et les valeurs des luminances des pixels formant les blocs images élémentaires, ils sont tous superposés. La composition, obtenue par la somme de tous ces blocs, forme finalement le bloc image décodé.

A partir des médaillons d'image obtenus à la sortie de la chaîne de décodage l'image finale numérisée est reconstituée. Elle est convertie en signal vidéo par un convertisseur numérique / analogique. Le module de sortie est directement relié à un moniteur pour la visualisation.

VIII - F Architecture du système de codage

La principale caractéristique de ce système est la modularité. Elle se traduit par une répartition dans les composantes du dispositif de l'ensemble des moyens informatiques mis en oeuvre. Le codeur-décodeur est composé d'entités indépendantes, pouvant échanger des informations, conçues autour de transputers ou de processeurs de traitement du signal. Les différents processus du codage ou décodage sont exécutés indépendamment les uns des autres sur des modules capables d'échanger les informations.

Chapitre IX

PRESENTATION DES PERFORMANCES

Une des particularités de notre technique est d'utiliser **une méthode de codage hybride adaptatif**. Ce procédé exploite les avantages des trois techniques classiques de codage; codage prédictif, codage par transformation et quantification vectorielle.

La méthode de codage est fondée sur **un codage spatial et sur un codage temporel**. Le premier consiste à transformer l'image par la transformée en cosinus discrète et à appliquer une technique de classement visuelle pour identifier les fréquences spatiales pertinentes au codage. La quantification vectorielle permet de coder les vecteurs de coefficients transformés retenus. Elle exploite une distance particulière que nous avons élaborée et qui est elle aussi fondée sur des constatations visuelles.

Le codage temporel consiste, lui, en une estimation de mouvements et un codage différentiel de la moyenne des luminances des pixels formant l'image.

Une autre particularité essentielle de notre procédé est **l'optimisation de l'utilisation du dictionnaire** formé par les vecteurs de quantification. Elle découle des propriétés suivantes :

- **une exploitation maximale des propriétés visuelles des blocs image en vue d'une classification suivant des structures perceptibles par l'observateur,**

- **une vectorisation adaptative suivant la famille d'appartenance du bloc image,**
- **une distance exploitant les manipulations possibles sur les vecteurs formés par les fréquences spatiales,**
- **une optimisation totale de l'utilisation du dictionnaire. La robustesse de la méthode de codage est telle que même pour un dictionnaire brut et non classé la qualité des images décodées reste acceptable.**

Selon notre méthode de décompression rapide, le décodage est fondé sur la reconstruction des blocs élémentaires matérialisant les structures visuelles de l'image. La superposition des blocs fondamentaux restitue des images avec la **qualité vidéoconférence**.

IX - A Classification visuelle des blocs

IX - A - 1 Activité spatiale

Notre procédé de codage est fondé sur une méthode de classification des blocs image suivant l'importance des variations de luminance des pixels et suivant les structures qui les composent. Nous allons présenter des images prises dans la série de référence et sur lesquelles nous avons appliqué des traitements de façon à mettre en valeur les différentes classes d'appartenance. En premier lieu nous nous proposons de mettre en évidence les familles d'activité spatiale.

Les images utilisées sont : Femme, Lacornouaille et Couloir elle sont composées de (248x248) pixels codés sur 8 bits. Dans la série d'images N°1 sont présentées les images originales :

La série d'images N°2 met en valeur les blocs uniformes contenus dans chacune des images de référence. Ces blocs sont identifiés par l'intermédiaire du paramètre activité totale T . Si ce dernier a une valeur inférieure à un seuil pré-établi, le bloc est déclaré uniforme. Au codage, il sera représenté par sa moyenne qui correspond au premier coefficient de la matrice transformée.

Les trois séries d'images suivantes N°3, N°4 et N°5 mettent en évidence les différentes familles de blocs. L'appartenance à ces classes est

définie par l'activité spatiale de chaque bloc. Une sélection discriminatoire à l'aide de seuils détermine la catégorie de chaque bloc. Les seuils ont été établis subjectivement après plusieurs expérimentations (cf. Tab. N°1).

Images 248x248 p, 961 blocs	Femme	Lacornouaille	Couloir
Blocs homogènes	372 = 38%	217 = 22,5%	528 = 55%
Blocs à faible activité spatiale	310 = 32,2%	259 = 27%	235 = 24,5%
Blocs à moyenne activité	95 = 9,9%	148 = 15,4%	73 = 7,6%
Blocs à forte activité spatiale	184 = 19,4%	337 = 35%	125 = 13%

Tableau N°1 Distribution suivant l'activité spatiale des images de référence.

Le pourcentage par image du nombre de blocs, appartenant à une classe donnée, dépend essentiellement du contenu de l'image (cf. Tab. N°1). Le paramètre *T* d'activité spatiale quantifie l'importance des évolutions de luminance représentées sur un bloc image. En effet, les évolutions brutales, contenues dans le bloc, se traduiront par un spectre fréquentiel étendu dans la matrice transformée. Le spectre est matérialisé par les coefficients transformés pris en valeurs absolues.

IX - A - 2 Classification suivant les formes de transitions

La série d'images N°6 représente les blocs composés d'évolutions de luminance verticales ou horizontales. Cette catégorie de blocs est identifiée par la classification suivant des structures géométriques contenus dans les blocs image. Cette même classification permet de reconnaître les blocs symétriques par rapport à l'une de leurs diagonales, mis en évidence dans la série N°7. L'identification des structures géométriques est effectuée dans l'espace transformé. Elle exploite les propriétés retenues dans l'étude des correspondances entre l'espace image et l'espace transformé (cf. Tab. N2, 3 et 4).

nombre de blocs par classe	Evolutions verticales	Evolutions horizontale	Symétries / Diagonale 1	Symétries / Diagonale 2	Blocs non structurés
Blocs faible activité	27	23	51	124	85
Blocs activité moy	4	24	4	0	63
Blocs forte activité	8	35	1	0	140
Nombre de blocs par classe en %	4%	8,5%	6%	13%	30%

Tableau N°2 Distribution des structures géométriques pour Femme.

nombre de blocs par classe	Evo-lutions verticales	Evolutions horizontale	Symétries / Diagonale 1	Symétries / Diagonale 2	Blocs non structurés
Blocs faible activité	42	25	24	74	94
Blocs activité moy	11	53	3	0	81
Blocs forte activité	35	85	0	0	217
Nombre de blocs par classe en %	9,1%	17%	3%	8%	40,8%

Tableau N°3 Distribution des structure géométrique pour Lacornouaille.

nombre de blocs par classe	Evolutions verticales	Evolutions horizontale	Symétries / Diagonale 1	Symétries / Diagonale 2	Blocs non structurés
Blocs faible activité	72	97	16	24	26
Blocs activité moy	6	48	0	0	19
Blocs forte activité	52	55	0	0	18
Nombre de blocs par classe en %	13,5%	20,8%	2%	2,5%	6,5%

Tableau N°4 Distribution des structure géométrique pour Couloir.

Dans la série d'images N°8 sont matérialisés les blocs dont la diagonale transformée est importante. Cette classe renferme aussi les blocs symétriques par rapport à leur centre. Le nombre de ces derniers n'étant pas important nous avons décidé de les regrouper avec d'autres blocs. L'ensemble vérifie la condition précitée.

Nous remarquerons dans la série N°9, où nous avons représenté les blocs images par leurs moyennes respectives, que des plages uniformes se forment. Elles traduisent une redondance de la moyenne des blocs image. Il est judicieux alors d'exploiter cette constatation en codant la moyenne par un codage prédictif. L'estimateur optimal est une fonction simple des moyennes des blocs voisins et de la moyenne du bloc précédent dans le temps.

La série d'images N°10 est une représentation en niveaux de gris des coefficients transformées des image initiales. Nous notons l'activité forte dans les zones où l'évolution de la luminance des pixels composant l'image initiale est importante. Nous pouvons aussi remarquer la disposition des coefficients, sur la première ligne ou sur la première colonne, dans les blocs transformés correspondant aux blocs images contenant des évolutions verticales ou horizontales.

Nous avons mis en évidence dans la série N°11 des images sur lesquelles nous avons appliqué un filtrage. Pour chaque bloc transformé, les 21 coefficients $F(u,v)$ de rang faible dans la matrice transformée sont retenus et codés. Ces coefficients sont ceux dont les coordonnées vérifient ($u+v < 6$

avec u,v allant de 0 à 7).

Nous avons tenu a calculer un paramètre de distorsion nous permettant de comparer les différentes séries d'images. Nous avons quantifié les dégradations apportées par chaque traitement par un facteur SNR, ce facteur permet de comparer l'image initiale - formée de N pixels de luminance $f_1(i)$ - et l'image finale traitée - $f_2(i)$ -.

$$SNR = 10 \log_{10} \frac{\sum_{i=0}^{N-1} (f_1(i))^2}{\sum_{i=0}^{N-1} (f_1(i)-f_2(i))^2}$$

Après application du filtrage, la distorsion obtenue dans la série N°11 .

	Femme	lacornouaille	couloir
SNR (db)	26,3	26,2	23,2

Tableau N°5 SNR : série N°11

La classification suivant les structures visuelles pour retenir les coefficients transformés pertinents est présentée dans la série d'images N°12. Elle permet de repérer les blocs contenant des structures particulières et d'adapter les seuils de discrimination de façon à optimiser le codage.

	Femme	lacornouaille	couloir
SNR (db)	26	25,2	21,8

Tableau N°6 SNR : série N°12

IX - A - 3 Quantification vectorielle

La quantification vectorielle appliquée à tous les coefficients obtenus après le filtrage (cf. série N°13) exploite deux dictionnaires de 15 vecteurs chacun. Le premier est formé de vecteurs de dimension 5, utilisés pour les premières lignes et colonnes du bloc transformé. Le second dictionnaire est composé de vecteurs de dimension 3, il quantifie les coefficients transformés citués sur les secondes lignes, les secondes colonnes, il permet aussi de quantifier les coefficients, différents de F(0,0), de la première diagonale du bloc transformé.

	Femme	lacornouaille	couloir
SNR (db)	22,4	19,3	14,2

Tableau N°7 SNR : série N°13

Les séries d'images N°14, 15 et 16 montrent des images où les dictionnaires de différentes tailles sont utilisés :

- Dans la série N°14, les images sont codées par une quantification vectorielle utilisant deux dictionnaires de 15 vecteurs,

- dans la série N°15 nous utilisons deux dictionnaires de 64 vecteurs chacun,

- finalement, dans la série d'images N°16 nous avons exploité 2 dictionnaire de 256 vecteurs chacun, de dimensions 5 et 3.

SNR (db)	Femme	lacornouaille	couloir
2 Dictionnaires de 15 vecteurs série N°14	22,1	19,3	14,2
2 Dictionnaire de 64 vecteurs série N°15	22,8	20,5	16,1
2 Dictionnaire de 256 vecteurs série N°16	23,6	21,4	16,6

Tableau N°8 SNR : série N°14,15 et 16

Nous constatons que l'augmentation de la taille des dictionnaires exploités pour la quantification vectorielle, améliore le rapport signal sur bruit.

IX - B Taux de compression

A la suite de l'application de la méthode de codage par quantification vectorielle sur l'espace transformé, nous obtenons des taux de compression détaillés dans le tableau suivant :

Taux de compression	2 Dictionnaires de 15 vecteurs	2 Dictionnaires de 64 vecteurs	2 Dictionnaires de 256 vecteurs
Femme	20	18	15,8
Lacornouaille	16	14,5	13,7
Couloir	28,5	26	24,5

Tableau N°9 : Taux de compression .

débit en nombre de bits par pixel	2 Dictionnaires de 15 vecteurs	2 Dictionnaires de 64 vecteurs	2 Dictionnaires de 256 vecteurs
Femme	0,4	0,5	0,5
Lacornouaille	0,5	0,6	0,6
Couloir	0,3	0,3	0,3

Tableau N°10 : Nombre de bits par pixels après codage.

Les performances que nous avons obtenues sont intéressantes dans le sens où nous pouvons reconstruire l'image à partir d'un petit nombre d'éléments fréquentiels en gardant une qualité subjective acceptable.

L'amélioration de la qualité subjective passe par la formation de dictionnaires adaptés aux différentes classes d'activité. L'adaptation prend en compte un nombre variable de coefficients retenus dans chaque classe d'activité spatiale.

Le codage temporel peut amener une amélioration pour les blocs dont la qualité subjective n'est pas satisfaisante. En effet, cela se traduit par la prise en compte d'informations supplémentaires pour les zones d'images fixes. Pour les structures en déplacement les détails véhiculés par l'image ne sont pas perceptibles par l'observateur, il n'est donc pas nécessaire de coder toutes ces informations.

IX - C Comparaisons avec les techniques connues

Nous avons décrit dans ce mémoire plusieurs méthodes plus ou moins classiques basées sur une quantification vectorielle avec une adaptation visuelle. Elles font correspondre à chaque famille de structures visuelles un dictionnaire spécifique. Ce livre de codes est classé et constitué de vecteurs composés par les motifs perceptibles qui caractérisent la famille en question. Cette approche algorithmique amène, pour rendre fidèle l'adaptation, à décomposer chaque livre de code, affecté à une famille visuelle, en plusieurs sous-dictionnaires. Ces derniers sont justifiés par une nouvelle adaptation pour représenter toutes les variantes possibles dans une transition de

luminance donnée. Citons par exemple, la famille des transitions verticales; elle va être représentée par plusieurs sous-dictionnaires correspondant aux différentes positions de la transition verticale dans le bloc image à coder.

L'indexation d'un dictionnaire à chaque évolution particulière perceptible ne peut être justifiée que si les représentants de cette famille sont statistiquement relativement nombreux. Le nombre des blocs appartenant à une classe visuelle donnée doit, donc, nécessairement être compté sur plusieurs images de référence car la taille du dictionnaire représentant cette famille en dépend.

La technique que nous avons élaborée est effectivement fondée sur une identification des caractéristiques visuelles contenues dans l'image. Mais, notre reconnaissance des structures perceptibles aboutit à un algorithme différent des autres méthodes. En effet, notre classification visuelle permet de situer les informations pertinentes et de les traiter d'une façon particulière.

Nous n'appliquons pas la quantification vectorielle pour trouver un bloc ressemblant visuellement au bloc image à coder, mais pour identifier et coder séparément les informations qui correspondent aux structures visuelles qui le composent.

Les classes visuelles particulières que nous avons identifiées permettent dans le cas des transitions verticales et horizontales de négliger 90% des coefficients transformés TCD. Pour les blocs image ayant une symétrie par rapport à une de leurs diagonales, 17 coefficients transformés sont pertinents. Deux fonctions liant les coefficients situés des deux côtés de la diagonale principale du bloc transformé, sont identifiées à partir des correspondances entre l'espace image et l'espace transformé. Elles permettent de coder seulement 9 coefficients (14% du total) et de restituer la totalité des coefficients pertinents au décodage en fonction des neuf reçus.

Les classes d'activités identifiées sont exploitées pour effectuer une adaptation des seuils délimitant les groupes de blocs image. Cette adaptation limite les défauts qui apparaissent lors du décodage. En effet, une déformation parvenue lors du codage est plus perceptible dans un bloc fortement actif.

Une particularité algorithmique puissante différencie avantageusement notre technique des autres méthodes, il s'agit en effet du paramètre de distorsion utilisé pour la quantification vectorielle. Les

opérations effectuées sur la polarité des coordonnées formant les vecteurs de coefficients transformés sont une manière efficace d'augmenter la capacité de représentation du dictionnaire utilisé. Nous l'avons décrit : pour les techniques classiques, pour une famille d'évolutions particulières donnée, il faut autant de sous dictionnaires que de positions de transitions dans le bloc image.

A l'aide du paramètre de distorsion que nous avons élaboré, nous pouvons à partir d'un seul vecteur reformer quatre blocs image en modifiant seulement la polarité des composantes. Les blocs représentent la même évolution de luminance avec un gradient positif, négatif et les compléments à 255 de ces deux blocs image.

En attendant la découverte du dictionnaire "universel" les auteurs précités se sont attaqués à l'amélioration des livres de codes. Mais, l'augmentation de la représentativité des dictionnaires est un problème ardu pour les techniques de codage par quantification vectorielle. L'étendu des dictionnaires utilisés par les méthodes connues est trop importante pour permettre le maintien de la simplicité et la souplesse des équipements codeurs - décodeurs.

Marescq dans [MARE 86] initialise son dictionnaire avec 2438 vecteurs de dimension 16. Mais, pour garder une qualité subjective acceptable, il a été amené à regrouper plusieurs classes visuelles non représentatives. L'équiprobabilité de l'appartenance d'un bloc donné à une famille visuelle est indispensable pour maintenir des performances stables. Cependant, la robustesse d'un codeur de ce type est sérieusement affectée par la nature aléatoire du contenu des images à coder.

Marescq a présenté une méthode pour améliorer son procédé de codage, elle est basée sur le rafraîchissement des dictionnaires. Simultanément un algorithme effectue l'intégration de nouveaux vecteurs non représentés initialement dans le dictionnaire. Cette approche permet de compenser les évolutions aléatoires de l'image, mais elle fait passer la taille du livre de codes de 2000 vecteurs à 4000. Le débit moyen sur une séquence d'image est de 3,3bits/pixel avec un supplément de 1,5bits/pixel nécessaire pour le rafraîchissement, les nouveaux vecteurs étant transmis au décodeur.

Monet dans [MONE 89] a étudié l'évolution du nombre de blocs par classe visuelle. Ces travaux montrent que d'une image à l'autre le pourcentage de blocs ayant les mêmes caractéristiques visuelles peut passer

du simple au double. L'utilisation de dictionnaires de tailles fixes est donc à proscrire. En effet, la répartition des blocs en classes visuelles figée fait que l'élimination de certaines classes devenues non représentatives est indispensable. la taille maximale permise est 4096 éléments de façon à pouvoir les coder sur 12 bits par quantification vectorielle.

Dans une étude de synthèse menée par Monet, un livre de codes de taille fixe est utilisé. Il est constitué de plusieurs sous-dictionnaires dont les tailles sont adaptées aux taux d'appartenance des blocs image à coder :

Classe i	1	2	3	4	5	6
Taille T_i	16	128	1024	64	512	2048

Classe i	7	8	9	10	11	12
Taille T_i	1024	2048	32	128	1024	2048

Tableau N°11 Tailles des dictionnaires affectés aux classes visuelles.

Le débit estimé à partir de ce dictionnaire est de 1,43bit/pixel. La qualité des images en mouvement décodées obtenue a été considérée comme insuffisante. Cela a amené Monet et al. à améliorer le procédé pour aboutir à un dictionnaire de taille plus importante que le précédent, mais qui offre des possibilités de représentation en arborescence beaucoup plus avantageuse pour le codage. Cette structure du livre de codes est particulièrement adaptée au rafraîchissement temporel.

Arbre	Trames impaires	Trames paires
Taille totale	13462	14900
Nombre de feuilles	6538	7457

Tableau N°12 Evolution des dictionnaires par rafraîchissement.

Le débit moyen estimé varie autour de la valeur de 2,5bits/pixel. Il est fonction de la taille des dictionnaires et du nombre de vecteurs nouveaux affectés lors du rafraîchissement.

Avec une technique exploitant tour à tour la transformée en cosinus discrète et la quantification Buhler [BUHL 86] présente une approche avec adaptation visuelle. Dans un chapitre précédent nous avons décrit les points positifs de cette méthode de codage. D'un point de vue conceptuel, ce procédé est classique en quantification vectorielle adaptée visuellement. Il fait appel à des sous-dictionnaires formés à partir de blocs composés de structures perceptibles classiques. Les tailles des livres de codes utilisés sont formés de vecteurs de dimension 16 correspondant aux 16 pixels des blocs 4x4 d'image.

Le débit estimé pour cette technique de codage varie entre 0,6bit/pixel et 0,9bit/pixel.

Plusieurs dictionnaires ont été testés par Buhler. Ils sont inspirés des travaux de Ramamurthi et Gersho.

Les livres de codes se décomposent de la façon suivante :

Sous-livres de codes	Nombre de vecteurs
6 livres "contours verticaux"	120
6 livres "contours horizontaux"	120
16 livres "contours diagonaux"	120
1 livre "indéfinis" (sens vertical)	240
1 livre "indéfinis" (sens horizontal)	240
1 livre "indéfinis" (sens diagonal)	240
1 livre "activité moyenne"	1024
1 livre "uniforme"	128
nombre total des vecteurs dans le dictionnaire	5232

Tableau N°13 Dictionnaire classé N°1 :

Un dictionnaire N°2 a été utilisé par Buhler et al., il présente des dictionnaires dont les tailles sont le double de celles des sous-livres de codes du N°1. Le nombre total de vecteurs utilisé est de 10336.

Les vecteurs qui sont classés dans ces dictionnaires sont tous de dimension 16 correspondant aux luminances des blocs 4x4 d'image.

Une étude statistique a donné les pourcentages suivants :

- 25% des vecteurs sont classés "contours".
- 40% des vecteurs sont classés "activité moyenne".
- 35% des vecteurs sont classés "uniformes".

Les débits obtenus après la quantification vectorielle sont étroitement liés à la taille des dictionnaires exploités. Les mots codes représentant chaque bloc, dépassent 10 bits par mot. Un codage de Huffman s'impose donc, il permet d'augmenter avantageusement le taux de compression.

Le nombre de bits par pixel obtenu avec la technique décrite par Buhler [BUHL 86], varie entre 0,4bits/pixel et 0,8bit/pixel. Ces résultats découlent d'expérimentations effectuées sur une série d'images représentant

des visages.

Nous avons décrit les techniques qui nous ont semblé les plus originales, en mettant en valeur leurs points forts. Nous avons surtout insisté sur les particularités qui les différencie de la méthode de codage que nous avons élaborée. Nous nous proposons ici d'énumérer les avantages de notre méthode de codage :

- Identification des structures perceptibles dans l'espace transformé; les paramètres de décision sont calculés simplement en tenant compte des correspondances entre l'espace pixel et celui des fréquences. En opposition avec les paramètres classiques fondés sur le calcul des gradients locaux ou autres.
- Notre répartition en familles ne nécessite pas une équiprobabilité du facteur d'appartenance d'un bloc image aux classes visuelles. En effet, contrairement aux autres techniques l'aspect aléatoire du contenu de l'image est pris en compte. Nous n'essayons pas de restituer les blocs image à coder, mais nous reproduisons séparément les évolutions primaires de luminance qui les composent.
- La classification des particularités visuelles d'un bloc n'est pas utilisée pour affecter le bloc incident à un livre de code particulier, mais, elle est exploitée pour identifier les coefficients transformés pertinents pour le codage. Les correspondances entre le domaine des pixels et celui des coefficients transformés permettent d'élaborer des fonctions pour restituer la totalité des coefficients pertinents à partir d'un petit groupe restreint.
- Le dictionnaire manipulé pour la quantification vectorielle est construit par des vecteurs dont les composantes traduisent dans le domaine fréquentiel les évolutions de luminance formant un bloc image. Les différentes positions que vont avoir les vecteurs de coefficients dans les matrices transformées sont essentielles. Elles permettent la composition des blocs, offrant multiples possibilités de variations de luminance que présentent les images décodées.
- Le paramètre de distorsion élaboré manipule la polarité des composantes des vecteurs. Il s'adapte parfaitement à la correspondance entre les fréquences spatiales et les évolutions de luminance qu'elles représentent. Ce paramètre traduit en partie le signe variable des gradients calculés sur les blocs de pixels. Il permet d'éviter d'utiliser

plusieurs sous-dictionnaires représentant toutes les possibilités de positions pour la même transition dans un bloc d'image.

L'accumulation de ces particularités avantageuses a permis la reconstitution d'image au décodage à partir de dictionnaires de tailles très réduites, en comparaison avec celles utilisées par les auteurs précités.

Le débit que nous avons présenté n'est pas estimé mais c'est un débit réel obtenu à la suite du codage d'images de référence. Nous pouvons améliorer notre taux de compression en utilisant un codage de Huffman exploité couramment dans ce type d'applications.

Rappelons que la technique de Huffman est applicable à chaque fois que nous avons affaire à des séquences de données répétitives. Elle exploite pour cela, la fréquence d'apparition de chaque valeur à transmettre pour évaluer le nombre de bits nécessaires pour former son nouveau mot code.

Le codage par quantification vectorielle est parfaitement adapté à cette amélioration. Les valeurs à transmettre pour chaque bloc images sont essentiellement les adresses des vecteurs les plus proches dans le dictionnaire. La fréquence d'apparition de chaque adresse est fonction de la pertinence du vecteur correspondant. Il est donc judicieux de coder les adresses sur des mots code de longueurs variables. La réduction de débit apportée par le codage de Huffman est utilisée aussi dans les technique de codage par quantification scalaire. Il offre un gain important en taux de compression.

IX - D Extensions et perspectives pour le procédé de codage

Nous n'avons pas exploité la réduction de débit offerte par le codage de Huffman dans notre technique de compression d'images. Une version définitive de notre méthode devra tenir compte de cet apport supplémentaire.

L'objectif que nous nous sommes fixé initialement été de concevoir une méthode de codage d'image de qualité visioconférence. L'amélioration de la qualité subjective des images codées est envisageable. En effet, plusieurs possibilités sont offertes:

- La classification des dictionnaires peut être optimisée par une technique qui exploiterait les paramètres caractéristiques des blocs image ; activité spatiale des vecteurs. L'adaptation des dictionnaires aux classes d'activité se traduira par une évaluation de la norme des

vecteurs pour mieux les utilisés dans la quantification vectorielle.

- Une adaptation prenant en compte un nombre plus important de coefficients transformés obtenus après la TCD est un facteur non négligeable d'amélioration de la qualité subjective. Si le nombre de coefficients retenus est en fonction de la classe d'activité le débit total peut être maintenu à des valeurs basses.
- Les seuils de décision pour la classification visuelle présentent une possibilité d'ajustabilité, en fonction de la qualité subjective et du taux de compression, qui offre une souplesse supplémentaire au procédé de codage.

La petite taille des vecteurs et des dictionnaires que nous exploitons pour la quantification vectorielle offrent des perspectives intéressantes. En effet, dans ce cas de figure la réalisation d'un codeur - décodeur intégré est matériellement envisageable. Une première étude pour la réalisation d'un composant ASIC effectuant la quantification vectorielle a aboutit un prototype utilisant un dictionnaire de 64 vecteurs de dimension 5. Ce module est une composant important dans un codeur-décodeur intégré pour la compression d'images.

CONCLUSION

Nous avons présenté dans ce mémoire, une nouvelle technique de compression d'images, applicable à chaque fois qu'il est nécessaire de réduire la quantité d'informations en vue d'une transmission à bas débit.

La conception de systèmes de codage d'image est étroitement liée à l'application visée initialement. En effet, la complexité des terminaux de codage-décodage et la qualité subjective des images restituées après décodage dépendent étroitement des contraintes imposées par l'application.

Le procédé de codage d'images que nous avons conçu est fondé sur une analyse du contenu de l'image. Des structures géométriques composant l'image sont reconnues à partir de leurs impacts sur l'espace fréquentiel. L'originalité de notre méthode de codage réside dans l'extraction de correspondances établies entre l'espace image formé de pixels et l'espace fréquentiel constitué par les coefficients transformés en TCD. Une connaissance approfondie des répercussions des différentes évolutions de luminance sur l'espace transformé a permis d'optimiser la méthode de compression.

Des méthodes utilisant le même principe de classification visuelle sont handicapées par la complexité de leurs traitements. Elles effectuent l'identification visuelle dans l'espace des pixels en calculant par exemple des gradients locaux de luminance.

Un avantage de la méthode hybride de codage que nous présentons ici, est la classification visuelle des blocs incidents. Nous établissons la répartition des blocs par identification des structures régulières de variation de la luminance dans l'espace transformé. Une méthode nécessitant des calculs simples sur les coefficients transformés nous permet de reconnaître

des évolutions verticales, horizontales, obliques, ou des zones uniformes. Cette approche facilite l'identification des coefficients transformés pertinents et de leurs positions respectives dans la matrice transformée en fonction des formes géométriques qu'ils représentent. Elle nous a permis de reconnaître des structures jamais identifiées jusqu'alors.

Un autre avantage de notre procédé de compression est d'appliquer la quantification vectorielle sur l'espace transformé. Cette technique a déjà été utilisée, mais nous proposons une nouvelle approche fondée sur l'identification de l'impact subjectif des vecteurs de fréquence sur l'espace des pixels. Cette méthode originale a permis de réduire avantageusement la taille des dictionnaires ainsi que la dimension des vecteurs qui les composent. Nous revendiquons pour notre méthode de codage l'utilisation judicieuse de la polarisation des coefficients transformés. Cette polarisation permet d'utiliser un même vecteur du dictionnaire dans plusieurs configurations géométriques pour composer les différentes évolutions de luminance.

Une autre caractéristique importante de notre système de compression est la possibilité d'effectuer le décodage d'une façon rapide en évitant la transformée en cosinus discrète inverse.

L'exploitation des propriétés précitées pour l'optimisation de la méthode de codage se traduit par l'obtention de performances intéressantes. Le taux de compression moyen obtenu sur des images de référence atteint 20, pour une qualité subjective de la visioconférence. L'image et ses structures perceptibles sont reconstituées par un nombre très réduit de composantes fréquentielles.

Une des extensions envisageables pour notre méthode est éventuellement l'amélioration de la qualité subjective pour la compression des images fixes, par exemple. Cela est possible par l'adaptation des dictionnaires par classes d'activité ou par l'extension de leurs tailles. Le codage temporel offre, lui aussi, une possibilité d'amélioration de la qualité. En effet, pour les zones dont les détails sont porteurs d'informations importantes, des vecteurs supplémentaires seront transmis au décodeur. L'algorithme rapide de décompression permet d'ajouter à l'image restituée le ou les vecteurs contenant les détails supplémentaires, sans remettre en cause ou retraiter les informations déjà visualisées.

L'évolution des méthodes de codage et l'amélioration de leurs performances passent inévitablement par le développement de techniques

adaptatives fondées sur l'analyse du contenu de l'image. Ces procédés permettent simultanément de cerner les informations pertinentes au codage et de quantifier leur impact subjectif sur l'observateur.

Liste des références

- [AHME 74] N. Ahmed, T. Natarajan, K.R. Rao, "Discrete Cosine Transform", IEEE Trans. on Comm., Vol. COM.- 22, N°1, pp. 90-93, January 1974.
- [PONC 74] J. Poncin, "Codage numérique d'images à débit réduit (système intra-images]", Rapport interne C.C.E.T.T, Février 74.
- [KRET 77] F. Kretz, J.L. Boudeville, P. Sallio. "Optimization of DPCM video coding scheme using subjective quality criterium" Proc. of IEEE conference N°37, Sept 77.
- [CHEN 77] W. H. Chen, C. H. Smith : "Adaptive coding of monochrome and color images", IEEE Trans. on Com., Vol COM - 25, n°11, Nov. 1977, pp. 1285-1292.
- [LIND 80] Y. Linde, A. Buzo, R. M. Gray, "An algorithm for vector quantizer design", IEEE Trans. on Com., vol. com-28, n°1, janvier 1980.
- [DEWI 81] j. Dewitte, "Un système de codage MIC différentiel adaptatif avec critères psychovisuels de qualité," Thèse Docteur-Ingénieur, Rennes, INSA, Dec 1981.
- [GERS 82] A. Gersho, " On the structure of vector quantizers," IEEE Trans. Inform. Theory, vol. IT-28 , pp. 157-166, Mar. 1982.
- [WU 82] J. K.Wu, R. E. Burge, "Adaptive Bit Allocation for Image Compression", Comp. Grap. and Image Processing, Vol. 19, pp. 392-400. 1982.
- [BAKE 82] R. L. Baker, R. M. Gray, "Image compression using non adaptive

spatial vector quantization," in Proc. Conf. Rec. Sixteenth Asilomar Conf. Circuits, Syst., Comput., Oct 82, pp. 55-61.

- [KRET 83] F. Kretz, "Edges in visual scenes and sequences : application to filtering, sampling and adaptive DPCM coding," in Image Sequence Analysis : Part 2, ed : T. S. HUANG, 1983 (Springer Verlag).
- [RAMA 83] B. Ramamurthi, A. Gersho, "Image coding using segmented codebooks," presented at Proc. Int. Picture Coding Symp., Mar 1983
- [NASR 84] N. M. Nasrabadi, R. A. King, "A new image coding technique using transform vector quantization", in Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing, Mar. 1984.
- [GRAY 84] R. M. Gray, " Vector quantization," IEEE ASSP Mag., pp.4-29, Apr.1984.
- [KOHO 84] T. Kohonen, "Self organization and Associative memory," New York: Springer-Verlag. 1984.
- [RAMA 84] B. Ramamurthi, A. Gersho, "Image vector quantization with a perceptually-based cell classifier," presented at IEEE Poce. Int. Conf. ASSP, Mar 1984.
- [LOHS 84-1] H. Lohscheller, "Adaptive transform coding for still picture communication," Int. Seminar on Digital Communication, Zurich, pp. 25-31, Mar.1984.
- [LOHS 84-2] H. Lohscheller, "A subjectively adapted image communication system", IEEE Trans. on Com., Vol. Com-32, N°12, Dec 1984.
- [CLAR 85] R.J. Clarke : "On dynamic range of coefficients generated in transform processing of digitalised image data", IEEE Proc, vol 132, Pt F, n°2, April 1985, pp ;107-110.
- [BARB 85] D. Barba: "Méthodes de codage adaptatives d'images par

transformation discrète en cosinus", Rapport interne Centre Commun d'Etudes de Télédiffusion et Télécommunications, RTI/T/019/85, 1985.

- [NASR 85] N. M. Nasrabadi, "Use of vector quantizers in image coding", in Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing, Mar. 1985, pp. 125-128.
- [GUIC 86] J. Guichard, D. Nasse, "L'image numérique", L'écho de la recherche N°126, 4ème trimestre 1986.
- [RICH 86] C. J. Richard, "Compression du signal vidéo", Techniques de l'ingénieur 6-1986, E 5500.
- [GOLD 86] M. Goldberg, P. Boucher, S. Shlien, "Image Compression Using Adaptive Vector Quantization," IEEE Trans. on Com., Vol. COM-34, N°2, Feb 1986.
- [MURA 86] T. Murakami, K. Asai, A. Itoh, "Vector Quantization for color images". ICASSP 86, Vol. 1, pp. 133.
- [BUHL 86] Y. Buhler, " Etude d'une méthode hyerarchique pour le codage à faible débit des images numériques,"
- [MARE 86] J. P. Marescq, "Etude de schémas de quantification vectorielle adaptative multiclassés. Application au codage d'images télévisuelles." Thèse de Docteur-Ingénieur, Rennes, 1986.
- [MARE 87] J. P. Marescq, C. Labit, "Temporal adaptive vector quantization for image sequence coding," In SPIE, Vol. N°804, pp. 356-363, 1987.
- [VAIS 87] D.J. Vaissey, A. Gersho, "Variable Block Size Image Coding", Proc. IEEE Int Conf. ASSP, pp. 1051-1054, April 1987.
- [NASR 88] N. M. Nasrabadi, R. A. King, "Image Coding Using Vector Quantization: A Review", IEEE Trans. on Com., Vol. 36, N°8, Aug. 1988.
- [KHAL 88] M. H. C. Khalifi, "Etude et réalisation d'un module d'acquisition et de

traitement d'images destiné à la maintenance par ordinateur d'équipements électronique," Thèse d'Université, NANCY 1, Novembre 1988.

- [LABI 88] C. Labit, "Adaptativité et schéma de compression de données. Application au codage de séquences d'images télévisuelle," Thèse d'Université, Rennes I, 25 février 1988.
- [BELA 88] A. Belaïd, A. Boyer, "Techniques de reconnaissance des formes", Rapport CRIN-ESE, 1988.
- [ZHAO 89] Y. Zhao: "Etude et réalisation d'un système de transmission d'images numériques à bas débit", Thèse d'Université, NANCY 1, Octobre 1989.
- [DAVI 89] A. Davignon, " Classification en blocs de taille variable pour codage d'image par quantification vectorielle", Traitement du Signal, Vol. 6 - n°4, 1989.
- [MONE 89] P. Monet, C. Labit, "Etude de schémas de codage de séquences d'images par quantification vectorielle multiclasse adaptative temporelle," Rapport final de convention INRIA-CCETT 88ME27, Décembre 88.
- [HIRS 90] E. Hirsh, "Les transputers : application à la programmation concurrente," E. Eyrolles, 1990.
- [MATH 90] P. Mathieu, M. Barlaud, M. Antonini, "Compression d'image par transformée en ondelette et quantification vectorielle", Traitement du Signal, Vol. 7, N°2, 1990.

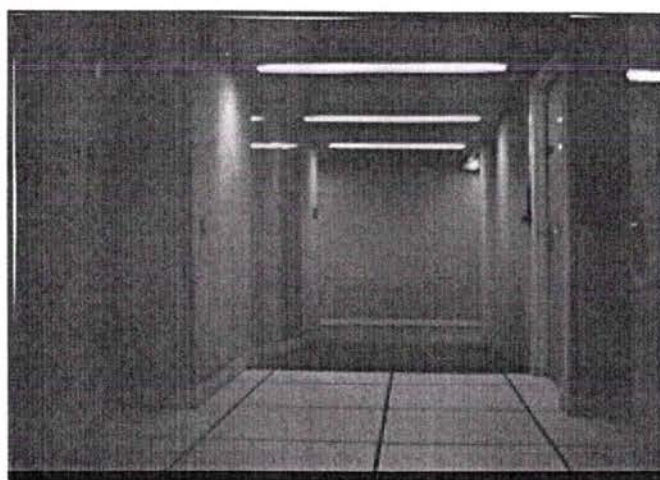
ANNEXES

1) séries d'images présentées:

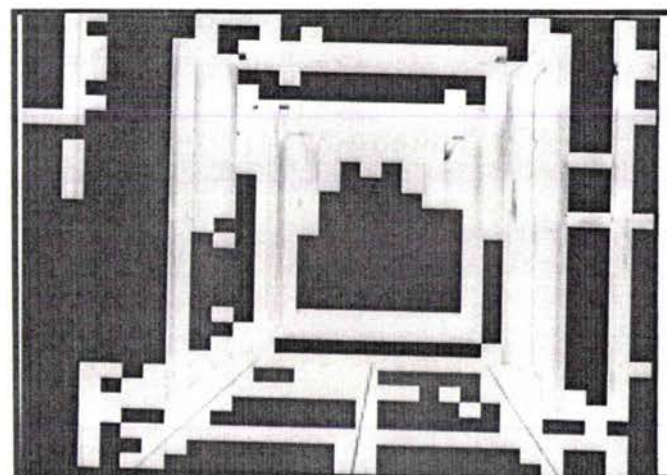
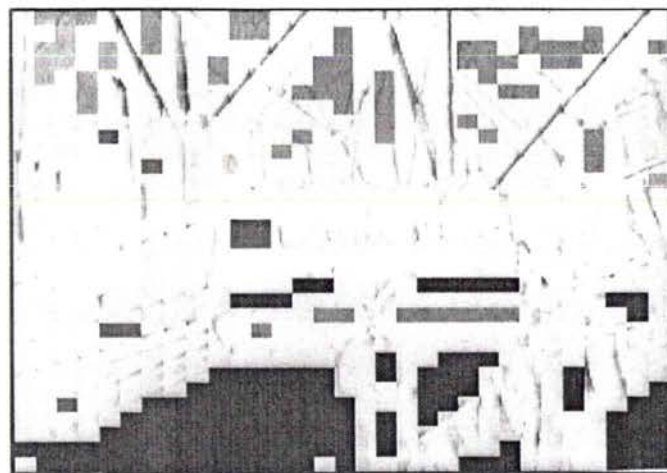
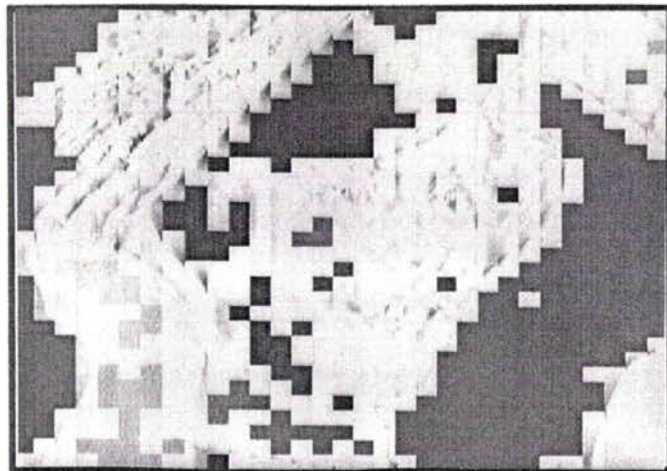
- Images initiales de référence,
- mise en valeur des blocs homogènes,
- bloc image de faible activité spatiale,
- bloc image d'activité spatiale moyenne,
- bloc image de forte activité spatiale,
- blocs image composés de structures verticales et horizontales,
- blocs image symétriques par rapport à l'une de leurs diagonales,
- blocs image présentant une symétrie centrale,
- moyennes des blocs de l'image,
- images des blocs transformés en cosinus discrète,
- images filtrées,
- images classées par familles visuelles,
- images codée par quantification vectorielle sans classification,
- images codée par quantification vectorielle avec classification visuelle et par un dictionnaire de 32 vecteurs,
- images codée par quantification vectorielle par un dictionnaire de 128 vecteurs,
- images codée par quantification vectorielle par un dictionnaire de 512 vecteurs.

2 - Démonstrations des correspondances les plus pertinentes, identifiées entre l'espace image et l'espace transformé. :

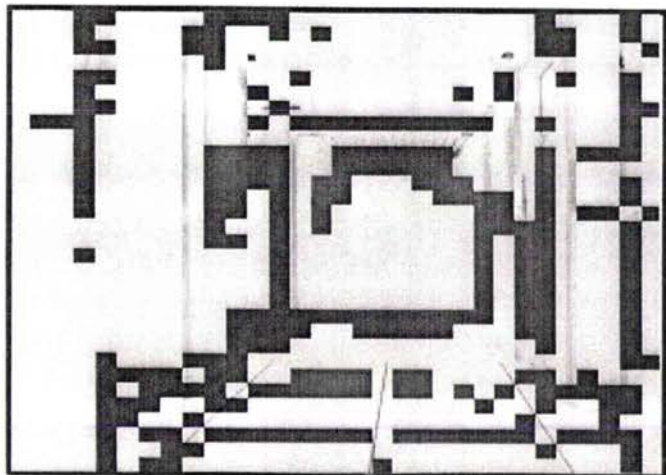
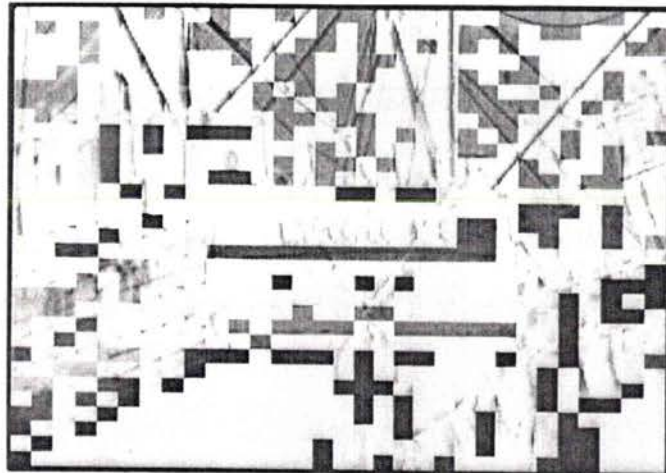
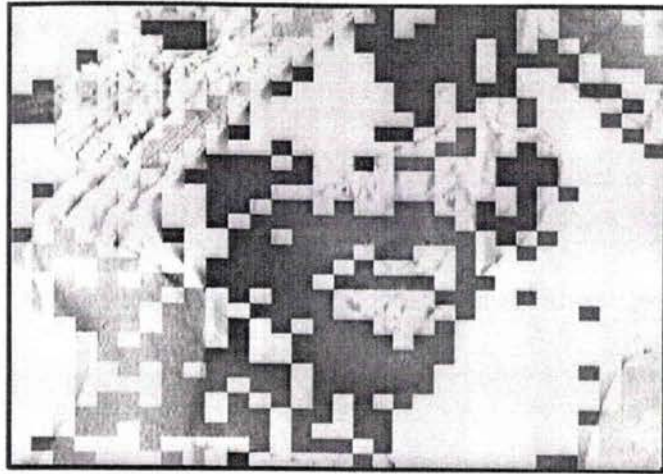
- Blocs symétriques par rapport à l'axe vertical central,
- blocs symétriques par rapport à l'axe horizontal central,
- bloc symétrique par rapport à sa première diagonale,
- bloc symétrique par rapport à sa seconde diagonale.



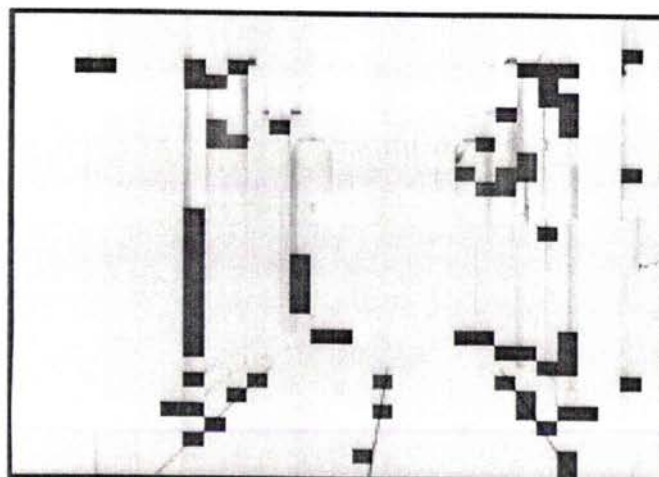
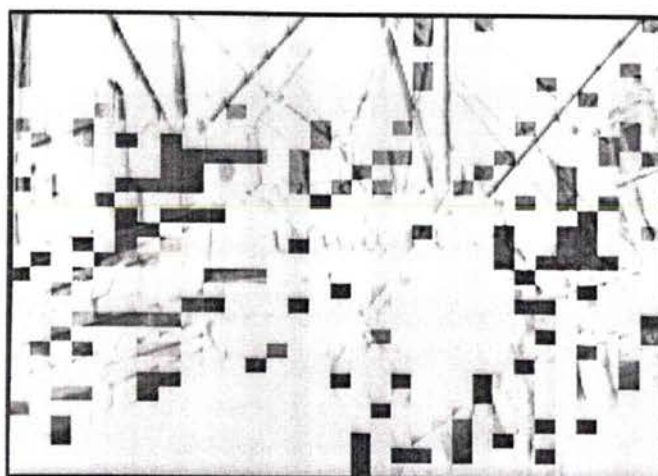
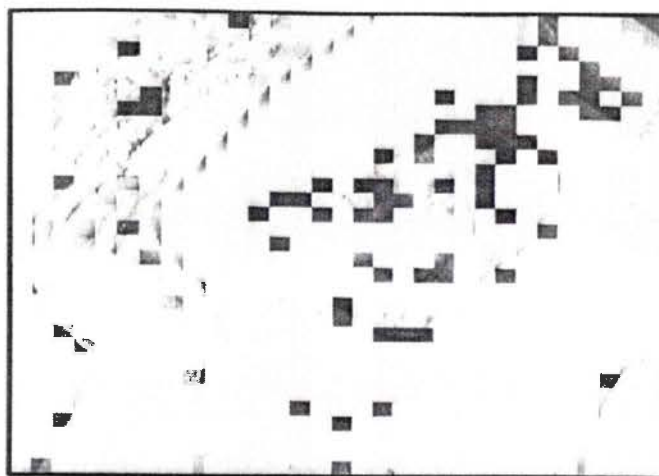
Série d'images N°1 : Images de références.



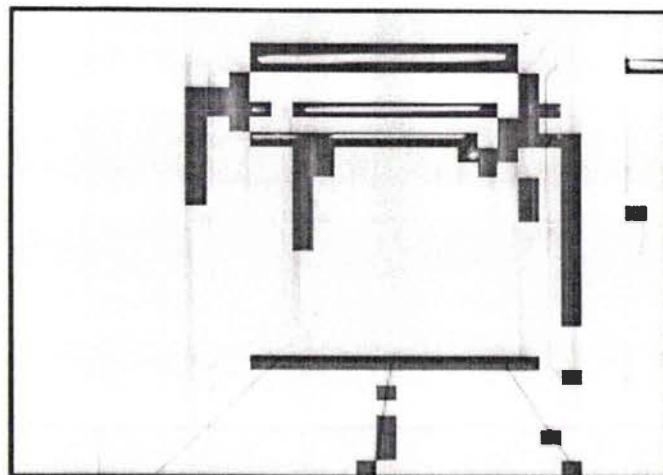
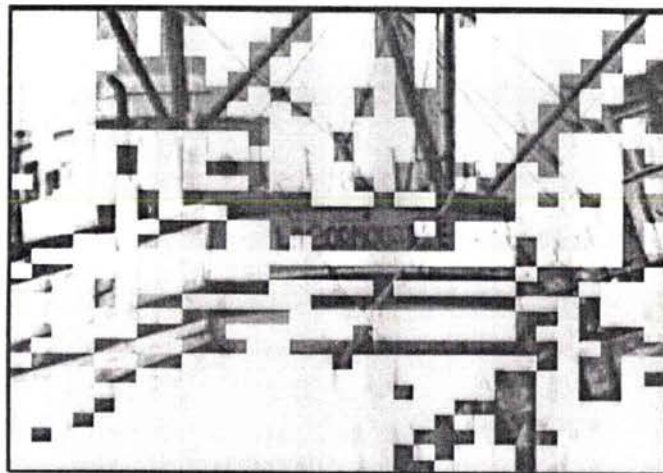
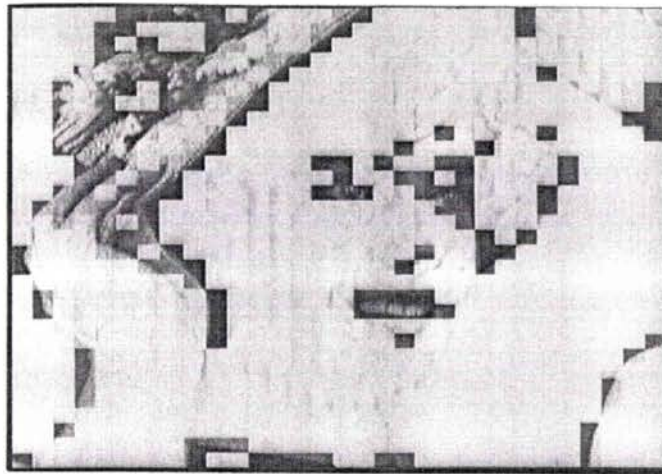
Série d'images N°2 : Blocs image homogènes.



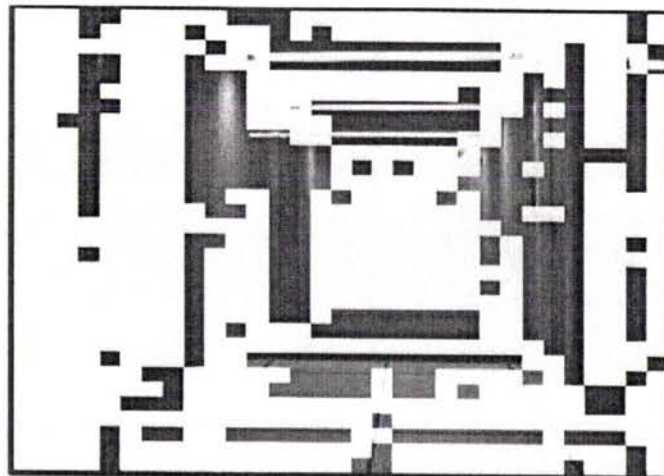
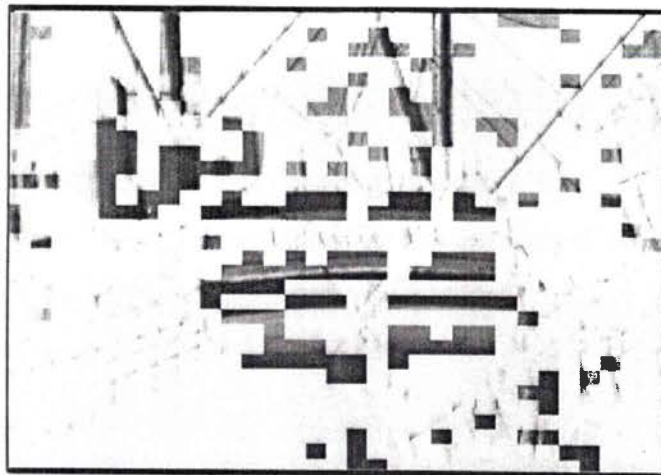
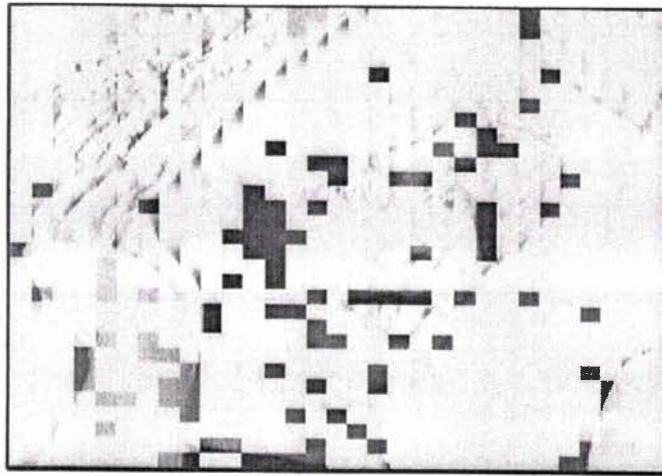
Série d'images N°3 : Blocs imagede faible activité spatiale.



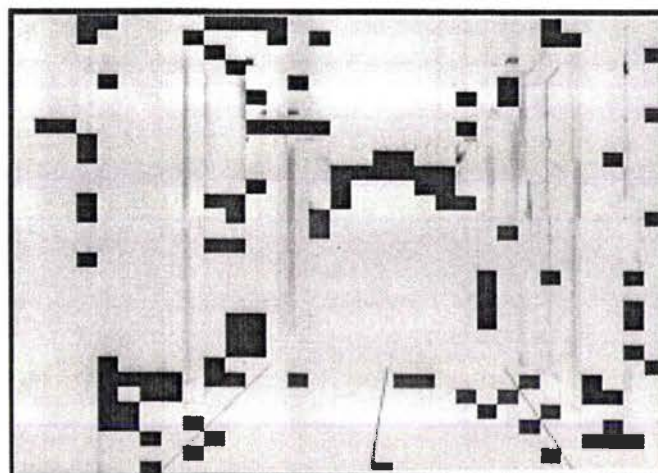
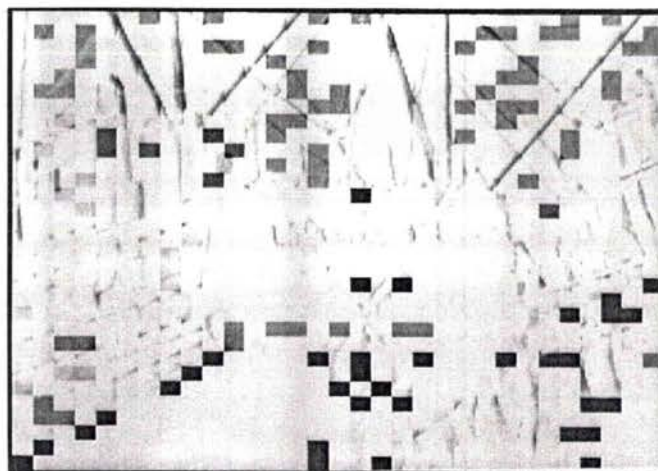
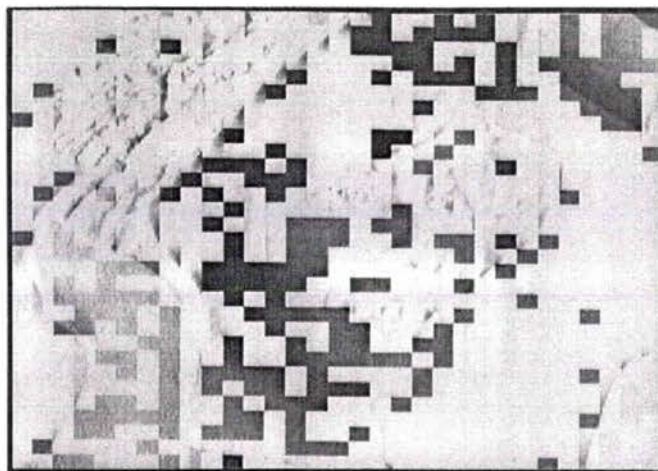
Série d'images N°4 : Blocs image dont l'activité est moyenne.



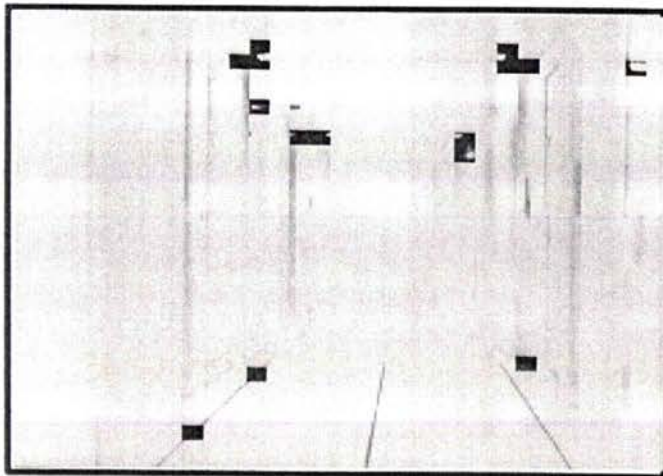
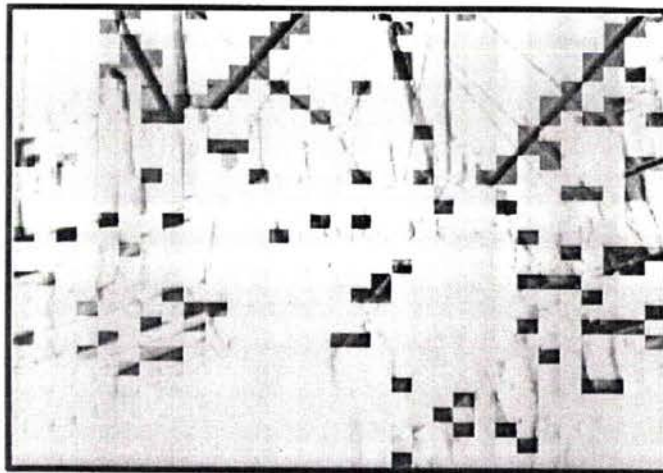
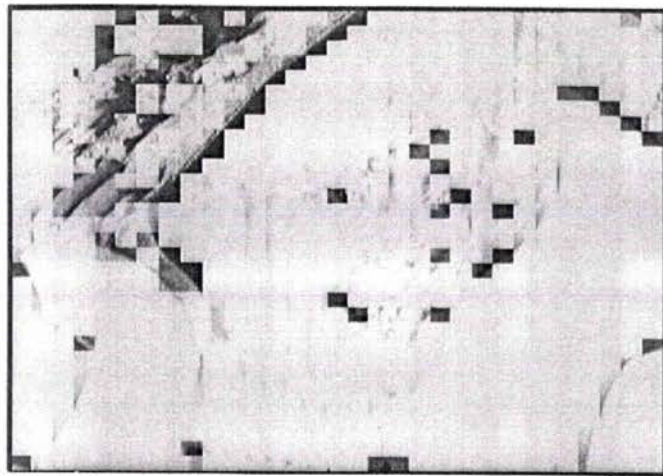
Série d'images N°5 : Blocs image dont l'activité spatiale est forte.



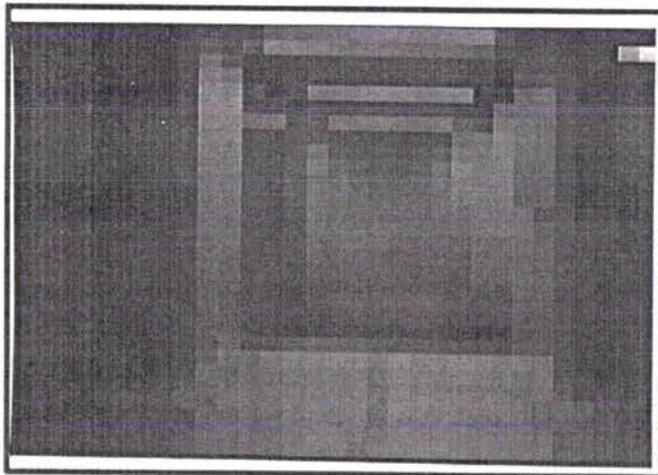
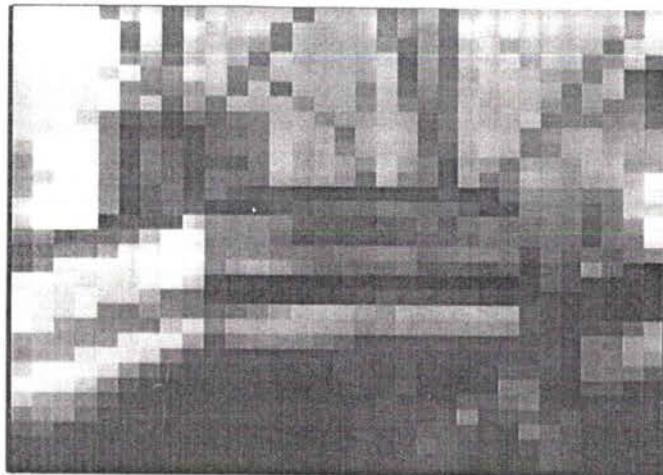
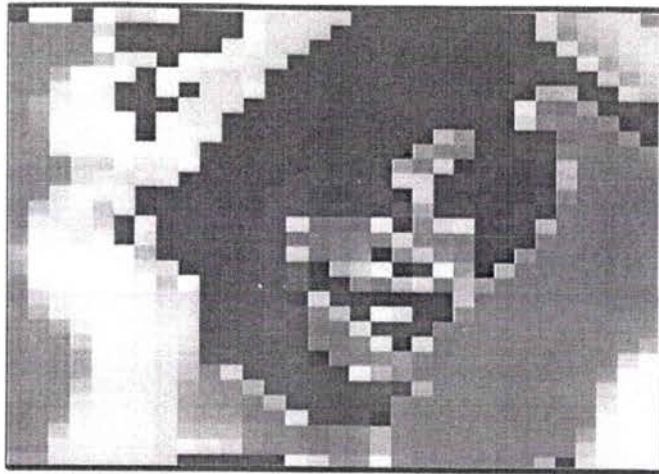
Série d'images N°6 : Blocs Imagecomposés de structures verticales ou horizontales.



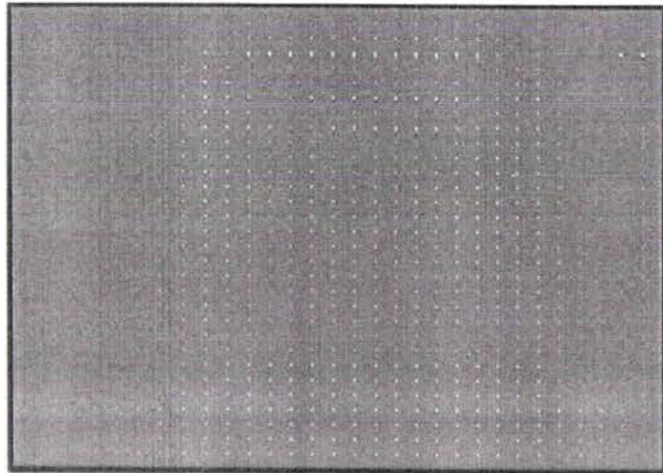
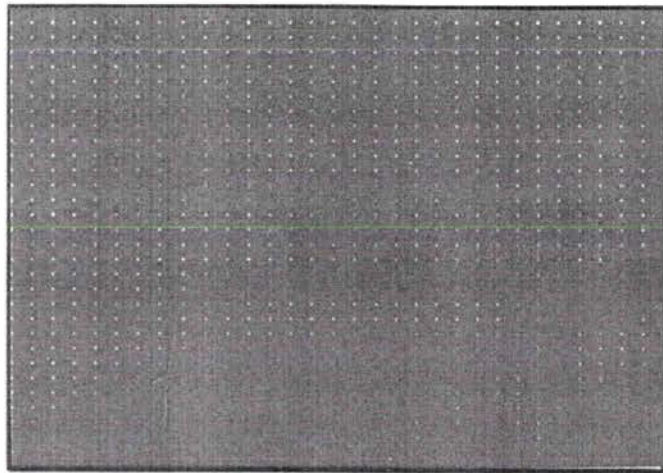
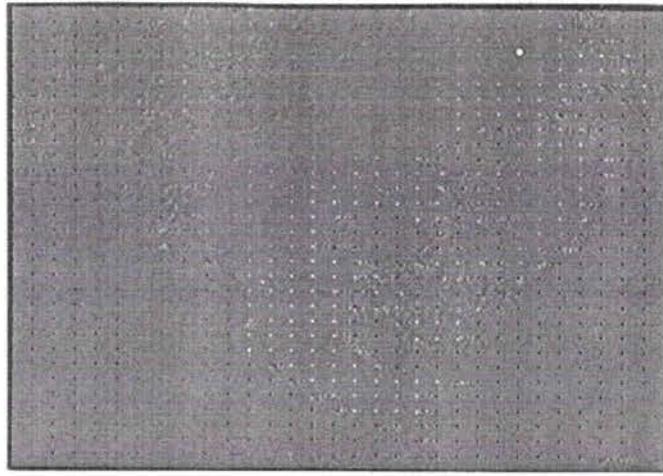
Série d'images N°7 : Blocs image symétriques par rapport à l'une de leurs diagonales.



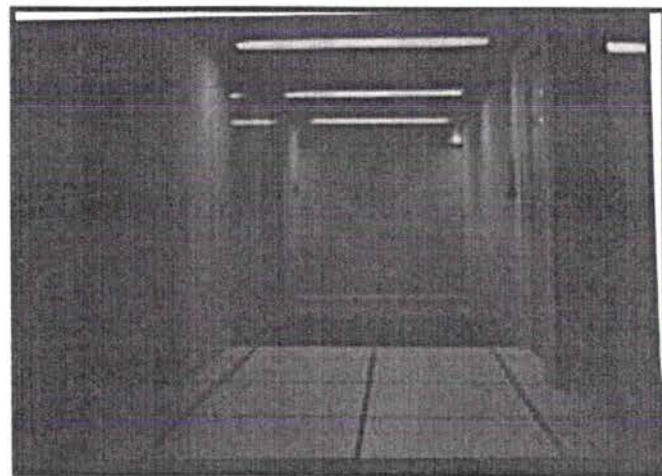
Série d'images N°8 : Blocs images dont la diagonale de la matrice transformée est importante.



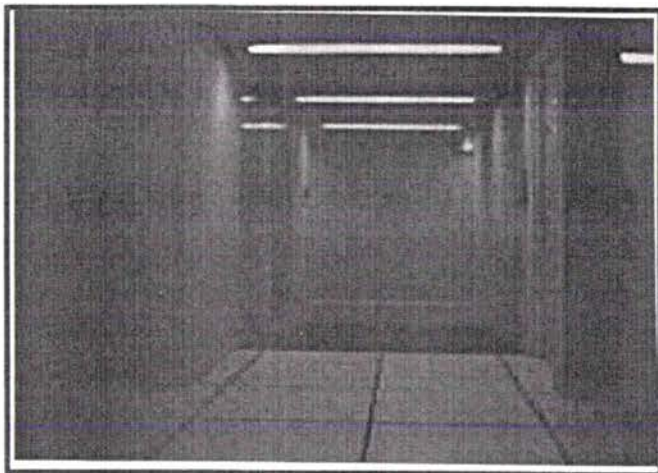
Série d'images N°9 : Blocs image représenté par leur moyennes.



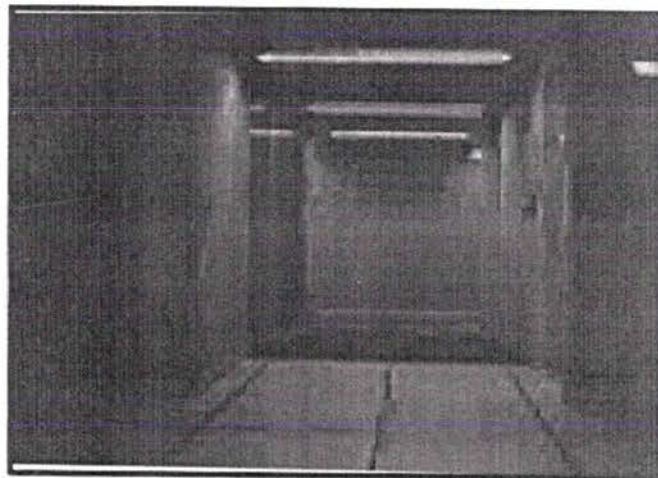
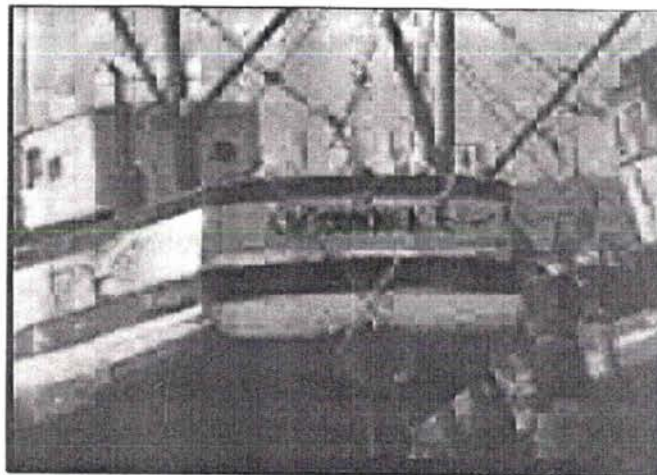
Série d'images N°10 : Images transformées en cosinus discrète.



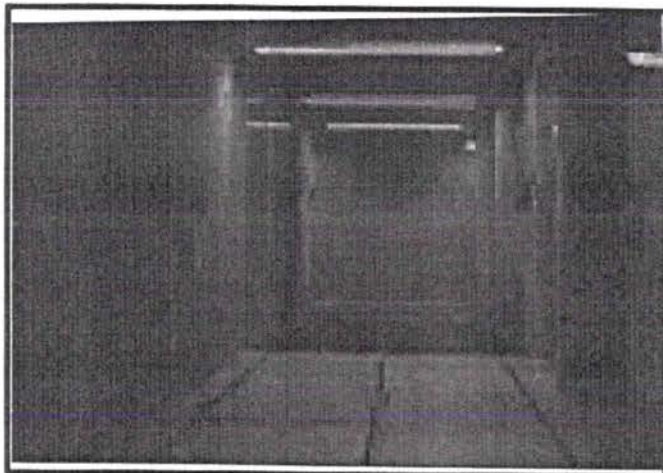
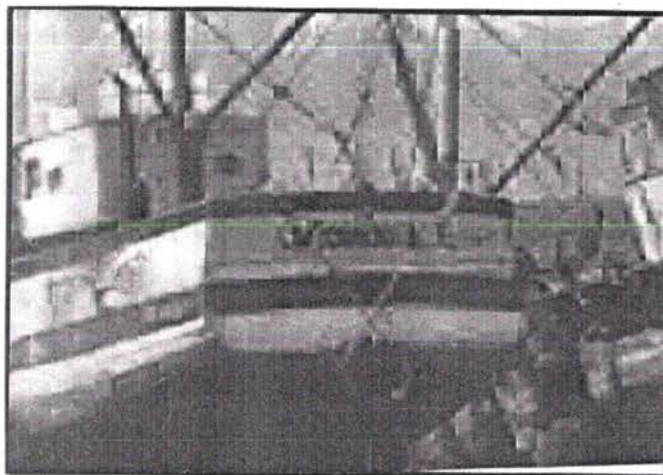
Série d'images N°11 : Images filtrées dans l'espace transformé.



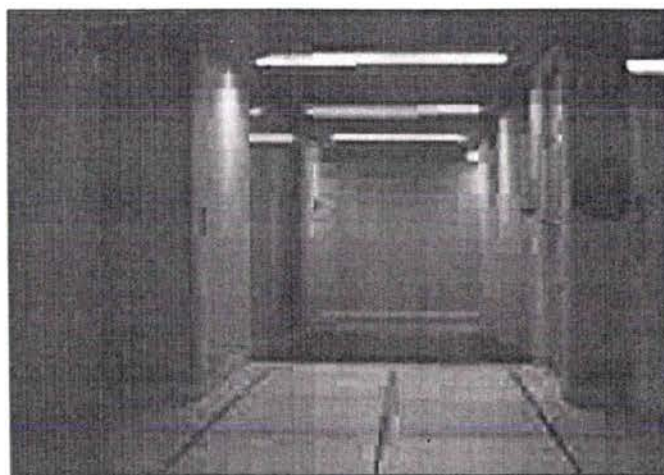
Série d'images N°12 :Images regroupées par classes de structures visuelles.



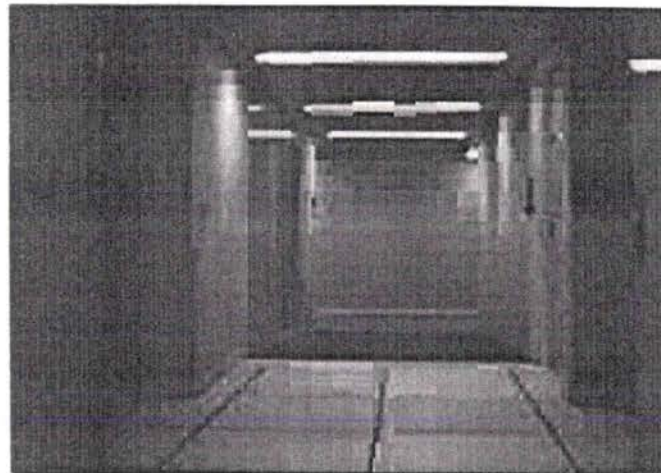
Série d'images N°13 :Images codées par quantification vectorielle sans classification, dictionnaire formé de 32 vecteurs.



Série d'images N°14 :Images codées par quantification vectorielle avec classification visuelle, avec deux dictionnaires de 15 vecteurs.



Série d'images N°15 :Images codées avec deux dictionnaires de 64 vecteurs.



Série d'images N°16 :Images codées avec deux dictionnaires de 256 vecteurs.

Blocs Symétriques par rapport à l'axe vertical central :

Pour tout pixel du bloc image $B_1(8 \times 8)$, la valeur de la luminance est $f_1(x,y)$; avec x et y coordonnées du pixel dans le bloc variables entre 0 et 7. Pour B_2 nous avons $f_2(x,y)$

Les luminances sont symétriques par rapport à l'axe central verticale implique que :

$$f_1(x,y) = f_2(x, 7 - y).$$

La formulation de la transformée en cosinus discrète étant :

$$F(u,v) = \frac{4.c(u).c(v)}{N^2} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x,y) \cos \frac{(2x+1)u\pi}{2N} \cdot \cos \frac{(2y+1)v\pi}{2N} \quad .(1)$$

$$\begin{aligned} c(u) &= 1/\sqrt{2} \quad \text{si } u = 0 \\ c(u) &= 1 \quad \text{si } u = 1, 2, \dots, N-1. \end{aligned}$$

$$K = \frac{4.c(u).c(v)}{N^2}$$

$f(x,y)$ représente la luminance du pixel au point (x,y) dans l'image.

$F(u,v)$ représente le coefficient transformé au point (u,v) .

avec u et v allant de 0 à $N-1$.

La transformée inverse est définie par :

$$f(x,y) = K \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} F(u,v) \cos \frac{(2x+1)u\pi}{2N} \cdot \cos \frac{(2y+1)v\pi}{2N} \quad .(2)$$

avec x et y allant de 0 à $N-1$.

Pour le bloc B_1 (1) devient :

$$\begin{aligned}
F_1(u,v) &= K \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f_1(x,y) \cos \frac{(2x+1)u\pi}{2N} \cdot \cos \frac{(2y+1)v\pi}{2N} \\
&= K \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f_2(x, N-1-y) \cdot A(x,u) \cdot \cos \frac{(2(N-1-y)+1)v\pi}{2N} \cdot (3).
\end{aligned}$$

$$A(x,u) = \cos \frac{(2x+1)u\pi}{2N}.$$

Avec un changement de variable $Y = N-1-y$, (3) devient :

$$\begin{aligned}
F_1(u,v) &= K \sum_{x=0}^{N-1} \sum_{Y=0}^{N-1} f_2(x,Y) \cdot A(x,u) \cdot \cos \frac{(2(N-1-Y)+1)v\pi}{2N} \\
&= K \sum_{x=0}^{N-1} \sum_{Y=0}^{N-1} f_2(x, N-1-y) \cdot A(x,u) \cdot \cos \left[v\pi - \frac{(2(Y+1))v\pi}{2N} \right] \cdot (4)
\end{aligned}$$

si v est pair; $v = 2p$ avec $0 \leq p \leq \frac{N}{2} - 1$.

$$\cos \left[(2p)\pi - \frac{(2(Y+1))v\pi}{2N} \right] = \cos \left[\frac{(2(Y+1))v\pi}{2N} \right]$$

L'équation (4) devient :

$$\begin{aligned}
F_1(u, 2p) &= K \sum_{x=0}^{N-1} \sum_{Y=0}^{N-1} f_2(x,Y) \cos \frac{(2x+1)u\pi}{2N} \cdot \cos \left[\frac{2(Y+1)2p\pi}{2N} \right] \\
&= F_2(u, 2p).
\end{aligned}$$

si v est impair; $v = 2p+1$ avec $0 \leq p \leq \frac{N}{2} - 1$.

$$\cos \left[(2p)\pi - \frac{(2(Y+1))v\pi}{2N} \right] = - \cos \left[\frac{(2(Y+1))v\pi}{2N} \right]$$

L'équation (4) devient :

$$\begin{aligned}
F_1(u, 2p+1) &= K \sum_{x=0}^{N-1} \sum_{Y=0}^{N-1} f_2(x, Y) \cos \frac{(2x+1)u\pi}{2N} \cdot \cos \left[\frac{2(Y+1)(2p+1)\pi}{2N} \right] \\
&= -F_2(u, 2p+1) .
\end{aligned}$$

Deux blocs B_1 et B_2 symétriques par rapport à l'axe vertical central sont liés dans l'espace transformé par :

$$F_1(u, 2p) = F_2(u, 2p) .$$

$$F_1(u, 2p+1) = -F_2(u, 2p+1) .$$

Blocs Symétriques par rapport à l'axe horizontal central :

La même démonstration est valable pour deux blocs B3 et B4 symétriques par rapport à l'axe horizontal central :

$$f_3(x, y) = f_4(N-1-x, y)$$

B3 et B4 sont liés dans l'espace transformé par :

$$F_3(2q, v) = F_4(2q, v) .$$

$$F_3(2q+1, v) = - F_4(2q+1, v) .$$

$$\text{avec } 0 \leq q \leq \frac{N}{2} - 1.$$

Bloc B symétrique par rapport à sa première diagonale :

$$f(x,y) = f(y,x) .$$

L'équation (1) devient :

$$\begin{aligned} F(u,v) &= \frac{4.c(u).c(v)}{N^2} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(y,x) \cos \frac{(2x+1)u\pi}{2N} . \cos \frac{(2y+1)v\pi}{2N} \\ &= F(v,u). \end{aligned}$$

La matrice transformée en cosinus discrète est aussi symétrique par rapport à sa première diagonale.

Bloc B symétrique par rapport à sa seconde diagonale :

$$f(x,y) = f(N-1-y, N-1-x)$$

$$F(u,v) = \frac{4.c(u).c(v)}{N^2} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(N-1-y, N-1-x) \cos \frac{(2x+1)u\pi}{2N} . \cos \frac{(2y+1)v\pi}{2N} \quad (1)$$

Un changement de variable s'impose ; $X = N-1-y$, $Y = N-1-x$.

$$A(x,u) = \cos \frac{(2x+1)u\pi}{2N}$$

$$B(y,v) = \cos \frac{(2y+1)v\pi}{2N}$$

$$A(x,u) = \cos \frac{(2(N-1-Y)+1)u\pi}{2N} = \cos \left[u\pi - \frac{(2Y+1)u\pi}{2N} \right]$$

$$B(y,v) = \cos \frac{(2(N-1-X)+1)v\pi}{2N} = \cos \left[v\pi - \frac{(2X+1)v\pi}{2N} \right]$$

Si u et v sont de même parité, le produit $A(x,u) . B(y,v)$ devient :

$$A(x,u) . B(y,v) = \cos \left[\frac{(2Y+1)u\pi}{2N} \right] . \cos \left[\frac{(2X+1)v\pi}{2N} \right].$$

L'équation (1) se traduira par :

$$F(u,v) = \frac{4.c(u).c(v)}{N^2} \sum_{Y=0}^{N-1} \sum_{X=0}^{N-1} f(X,Y) \cdot \cos \left[\frac{(2Y+1)u\pi}{2N} \right] \cdot \cos \left[\frac{(2X+1)v\pi}{2N} \right].$$

$$= F(v,u).$$

Dans le cas où u et v sont de parités différentes, le produit $A(x,u).B(y,v)$ devient :

$$A(x,u) \cdot B(y,v) = - \cos \left[\frac{(2Y+1)u\pi}{2N} \right] \cdot \cos \left[\frac{(2X+1)v\pi}{2N} \right].$$

L'équation (1) se traduira par :

$$F(u,v) = - F(v,u).$$

La supposition inverse est vraie et est démontrable. Si un bloc transformé présente les particularités suivantes :

$$\begin{aligned} F(2q, 2p) &= F(2p, 2q); \\ F(2q+1, 2p+1) &= F(2p+1, 2q+1); \\ F(2q, 2p+1) &= - F(2p+1, 2q); \\ F(2q+1, 2p) &= - F(2p, 2q+1); \end{aligned}$$

Alors le bloc image incident présente une symétrie par rapport à sa seconde diagonale :

$$f(N-1-y, N-1-x) = f(x,y).$$

Ce rapport présente un travail qui a été réalisé au Centre d'Etudes et de Recherche de Lorraine de TéléDiffusion de France (CERLOR) et au laboratoire Mécatronique Industrielle de l'Université de Metz.

Je suis très reconnaissant à Monsieur Didier Flaender, directeur du CERLOR, de m'avoir accueilli dans ce centre de recherche et d'avoir accepté d'être rapporteur de cette thèse.

Que Messieurs Hervé Caudron, directeur adjoint du CERLOR et responsable du laboratoire T.M.I. dans lequel j'ai fini mes travaux, et Philippe Jean, responsable du laboratoire T.E.C., qui m'a permis de commencer cette étude, soient remerciés pour les importants moyens et les facilités qu'ils ont mis à ma disposition. La confiance qu'ils me témoignent en siégeant en tant qu'invités à ce jury, me touche profondément.

Je souhaite que Monsieur le Professeur André Tosser-Roussey, professeur à l'Université de Metz et directeur du laboratoire Mécatronique Industrielle, reçoive ma profonde gratitude pour m'avoir intégré à son équipe de recherche et pour la confiance qu'il m'a accordée et qui se traduit aujourd'hui par la direction de ce jury.

Je suis très honoré de la participation de Monsieur Henri Maître, directeur du département Image à l'E.N.S.T., à mon jury et je le remercie très sincèrement pour ses conseils et pour son rapport.

Que Monsieur Michel Bruynoogh soit remercié d'être rapporteur de cette thèse.

Monsieur Vincent Thomas, chef de la cellule Codage d'Images au C.C.E.T.T., a bien voulu siéger dans ce jury, je l'en remercie vivement.

Je remercie Monsieur Claude Lhermitte, chef du service Informatique à Supélec, pour sa présence dans ce jury.

Monsieur Mustapha Bennouna, chef du département Traitement du Signal et Automatique à l'Ecole Mohammadia d'Ingénieurs, est venu de très loin

pour représenter un pays qui est cher à mon coeur, je l'en remercie beaucoup et j'espère que sa venue est l'aube d'une belle coopération entre le CERLOR et l'E.M.I.

Je remercie Monsieur Roland Kleim, Professeur à l'Université de Metz, de bien avoir voulu voulu siéger dans ce jury.

J'ai le vif plaisir de remercier Monsieur Arnaud Laprèvote avec lequel j'ai formé une équipe de choc tant pour la créativité que pour la bonne humeur.

Je désire que l'ensemble du personnel du CERLOR ainsi que chaque membre du laboratoire Mécatronique Industrielle soient remerciés pour l'excellente ambiance dans laquelle ce travail a été réalisé.

Les dispositifs technologiques et l'algorithme de traitement présentés dans ce mémoire sont cités dans deux demandes de brevets d'inventions, dont les intitulés sont :

- "***Une méthode et un système pour la mise au point automatique d'une caméra utilisant la TCD***". Brevet déposé le 11 avril 1991 au nom de TéléDiffusion de France S.A.. N°91 04420.
- "***Un procédé de codage de signaux d'image hybride adaptatif***". Brevet déposé le 24 juillet 1991 au nom de TéléDiffusion de France S.A.. N°91 09359.

"Entia non sunt multiplicanda praeter necessitatem * "
Guillaume d'Ockham philosophe franciscain 1270-1349.

*** Il ne faut pas multiplier les entités au-delà de ce qui est
nécessaire**

Chapitre I.....	1
INTRODUCTION.....	1
I - A Contraintes.....	1
I - B Solutions retenues	2
Chapitre II	6
TECHNIQUES CLASSIQUES DE CODAGE D'IMAGES	6
II - A Méthodes prédictives	6
II - A - 1 Le prédicteur des échantillons du signal image.....	8
II - A - 2 Le quantificateur de l'erreur de prédiction	8
II - A - 3 Défauts générés par la MICD	10
Le bruit granulaire	10
Le bruit de faux contour	10
Le bruit de traînage.....	10
Le flottement de contour.....	11
II - B Codage par transformations mathématiques	12
II - B - 1 Formulations mathématiques	12
II - B - 2 Transformée en cosinus discrète	14
Définition	14
Evaluation de la dynamique des composantes de la matrice transformée	15
II - B - 3 Codage par seuillage	17
II - B - 4 Images adaptées au codage par transformation	18
II - C Codage par quantification vectorielle	18
II - C - 1 Principe.....	18
II - C - 2 Formation des vecteurs.....	20
Techniques de normalisation pour la quantification spatiale	21
II - C - 3 Construction du dictionnaire.....	23
Nuées dynamiques	23
Algorithme de classification du dictionnaire LBG	25
Algorithme de Linde Buzo Gray.....	27
Méthode de Kohonen.....	28
II - C - 4 Distance déterminant la distorsion entre les vecteurs.....	29
Chapitre III.....	30
METHODES ADAPTATIVES DE CODAGE D'IMAGES	30

III - A Méthode prédictives	30
III - A - 1 MICDA-ER	31
III - B Techniques adaptatives de codage par transformation TCD	34
III - B - 1 Adaptation par variation de la taille des blocs transformés	34
III - B - 2 Stratégies adaptatives de codage de coefficients transformés	35
III - B - 3 Classification basée sur des paramètres énergétiques	35
III - C Approches adaptatives pour la quantification vectorielle.....	43
III - C - 1 Accommodation des tailles de bloc aux images codées	43
III - C - 2 Quantification vectorielle multi-étages	45
III - C - 3 Quantification vectorielle à dictionnaire rafraîchi	45
Chapitre IV	47
TECHNIQUES DE CODAGE PAR CLASSIFICATION VISUELLE	47
IV - A Quantification vectorielle	48
IV - B Codage par transformé en cosinus discrète.....	55
IV - C Codage Différentiel	58
Chapitre V	61
METHODES HYBRIDES DE CODAGE.....	61
V - A La quantification vectorielle appliquée dans le domaine transformé	61
V - B Codage hybride; adaptation visuelle.....	63
V - B - 1 Quantification vectorielle dans l'espace transformé.....	63
V - B - 2 Quantification vectorielle dans l'espace image.....	68
Classification visuelle	68
Résumé	70
Codage par transformée en cosinus discrète	71
Codage par quantification vectorielle.....	71
Chapitre VI.....	76
CORRESPONDANCE IMAGE - ESPACE TRANSFORME	76
VI - A Introduction.....	76

VI - B Correspondance pour les transformées mathématiques.....	77
VI - B Espace Transformé en Cosinus Discrète	77
VI - B - 1 Blocs homogènes	78
VI - B - 2 Transitions primaires	79
VI - B - 3 Manipulations du bloc image	80
Transformation qui donne à un pixel la valeur complémentaire à 255 de sa luminance.....	80
Symétrie par rapport à la médiane verticale	80
Symétries par rapport à la médiane horizontale.....	81
Symétrie par rapport à la première diagonale du bloc image	81
Symétrie par rapport à la seconde diagonale	82
Translation des luminances d'un bloc image	82
Chapitre VII	84
CODAGE PAR QUANTIFICATION VECTORIELLE DANS L'ESPACE TRANSFORME AVEC CLASSIFICATION VISUELLE.....	84
VII - A Classification des blocs image.....	84
VII - A - 1 Paramètres de classification.....	85
Activité spatiale totale	85
Activité partielle matérialisant les transitions verticales	86
Activité partielle correspondant aux transitions horizontales	86
Bloc présentant une symétrie par rapport à la première diagonale	87
Bloc image symétrique par rapport à la deuxième diagonale	87
Bloc image ayant une symétrie par rapport à son centre.....	88
VII - B Arbre de décision	88
VII - B - 1 Classes d'activité spatiale.....	88
VII - A - 2 Classes de composantes visuelles particulières.....	90
VII - A - 3 Formation de vecteurs de coefficients transformés	93
VII - C Quantification vectorielle des coefficients transformés.....	95
VII - C - 1 Distorsion entre deux vecteurs	96

VII - C - 2 Création du dictionnaire	101
VII - D Description du codage temporel	101
VII - D - 1 Estimation de mouvements	102
VII - D - 2 Codage de la composante continue par prédiction temporelle	103
VII - E Tests d'évaluation	103
VII - F Décodage rapide	104
Chapitre VIII	108
REALISATION MATERIELLE	108
VIII - A Modules de traitements parallèles	108
VIII - B Acquisition	111
VIII - C Le calculateur TCD	112
VIII - D Le codeur	114
VIII - D - 1 Description du schéma synoptique du système de transmission d'images animées	116
VIII - D - 2 Modules de codage	117
VIII - E Le décodeur	118
VIII - F Architecture du système de codage	119
Chapitre IX	120
PRESENTATION DES PERFORMANCES	120
IX - A Classification visuelle des blocs	121
IX - A - 1 Activité spatiale	121
IX - A - 2 Classification suivant les formes de transitions	122
IX - A - 3 Quantification vectorielle	125
IX - B Taux de compression	126
IX - C Comparaisons avec les techniques connues	127
IX - D Extensions et perspectives pour le procédé de codage	133
CONCLUSION	135
Liste des références	138
ANNEXES	142

Chapitre I

INTRODUCTION

L'image est devenue un support incontournable de communication. La richesse des informations qu'elle rassemble est sans comparaison; formes, couleurs et signes en déplacement. L'impact visuel que provoque l'image chez l'observateur n'a d'égal que la complexité de la symbolique qu'elle véhicule. En effet, les messages visuels sont les plus persuasifs et le plus riches. C'est une des principales raisons pour lesquelles nous constatons la multiplication d'équipements de communications visuelles. De nombreux services basés sur la transmission d'images ont vu le jour : la vidéoconférence, le visiophone, le photo-vidéotex, l'audio-vidéographie, le multimédia etc..

Ces nouveaux modes de communication nécessitent la création, la manipulation, l'archivage et l'échange d'images. Ils nécessitent aussi l'exploitation des voies numériques de transmissions nationales et internationales. L'extension du marché potentiel de ces services passe par la minimisation de leur prix de revient. Il s'agit de baisser le coût des terminaux, ainsi que le temps d'occupation des canaux de transmission.

I - A Contraintes

Les systèmes de communication basés sur l'image sont soumis à de sévères contraintes découlant de la nature même des images et des scènes qu'elles représentent.

L'image représentée numériquement est composée d'une matrice de pixels. Ces éléments picturaux représentent par les valeurs de leurs

luminances les différentes structures ou formes contenues dans l'image. Le volume des informations constitué par une image noir et blanc fixe de (512 x 512) est de 256 Koctets, il requière plus de 32 secondes de temps de transmission sur le Réseau Numérique à Intégration de Services (le RNIS possède deux canaux à 64 Kbits/s de débit). Pour les séquences animées la quantité d'informations est multipliée par le nombre imposé d'images par seconde.

Tout système de transmission d'images doit donc supporter une cadence soutenue, acquérir, traiter, transmettre et visualiser des séquences d'images numérisées dans un temps raisonnable. La notion de vitesse, ici, est liée à la capacité de l'équipement à traiter le flot de données imposé par le service étudié, pour la télévision 25 images par seconde sont nécessaires, mais pour le visiophone quelques images par seconde suffisent.

I - B Solutions retenues

La quantité d'information contenue dans une image est telle qu'elle pose de très sérieux problèmes. Les conséquences sont ressenties sur le temps de traitement et de transmission de l'image et sur la taille mémoire nécessaire pour son stockage.

Afin d'extraire les informations pertinentes et de réduire la quantité globale des données, nous appliquons à l'image une compression ou un codage. Depuis que la numérisation est technologiquement possible, plusieurs techniques de codage existent. Elles exploitent la représentation matricielle des pixels de l'image, la redondance et la corrélation entre les pixels voisins. Mais ces méthodes restent étroitement liées aux applications auxquelles elles sont destinées.

Pour les systèmes de transmission d'images, la contrainte essentielle est leur rapidité de traitement en temps réel et leur capacité à supporter une cadence soutenue d'informations. Ce qui impose souvent une limitation au niveau de la complexité de l'algorithme de compression utilisé en vue d'une implémentation matérielle.

Trois aspects comparatifs sont exploités pour les codeurs d'images :

- Coût des équipements de codage - décodage.**
- Gain en débit.**
- Qualité subjective des image restituées à la réception.**

A la réception la restitution de l'image initiale est faite. Cette reconstruction doit être d'une grande fidélité et doit présenter un minimum de dégradations perceptibles par l'oeil de l'observateur.

Le taux de défauts tolérés dans les images reconstituées est indissociable de l'application étudiée. Une déformation d'un contour sur une radiographie archivée est grave de conséquences, donc intolérable. Cependant, la disparition des pois sur la cravate d'un interlocuteur par visiophone est tout à fait acceptable.

Les techniques de compression d'images répondent chacune à des contraintes particulières imposées par l'application étudiée. Le choix de la méthode de compression, la mieux adaptée nécessite donc une connaissance préalable du type d'images transmises.

Nous avons effectué une investigation pour identifier les techniques optimales de codage. La démarche scientifique prise nous a permis de faire le tour des méthodes connues et de déterminer leurs particularités propres, de façon à les exploiter de la meilleure manière possible.

Nous avons étudié en première partie les trois techniques classiques de codage :

- Codage prédictif différentiel,**
- Codage par transformations mathématiques,**
- Codage par quantification vectorielle.**

A la suite de cette étude nous nous sommes intéressé aux méthodes adaptatives. Plusieurs approches algorithmiques concernant des adaptations particulières des techniques classiques sont citées dans la littérature. Pour décrire l'état de l'art en matière d'adaptabilité, nous avons choisi de décrire ces différentes formes en détaillant les techniques d'adaptation les plus originales.

Tenant compte des recommandations des spécialistes en ce qui concerne les extensions algorithmiques, il nous a paru avantageux d'approfondir notre connaissance des adaptations visuelles. Celles ci sont fondées sur l'extraction des structures perceptibles contenues dans l'image. Le codage visuel identifie les informations pertinentes en localisant les composantes de l'image qui agissent sur la qualité subjective après reconstitution des images décodées.

A la suite de cette recherche permettant une meilleure connaissance des algorithmes de compression d'images, nous avons imaginé un nouveau procédé de codage.

Notre méthode est une adaptation visuelle d'une technique de codage hybride exploitant une transformation mathématique et la quantification vectorielle. L'objectif initialement fixé était la proposition d'une méthode de codage qui offrirait des performances supérieures à celles offertes par les méthodes connues.

Notre proposition présente une approche originale permettant d'optimiser les différentes étapes du codage - décodage. Elle est fondée sur l'exploitation de la signification visuelle des composantes de l'image.

Il s'agit d'extraire le maximum d'informations visuelles contenues dans l'image. Les structures identifiées sont modélisées et leurs impacts subjectifs sur l'observateur déterminent les seuils de visibilité. Cette étude préliminaire a abouti à une série de constatations jamais identifiées ni utilisées dans le codage d'images. L'exploitation de ces caractéristiques visuelles a amélioré de beaucoup le taux de compression ainsi que la qualité subjective des images décodées.

Le procédé que nous avons conçu présente plusieurs caractéristiques avantageuses, la principale étant la simplicité des traitements nécessaires pour le codage - décodage. Elle agit directement sur la complexité des équipements.

Pour réaliser un système électronique exécutant notre méthode de codage, nous avons opté pour un équipement modulaire, évolutif et d'une grande puissance de calcul. Notre choix s'est fixé sur un ensemble de processeurs de traitements de signal, doublés de processeurs parallèles de type : transputer. Cette configuration ouverte nous a permis d'améliorer les performances en temps d'exécution de notre technique de codage.

Ce mémoire présente le même cheminement que celui de nos travaux de recherche. Nous y détaillons dans le chapitre II l'essentiel des techniques de codage classiques. Dans le chapitre III nous avons exposé des exemples de procédés adaptatifs de codage d'images. Les méthodes utilisant des adaptations visuelles sont rassemblées dans le chapitre IV. Dans le chapitre V nous avons décrit quelques procédés hybrides. Ces techniques exploitent les avantages de plusieurs méthodes classiques pour améliorer les performances de codage.

Pour mettre en valeur les constatations sur les correspondances entre l'espace Image et l'espace Transformé, nous les avons exposées dans le chapitre VI.

Un développement de notre procédé de codage d'image, fondé sur l'utilisation de la signification visuelle des composantes de l'image, est effectué dans le chapitre VII. Le système de codage - décodage que nous avons développé est décrit dans le chapitre VIII. Cet équipement a été conçu pour pouvoir tester plusieurs types de codage et pour exécuter la nouvelle méthode de compression que nous avons développée.

Chapitre II

TECHNIQUES CLASSIQUES DE CODAGE D'IMAGES

La bonne utilisation des différentes techniques de codage passe par une maîtrise de leurs particularités. Il nous a semblé important de traiter les méthodes classiques de compression d'images, de manière à identifier leurs avantages respectifs et leurs domaines d'application.

II - A Méthodes prédictives

Une grande partie de l'information véhiculée par l'image est dite redondante. Son absence dans l'image n'altère cette dernière que d'une façon superficielle, sans pour autant porter atteinte à son contenu informatif. La nature de l'image noir et blanc numérisée; matrice composée d'éléments picturaux ou pixels, chacun d'eux étant codé par sa luminance, et le contenu de l'image composé de plages uniformes ou de formes font que les variations entre les valeurs de luminance des pixels voisins sont faibles.

La nature matricielle de la représentation de l'image et la dépendance existant entre chacun des pixels et ses voisins traduisent une corrélation très forte sur l'image. Cette corrélation fait qu'on ne peut traiter un pixel sans se soucier des valeurs de luminance de ses voisins.

Le principe de base des techniques de codage prédictives consiste à exploiter la redondance existant entre les valeurs du signal à transmettre. Il

repose sur la constatation de la faible valeur de la différence entre les luminances de deux pixels voisins ou se succédant sur une séquence d'images.

Une prédiction du signal à coder est réalisée au niveau de l'émetteur à chaque instant. Afin d'exploiter la corrélation existante, l'écart entre la valeur réelle de la luminance et la valeur de prédiction est quantifié et transmis au récepteur sous forme de code numérique (cf. Fig. N°1).

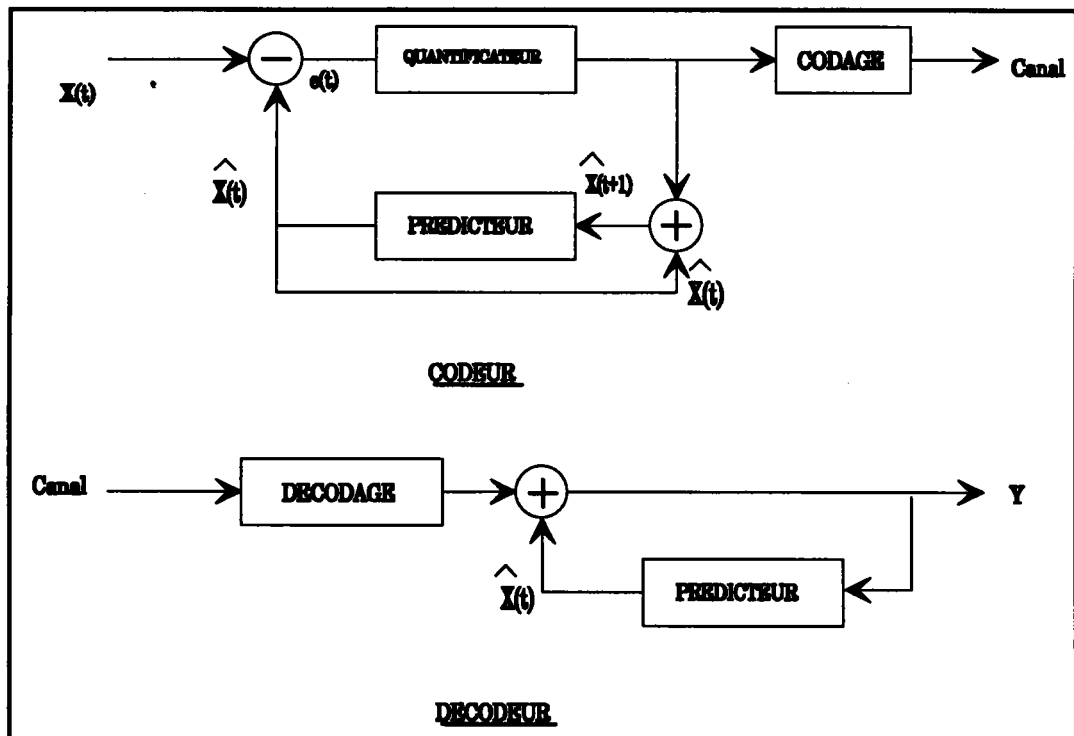


Figure N°1 Codage par prédiction différentielle

Dans le récepteur, en rajoutant la valeur de l'erreur décodée à une valeur de prédiction calculée de la même manière qu'à l'émission, c'est à dire à partir des échantillons du signal correspondant aux valeurs passées et déjà reçues au récepteur, on peut reconstituer le signal image, aux erreurs près de quantification et de transmission (cf. Fig. N°1).

Les systèmes de codage basés sur la MICD : Modulation par Impulsions Codées Différentielles, ou en Anglais DPCM : Differential Pulse Coding Modulation, sont donc essentiellement constitués par un prédicteur du signal

et un quantificateur de l'erreur de prédiction.

Nous allons maintenant détailler les différents éléments de cette chaîne de codage. En mettant en valeur les différents problèmes que pose chacun d'eux.

II - A - 1 Le prédicteur des échantillons du signal image

L'estimateur effectue une prédiction de chaque élément à transmettre (point, bloc, image...). Il n'utilise que des informations déjà obtenues et transmises, ce qui permet au décodeur de recalculer et de reconstituer le signal image initial [GUIC 86]. Le prédicteur optimal est celui qui calcule la valeur de prédiction qui s'approche le mieux de la valeur réelle du signal à coder.

La plupart des systèmes de codage d'images différentiels utilisent la valeur, plus ou moins pondérée, du point précédent de la même ligne comme valeur simple de prédiction. D'autres prédicteurs donnent une valeur en fonction des luminances des pixels voisins du point courant.

Nous pouvons répertorier les prédicteurs dans deux classes. Ceux qui exploitent les valeurs de luminance des voisins directs du pixel à coder sur l'image et ceux qui utilisent les valeurs de luminance des pixels de l'image précédente. Les deux catégories exploitent respectivement la redondance spatiale et la redondance temporelle du signal image numérisé.

II - A - 2 Le quantificateur de l'erreur de prédiction

Le gain en débit du MICD résulte principalement de la quantification du signal d'erreur de prédiction. La précision nécessaire pour le codage du signal d'erreur transmis est variable. La visibilité du bruit de quantification varie avec le contexte local : dans les zones de contraste élevé, l'erreur de prédiction est importante puisque la différence entre la valeur réelle du signal et la valeur estimée est importante; au contraire les erreurs de prédiction faibles se produisent dans les zones à faibles évolutions de

luminance [KRET 77].

Le quantificateur est l'organe du système dans lequel est réalisé la compression. En effet, il réduit la dynamique totale du signal d'écart de prédiction en zones délimitées par des seuils de décision. Le quantificateur donne à chaque intervalle de valeurs d'erreur de prédiction, présenté à son entrée, une valeur de quantification. Cette valeur sera codée et transmise au décodeur. Pour chaque quantificateur, une caractéristique détermine les zones, les seuils et les valeurs de quantification (cf. Fig. N°2).

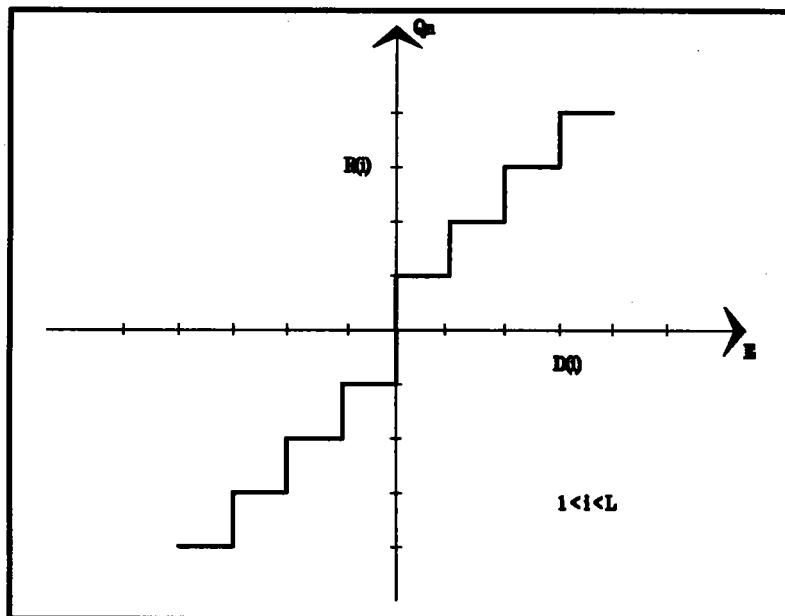


Figure N°2 Caractéristique d'un quantificateur $2L$ niveaux.
D domaine de décision, R niveau de quantification

La conception du prédicteur fait intervenir uniquement des propriétés statistiques du signal source. Celle du quantificateur fait appel aux propriétés de l'observateur, ce qui explique l'absence de méthode rigoureuse pour définir le quantificateur optimum.

Après la quantification, le codeur alloue un certain nombre de bits pour chaque valeur quantifiée de l'erreur de prédiction. Ce nombre est étroitement lié d'une part à l'efficacité de l'estimateur et du quantificateur et à leur capacité à minimiser les erreurs, d'autre part à la nature du signal image et aux évolutions temporelles et spatiales que présentent ses échantillons.

La compression se traduit par le codage sur un nombre limité de bit des erreurs de prédiction. Avec un estimateur optimal l'erreur reste faible.

II - A - 3 Défauts générés par la MICD

En l'absence d'erreurs de transmission, les systèmes basés sur la MICD sont susceptibles de générer un certain nombre de défauts dans l'image restituée au décodage. Du point de vue de l'observateur, ces dégradations se manifestent sous des apparences diverses. Elles ont cependant une origine commune : les erreurs de quantification.

Nous nous proposons de faire une liste non exhaustive de ces bruits de codage et de leurs causes, dans le but de mieux cerner les particularités des systèmes de codage MICD.

Le bruit granulaire

Il est visible principalement dans les zones uniformes de l'image. Lorsque l'évolution de luminance des pixels n'est pas significative, les valeurs successives de quantification de l'erreur de prédiction présentent une alternance dans leur signe. Ainsi, les valeurs de quantification oscillent autour du signal image. Sur l'image reconstituée apparaissent, dans des zones qui sont initialement uniformes, des groupes de pixels dont les luminances erronées laissent penser que nous sommes en présence de fortes variations du signal image.

Le bruit de faux contour

Ce bruit peut apparaître sur des plages de l'image où la luminosité varie de manière régulière et lente. Il est avant tout l'effet de l'erreur provoquée par la quantification, lors du réajustement de la valeur quantifiée par un passage à une valeur plus importante. Du point de vue de l'observateur, apparaît une transition de luminances inexistante dans l'image initiale.

Le bruit de traînage

Le "slope overload" est dû à une saturation du quantificateur lorsque

le signal subit des variations brutales et successives : les erreurs de prédiction dépassent les valeurs maximales de quantification. Le bruit de traînage est visible pendant le temps d'adaptation que met le quantificateur pour minimiser l'écart entre les valeurs en entrée et celles de quantification [PONC 74].

Le flottement de contour

Celui-ci se traduit par un sautiller ou mouvement des contours. Il est surtout gênant quand les transitions sont marquées. Les causes de ce bruit sont celles du bruit granulaire combinées avec celles du faux contour.

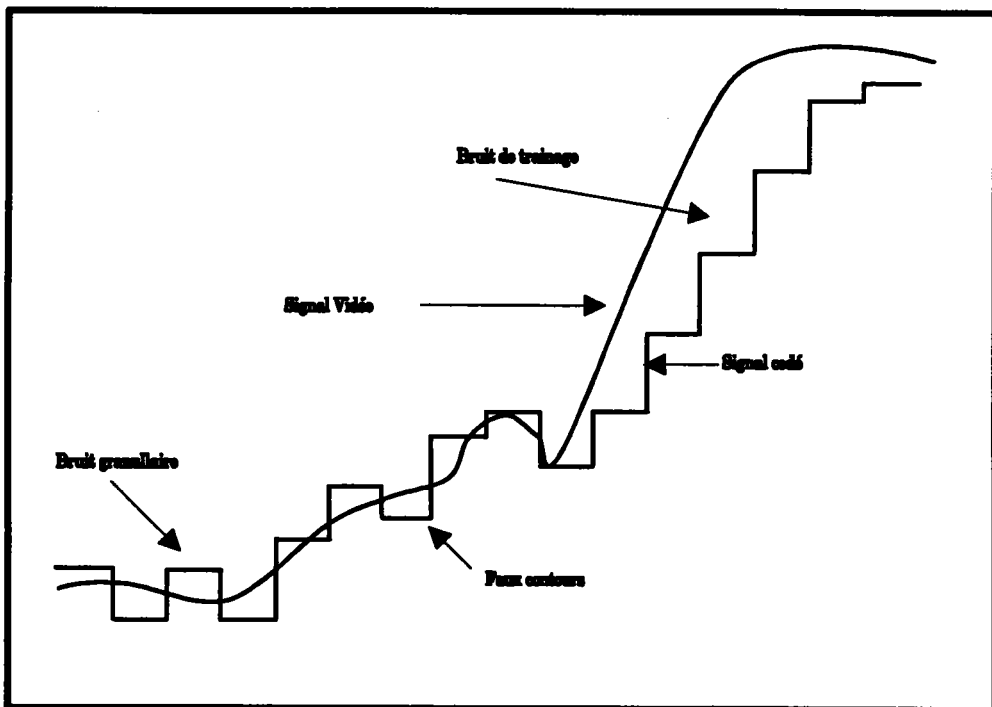


Figure N°3 Bruits de quantification identifiés dans le codage MICD

Rappelons que les techniques classiques de Modulation par Impulsions Codées Différentielles sont efficaces pour un signal image très redondant. En effet, les signaux dont les évolutions temporelles et/ou spatiales sont petites, permettent d'obtenir des prédicteurs optimaux et des erreurs d'estimation minimales. Pour cette catégorie d'images, la technique MICD permet d'obtenir une bonne qualité subjective et un bon taux de compression. Mais, pour des images contrastées et d'une grande activité temporelle et/ou spatiale, d'autres systèmes de codage sont mieux adaptés.

II - B Codage par transformations mathématiques

La représentation matricielle de l'image numérisée est source d'une importante redondance des valeurs des luminances des pixels. Elle est aussi génératrice d'une fonction d'interdépendance entre ces même valeurs. Cette représentation traduit la redondance de l'image par des zones plus ou moins étendues où la valeur de la luminance varie peu ou reste uniforme.

Plusieurs techniques de codage contournent cette représentation, en appliquant une transformation sur le plan de l'image. Cette opération est effectuée dans le but d'obtenir une meilleure décorrélation de représentants de l'image et une minimisation de la redondance (cf. Fig. N°4).

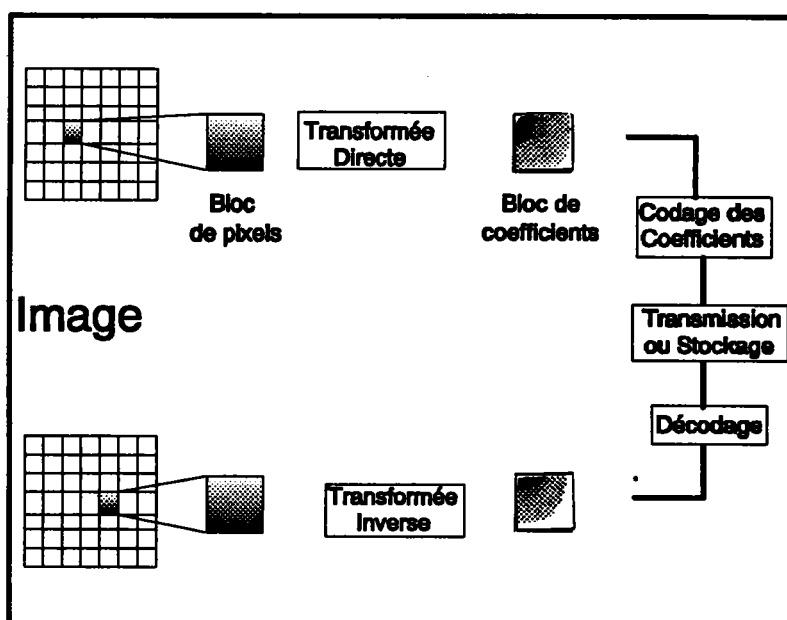


Figure N°4 Codage par transformation mathématique

Afin d'obtenir une meilleure réduction de l'information, plusieurs transformations mathématiques sont utilisées. Certaines étaient connues et répertoriées bien avant que n'existe l'image numérique. Elles ont, pour la plupart, fait leurs preuves en codage du signal mono-dimensionnel [RICH 86].

II - B - 1 Formulations mathématiques

L'image initiale est découpée en médaillons habituellement carrés de

dimension ($N \times N$) (cf. Fig. N°4). La luminance de chaque pixel est une fonction de deux dimensions suivant ses coordonnées : $f(x,y)$, [x et y sont comprises entre 0 et $N-1$].

La transformée mathématique directe et discrète est définie par l'expression linéaire:

$$F(u,v) = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x,y) a(x,y,u,v) . \quad (1 - a)$$

et la transformation inverse :

$$f^*(x,y) = \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} F(u,v) b(x,y,u,v) . \quad (1 - b)$$

Les fonctions a et b sont prises séparables suivant (x,u) et (y,v) .

Il est possible d'avoir une approche matricielle des transformations; les fonctions déterminées précédemment sont matérialisées par des matrices ($N \times N$) de dimensions deux.

La relation (1 - a) s'écrit alors:

$$[F] = [A] [f] [A]^t. \quad (2 - a)$$

$$[f^*] = [B] [F] [B]^t. \quad (2 - b)$$

si l'on prend $[B] = [A]^{-1}$ transformation inverse.

Dans le cas où la transformation $[A]$ est orthogonale et symétrique nous avons :

$$[A]^{-1} = [A]^t = [A].$$

si l'on prend $[B] = [A]^{-1}$, l'expression (2 - b) devient

$$[f^*] = [A] [F] [A] = [f]. \quad (3)$$

On remarquera que les mêmes opérations sont appliquées pour le

calcul de la transformation directe et inverse. Dans la mesure où les calculs sont faits avec la précision nécessaire et où aucune discrimination de coefficients transformés n'est effectuée, la double transformation directe - inverse n'introduit aucune dégradation sur l'image restituée [PONC 74].

Les principales transformations qui ont été utilisées dans le domaine de codage d'images sont les transformées de Fourier, de Haar, de Hotelling ou Karhunen-Loève, Walsh, Hadamard et pour finir la transformée en Cosinus Discrète (TCD).

Les performances offertes par la TCD ont fait que les autres transformées ont été écartées pour le codage d'images.

Une autre transformation qui a un avenir prometteur - dans le domaine des traitements d'images en général et la compression en particulier - est la transformée en ondelettes [MATH 90].

II - B - 2 Transformée en cosinus discrète

Définie en 1974 par Ahmed et al [AHME 74], la transformée en cosinus discrète suscite actuellement un intérêt remarquable.

Les propriétés mathématiques que présente la TCD : orthogonalité, parité et la simplicité des calculs qu'elle nécessite, lui confère une robustesse et des performances qui manquent aux autres transformées mathématiques.

L'essor technologique qu'a déclenché la TCD traduit la compétition ouverte que se livrent les industriels pour conquérir le grand marché de la communication par l'image.

Définition

A partir d'un bloc de $(N \times N)$ pixels corrélés, par une transformée en cosinus discrète nous obtenons un bloc de $(N \times N)$ coefficients transformés. La transformée s'exprime par la définition suivante:

Pour u et v de 0 à N-1 :

$$F(u,v) = \frac{4.c(u).c(v)}{N^2} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x,y) \cos \frac{(2x+1)up}{2N} \cdot \cos \frac{(2y+1)vp}{2N} \quad .(1)$$

$$c(u) = 1/\sqrt{2} \text{ si } u = 0$$

$$c(u) = 1 \text{ si } u = 1, 2, \dots, N-1.$$

$f(x,y)$ représente la luminance du pixel au point (x,y) dans l'image.

$F(u,v)$ représente le coefficient transformé au point (u,v) .

avec u et v allant de 0 à N-1.

La transformée inverse est définie par :

$$f(x,y) = \frac{4.c(u).c(v)}{N^2} \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} F(u,v) \cos \frac{(2x+1)up}{2N} \cdot \cos \frac{(2y+1)vp}{2N} \quad .(2)$$

avec x et y allant de 0 à N-1.

Evaluation de la dynamique des composantes de la matrice transformée

La distribution moyenne de l'énergie dans le domaine transformé (TCD) est concentrée sur un ensemble d'éléments transformés réduits. En effet, la répartition statistique de l'amplitude des fréquences spatiales $F(u,v)$ dépend fortement du rang (u,v) . Elle décroît fortement quand (u,v) s'éloigne de $(0,0)$.

Mais le passage dans le domaine transformé ne provoque pas à lui seul une réduction de l'information de l'image. Si le nombre des fréquences représentatives est restreint, leur dynamique, elle, est généralement supérieure à celle des luminances des pixels.

Ainsi, si la dynamique des luminances des pixels composant l'image est $[0, 255]$. La dynamique de l'élément $F(0,0)$ de la matrice transformée est égale à : $[0, 8 \times 8 \gamma(0,0) \times 255]$. La dynamique des autres éléments de la matrice transformée $F(u,v)$ avec (u,v) différent de $(0,0)$ est égale à $[- \gamma(u,v) \times 8 \times 8 \times 255, \gamma(u,v) \times 8 \times 8 \times 255]$, où $\gamma(u,v)$ est un facteur dépendant du rang de l'élément $F(u,v)$. Par exemple $\gamma(u,v)=0,676$ si: $(u,v)=(6,4)$ [CLAR 85].

Le codage classique par transformée en cosinus discrète est essentiellement basé, pour la réduction des informations de l'image, sur la quantification scalaire des fréquences transformées. Cette opération n'est efficace que si le quantificateur élaboré est optimal. En effet, si l'allocation du nombre de bits nécessaires pour coder un coefficient n'est pas adapté à sa dynamique, alors la réduction d'informations ne sera pas maximale.

Les coefficients transformés sont, en premier lieu, comparés à un seuil prédéterminé en fonction de la qualité désirée. Cela permet d'éliminer les valeurs non nécessaires à la restitution. Les coefficients restants sont quantifiés individuellement en parcourant le bloc transformé du haut à gauche vers le bas à droite, suivant un zigzag. Les fréquences sont codées successivement des plus basses aux plus hautes.

Sachant que la densité moyenne de l'énergie est concentrée autour de la composante continue, le quantificateur se limite aux zones de coefficients transformés représentatifs.

Le codeur quantifie chaque valeur non nulle de coefficient TCD. De plus, il code chaque plage de valeurs nulles par un mot code de longueur variable donné par une table de codage entropique de Huffman. Le nombre de bits alloués à la représentation de chaque bloc est non seulement fonction du nombre de coefficients non nuls, mais aussi de la fréquence d'apparition de chaque coefficient et de l'étendue de la plage des zéros dans la matrice transformée.

Chaque bloc transformé est alors codé par une suite de valeurs matérialisant les codes des coefficients transformés en cosinus discrète, ainsi que les positions des plages de zéros. Le quantificateur des coefficients utilise souvent un mot code de taille fixe. La longueur du mot code total affecté à chaque matrice transformée est variable. Elle dépend étroitement de la distribution spectrale obtenue à la suite de la transformation appliquée au bloc image initial, donc des variations de luminance des pixels qui le composent.

Rappelons la correspondance établie entre les évolutions des niveaux de gris des pixels et la distribution de leurs coefficients transformés. Pour un bloc d'image contenant des variations de luminances limitées, les coefficients transformés les plus représentatifs sont concentrés autour de la composante continue; $F(0,0)$ matérialisant la moyenne des luminances des pixels. Pour les images très fortement contrastées, celles dont les évolutions des

luminances sont brutales et nombreuses, le spectre énergétique et la dynamique des coefficients transformée sont étendus. En effet, de nombreux coefficients situés dans les hautes fréquences sont pertinents et représentent, avec la composante continue, l'essentiel de l'information véhiculée par l'image initiale.

Après avoir étudié la dynamique des coefficient, l'importance du spectre énergétique des blocs transformés et leurs signification par rapport à l'image initiale et aux évolutions de luminances qu'elle présente, nous nous proposons de donner un exemple simple de codage.

II - B - 3 Codage par seuillage

Une technique proposée dans [KHAL 88], consiste à appliquer un seuillage discriminant. Les coefficients dont la valeur absolue ne dépasse pas le seuil prédéfini sont négligés. La détermination de la valeur limite est un compromis entre la quantité d'information retenue pour représenter l'image et la qualité subjective de cette dernière.

La transformée en Cosinus Discrète est appliquée à des blocs (8x8). Une qualité satisfaisante a été obtenue pour un taux de compression moyen de 8. Elle a nécessité un seuil sélectif de 2, évalué par expérimentations successives.

Pour coder les coefficients dont la valeur absolue est supérieure à 2 sur un octet, une manipulation de la formulation mathématique est effectuée. La dynamique des coefficients se présente :

$$F(0,0) : [0,510];$$

$F(u,v) : [-254, 254]$ avec u,v variant de 0 à 7. pour u et v différents de zéro.

Une division par deux diminue l'étendue du domaine des valeurs. La transformation inverse réajuste les valeurs reçues par un facteur multiplicatif.

Les coefficients obtenus par la nouvelle transformée sont quantifiés et codés, sur un octet non signé pour $F(0,0)$ et sur un octet signé pour les autres coefficients transformés. La position de chaque coefficient dans la matrice TCD est codée sur un octet. Cette méthode permet de parcourir le domaine transformé à partir du coefficient $F(0,0)$ jusqu'au dernier coefficient pertinent. L'octet nul déterminant l'adresse de la composante continue

marque le début d'un nouveau bloc codé.

Cette technique, expérimentée sur plusieurs images, donne des taux de compression variant de 3 à 8. Le même procédé appliqué à des blocs d'images (16 x16) pixels offre de taux entre 6 et 10.

II - B- 4 Images adaptées au codage par transformation

Pour le codage par transformation TCD, une erreur qui modifie un coefficient se trouve répartie au niveau de tous les pixels du bloc d'image restitué. L'erreur de codage est donc "diluée" spatialement, ce qui est subjectivement favorable. Les déformations sur l'images sont plus acceptables si l'amplitude de l'écart dépasse la limite de la tolérance et si l'erreur est commise sur le coefficient représentant la composante continue, son effet sera visible sur tout le bloc image reconstituée.

Dans d'autres techniques de codage, les erreurs sont localisées et sont plus visibles à la restitution des images.

Un des principaux défauts du codage par la transformée TCD est L'effet de bloc. Il traduit le codage de chaque bloc indépendamment de ses voisins. Il est perceptible dans les plages de faible variation de luminance.

II - C Codage par quantification vectorielle

L'idée de la quantification vectorielle est de tirer parti de l'interdépendance présente entre les échantillons du signal, lors de la quantification.

D'après la théorie débit-distorsion établie par Shannon la quantification vectorielle offre de bonnes performances. En effet, le taux de compression et la qualité subjective des images obtenus par la quantification vectorielle sont meilleurs que ceux réalisés à l'aide de la quantification scalaire. Cette constatation est vérifiée même pour les sources débitant une séquence d'échantillons aléatoires et indépendants.

II - C - 1 Principe

Le codage utilisant la quantification vectorielle est une technique basée sur un découpage de l'image en médaillons. Les éléments obtenus, blocs ou matrices de pixels, sont étiquetés. L'indexation consiste à rechercher

pour chaque bloc de l'image l'élément le plus proche dans un dictionnaire pré-établi dans le codeur. L'étiquetage fait correspondre au bloc, présenté en entrée du codeur, l'adresse du vecteur voisin dans le dictionnaire.

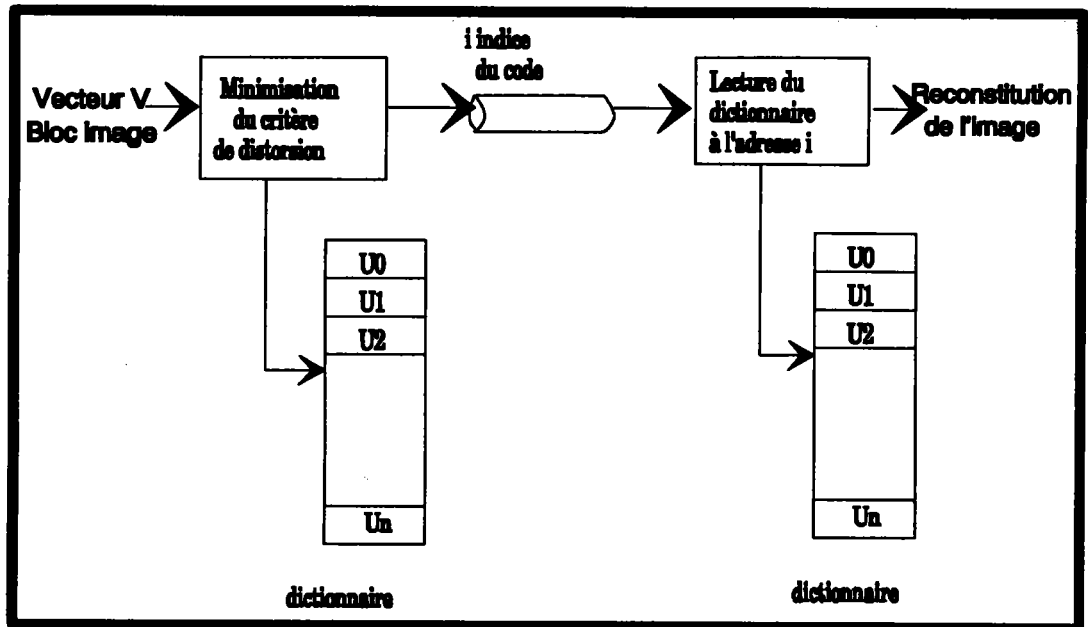


Figure N°5 Codage par quantification vectorielle

X est le bloc de pixel ou vecteur initial à coder, Y_i sont les éléments du dictionnaire; avec $i=1\dots n$, n nombre de vecteurs. Le codage par quantification vectorielle fait correspondre à X l'indice (i) de son plus proche voisin dans le dictionnaire. Cet indice est le mot code qui sera transmis.

$$QV(X) = i \quad \text{si } D(X, Y_i) \leq D(X, Y_j) \\ \text{pour tout } i \text{ différent de } j \text{ avec } j=1 \dots n$$

$D(X, Y)$ étant la distance qui sépare les vecteurs X et Y.

Nous parcourons, donc, tout le dictionnaire à la recherche du bloc de pixels qui pourra le mieux représenter le bloc incident. Une fois ce vecteur trouvé son adresse est transmise au décodeur. Au décodage, l'image restituée sera formée à partir des vecteurs indexés Y_i du dictionnaire (cf. Fig. N°5).

Le dictionnaire est mémorisé dans le codeur et aussi dans décodeur. A la réception du mot code, le décodeur fait une lecture du vecteur correspondant à l'adresse reçue et appartenant au livre de codes ou

dictionnaire. L'image finale est reconstituée par tous les vecteurs dont les composantes vont former les éléments de l'image décodée.

La notion de distance ou distorsion entre deux vecteurs est introduite pour la recherche du plus proche voisin dans le dictionnaire. Plusieurs distances ont été proposées pour optimiser la quantification vectorielle et pour s'approcher au maximum d'un modèle visuel idéal dont la fidélité serait presque sans reproche.

La distance ou distorsion appelée erreur quadratique, est parmi celles qui sont les plus utilisées pour la quantification vectorielle. L'appellation distance ici n'est pas exacte, il s'agit, en fait, du carré d'une distance au sens mathématique du terme.

$$D(U - V) = \sum_{i=0}^k (u_i - v_i)^2$$

Les vecteurs U et V sont de dimension k.

La problématique de la quantification vectorielle a été identifiée par différents auteurs: Gray et al, Goldberg et al, Nasrabadi et al etc. Nous nous proposons ici de la schématiser par la définition de ces trois composantes principales interdépendantes:

- la vectorisation ou la formation de vecteurs à partir des informations à coder,
- la formation du dictionnaire à partir d'un ensemble d'apprentissage,
- la recherche du plus proche voisin à l'aide d'une distance appropriée.

Nous allons détailler ces différentes étapes, en mettant en valeur les principaux problèmes que pose chacune d'elles, ainsi que les solutions proposées par des auteurs, ainsi que les démarches à suivre pour apporter une certaine amélioration ou évolution aux approches classiques.

II - C - 2 Formation des vecteurs

Cette étape rassemble tous les pré-traitements qui, partant de

l'information image à coder, forment des vecteurs. Les composantes de ces derniers doivent être d'une grande valeur informative. La vectorisation idéale est celle qui forme des vecteurs matérialisant exclusivement les données pertinentes du bloc image.

Les pré-traitements incluent le découpage de l'image en médaillons ainsi que l'opération d'extraction et de décorrélation des informations pertinentes à coder. Cette dernière étape est essentielle pour optimiser la quantification car elle agit directement sur la représentativité du dictionnaire ainsi que sur l'éventuelle complexité du codage et par conséquent du décodage.

L'information contenue dans l'image est traitée de façon à faciliter la quantification vectorielle. La décorrélation est basée sur des transformations simples sur les données initiales, pour aboutir à des données plus décorrélées et dont la densité d'informations par élément est plus importante.

Techniques de normalisation pour la quantification spatiale

Baker et al proposent dans [BAKE 82] une méthode qui permet la décorrélation des pixels du bloc image. En premier, la moyenne des luminances des pixels est calculée, ensuite une soustraction est effectuée entre cette moyenne et les luminances des pixels qui composent le bloc image. Le vecteur des écarts obtenus est quantifié. La moyenne, elle, est codée séparément par un quantificateur scalaire.

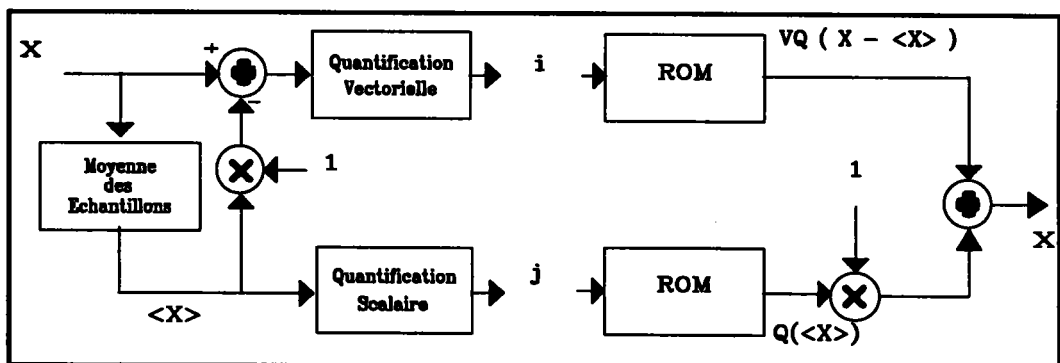


Figure N°6 Quantification vectorielle Moyenne / Gain

Une variante de cette technique consiste à prendre en compte l'erreur résiduelle de la quantification scalaire de la moyenne dans le codage vectoriel. Cela est possible en effectuant la soustraction entre les pixels et la valeur quantifiée de leur moyenne. Ainsi l'erreur de quantification est

introduite lors de la quantification vectorielle.

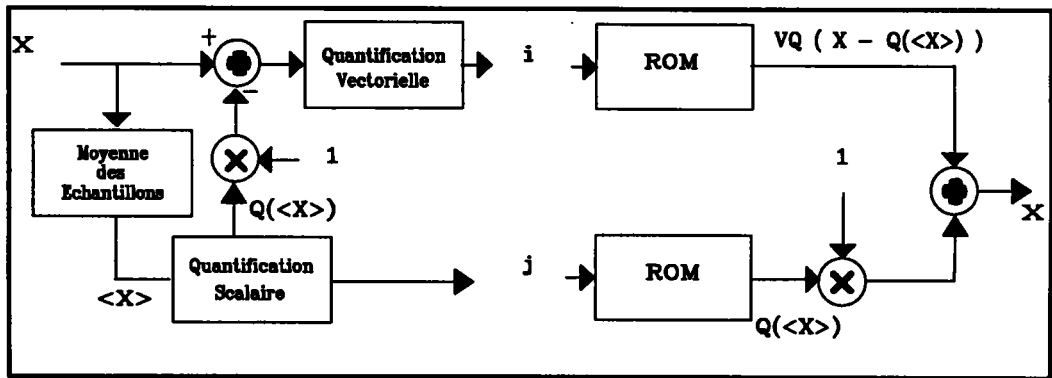


Figure N°7 Quantificateur vectorielle à moyenne résiduelle

Murakami et al proposent dans [MURA 86], une technique de codage utilisant la normalisation avant d'appliquer la quantification vectorielle (cf. Fig. N°6 et N°7). Elle offre ainsi un signal à quantifier peu dépendant de la scène à coder. Parmi les avantages de cette décorrélation l'augmentation de la représentativité du dictionnaire et donc l'optimisation du codage.

Soit un vecteur à coder $V(v_1, v_2, \dots, v_k)$, la moyenne (M) et l'écart-type (e) de chaque composante sont calculées. Le vecteur normalisé $U(u_1, u_2, \dots, u_k)$ est défini par:

$$u_i = \frac{v_i - M}{e}$$

avec

$$M = \sum_{i=0}^k \frac{v_i}{k} \quad \text{et} \quad e = \left[\sum_{i=0}^k \frac{v_i - M}{k} \right]^{1/2}$$

Les vecteurs normalisés sont quantifiés, leurs moyennes et les écarts types sont codés séparément par des quantificateurs scalaires.

Pour chaque vecteur quantifié transmis, la moyenne et l'écart-type le sont aussi, ainsi que la position du vecteur dans l'image. Le dictionnaire est construit de telle façon que la probabilité d'apparition des vecteurs qui le forment soit constante pour toute image en entrée du codeur.

Etant donné la rigidité du quantificateur vectoriel utilisé par Murakami et al. Les auteurs ont amélioré la technique de manière à la rendre adaptative.

Les valeurs des composantes vectorielles et leurs fonctions

d'interdépendances, ont une action déterminante sur l'optimisation de la quantification vectorielle. La taille des vecteurs à coder est aussi très influente sur les performances du codage. En effet, elle intervient directement sur la représentativité du livre de code, elle agit aussi sur le taux de compression et sur la qualité visuelle subjective des images restituées au décodage.

L'utilisation de vecteurs de petite dimension permet de minimiser l'erreur de quantification vectorielle et donc d'améliorer les images décodées, mais son effet se ressent négativement sur le taux de compression.

II - C - 3 Construction du dictionnaire

Partant d'un ensemble de vecteurs très large, le dictionnaire vectoriel optimal est celui qui emploie un groupe de vecteurs présentant la plus petite distorsion moyenne, parmi tous les autres groupes de vecteurs possibles. La construction du dictionnaire est basée sur la recherche des meilleurs représentants des blocs image. A l'aide de ce dictionnaire, un quantificateur vectoriel optimal pourra être élaboré.

Des algorithmes de classification sont utilisés, de façon à élaborer un dictionnaire de vecteurs représentatifs à partir d'un ensemble initial; ce dernier est appelé "training set" ou ensemble d'apprentissage.

L'ultime objectif est de créer un "codebook" ou dictionnaire universel. La principale particularité du livre de codes, est d'être parfaitement adapté à toutes les images incidentes. Ce dictionnaire permettrait de coder tous les types d'images avec un taux de compression acceptable et une bonne qualité subjective. En attendant la découverte de ce dictionnaire idéal, plusieurs auteurs, Gray et al dans [GRAY 84], Linde et al dans [LIND 80], Gersho et al dans [GERS 82] et Kohonen et al dans [KOHO 84], se sont penchés sur le problème. Plusieurs solutions ont été proposées :

Nuées dynamiques

Cette méthode exige la définition d'un espace des représentation E et un paramètre de distorsion $f(x,w)$ qui associe à tout point x de E et à tout sous ensemble w de E une valeur de R^+ :

$$\forall x \text{ de } E, \forall w \text{ de } P(E) \rightarrow f(x,w) \text{ de } R^+$$

f exprime le degré de dissemblance entre la classe w et le point x . Plus cette valeur est faible, plus x est considérée proche de w .

L'algorithme des nuées dynamique suppose connu J , le nombre des classes.

idée de base de l'algorithme

Il s'agit de dire que la meilleure partition $\Pi = \{w_1, w_2, \dots, w_j\}$ de E est celle pour laquelle la somme des distorsion entre les points et les classes auxquelles ils appartiennent est la plus faible. Cela s'écrit ainsi :

A chaque partition, nous associons la quantité :

$$g(\Pi) = \sum_{w \text{ de } \Pi} \sum_{x \text{ de } \Pi} f(x, w)$$

La meilleure partition de l'espace E en J classes est celle pour laquelle $g(\Pi)$ est minimum.

La majorité des algorithmes recherchent des minimum locaux.

L'algorithme suivant est fréquemment utilisé :

- 1) Choisir au hasard une partition Π^0 en classes $\{w_1^0, x_2^0, \dots, x_j^0\}$
- 2) $n = 0$
- 3) $n = n+1$
- 4) Pour chaque valeur de j , calculer

$$w_j^n = \{x \text{ de } E \mid \forall k \text{ différent de } j, f(x, w_j^{n-1}) \leq f(x, w_k^{n-1})\}$$

- 5) retourner en 3) tantque J_j, w_j^n différent de w_j^{n-1} et que $n \leq n^0$

la condition $n \leq n^0$ fixe un nombre maximum d'itérations a priori et évite que l'algorithme ne boucle indéfiniment. L'idée exprimée en 4) consiste à affecter successivement chaque point x de E à la classe dont il est le plus proche et à définir ainsi peu à peu les classes.

La convergence de l'algorithme n'est démontrable que pour certaines fonctions f .

La définition d'une fonction de distorsion se fait en deux étapes :

La première consiste à choisir une représentation de la classe appelée noyau. Le noyau peut être le centre de gravité de la classe, un ensemble de p points minimisant une fonction de ressemblance [BELA 88] .

Exemple de $f(x, w)$:

Le noyau de chaque classe est son axe principal d'inertie Δw et en définissant la distorsion f par :

$$f(x,w) = d(x, \Delta w) = \inf_y \text{ de } \Delta w \ d(x,y)$$

Algorithme de classification du dictionnaire LBG

Linde et al proposent une technique couramment appelée LBG. Elle nécessite une pré-sélection des vecteurs constituant le "training set", suivant leur nombre d'apparitions dans le codage. Rappelons que la méthode des centroïdes exploite quant à elle le maximum de vecteurs d'apprentissage [LIND 80].

Le LBG effectue itérativement des tests, de façon à former le dictionnaire le plus représentatif à partir de la série d'apprentissage. Pour un critère de distorsion donné, l'algorithme vise itérativement à optimiser la partition de l'espace des vecteurs en fonction du dictionnaire courant et vice-versa. Successivement, les éléments du dictionnaire sont réajustés, l'amélioration de représentativité des partitions de l'espace est vérifiée à chaque itération. le dictionnaire est considéré comme optimal lorsqu'une itération supplémentaire du processus n'amène plus de diminution significative de la distorsion moyenne totale.

La synthèse du dictionnaire initial nécessite une pré-sélection réalisée à la suite d'une étude statistique sur les vecteurs de l'ensemble d'apprentissage ou par la méthode des centroïdes. La création du dictionnaire initial est donc le fruit de plusieurs investigations effectuées sur les vecteurs dont le nombre d'apparitions est relativement plus important que celui des autres. Rappelons que la détermination de ce livre de codes initial est primordiale pour la suite des opérations de construction du dictionnaire.

La méthode des centroïdes, consiste à rassembler le maximum de vecteurs pour former un ensemble d'apprentissage. Les images de références sont souvent les sources qui fournissent les vecteurs. L'ensemble ainsi formé appartient à l'espace de dimension N; N étant la taille des vecteurs.

La détermination du centre de cet espace est entreprise; le résultat est un vecteur qui représente la valeur moyenne des vecteurs du "training set". Un hyperplan, passant par le centre, est ensuite formé de façon à fractionner l'espace en deux classes. La même opération est répétée plusieurs fois, décomposant l'espace en une multitude de nuages de vecteurs; chacun de ces groupes est représenté par son centre (cf. Fig. N°8 et 9).

Cette méthode, permet de répertorier toutes les sous-classes par leurs

centres respectifs. Ces centres seront une bonne représentation du "training set" et formeront ainsi un dictionnaire optimisé [GRAY 84].

Une autre particularité avantageuse de cette technique est qu'elle permet de minimiser le parcours du dictionnaire lors de la quantification vectorielle. L'arbre, formé à partir des barycentres des cellules, offre la possibilité d'être parcouru sur les seules branches voisines aboutissant au vecteur du dictionnaire. Ainsi, de proche en proche, la quantification vectorielle est réussie quand un des centres est suffisamment proche du bloc à coder, c'est à dire quand la distorsion entre les deux vecteurs est inférieure au seuil de tolérance fixé préalablement.

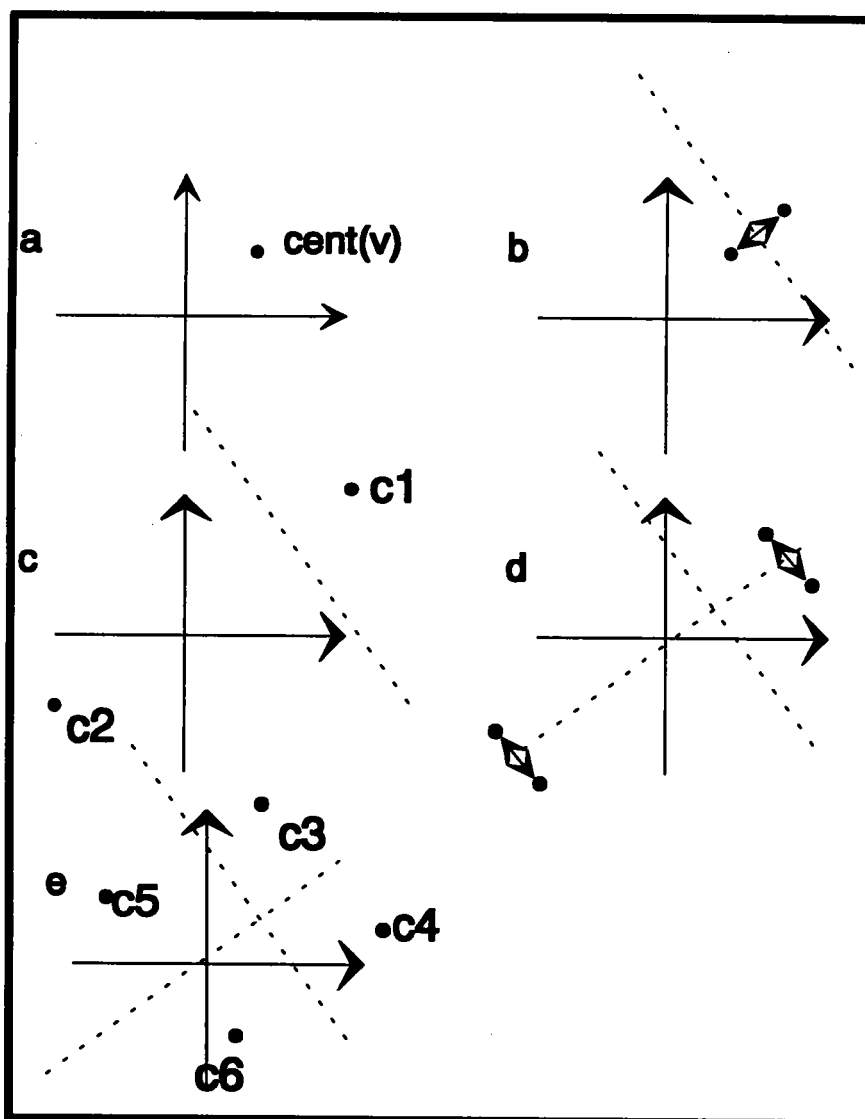


Figure N°8 Classification par la méthode des centroïdes

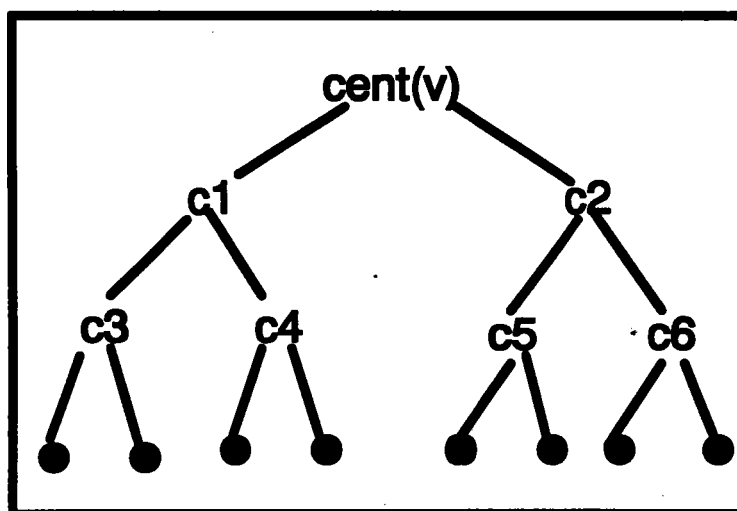


Figure N°9 Disposition des vecteurs dans le dictionnaire

Algorithme de Linde Buzo Gray

Il s'agit donc de trouver un vecteur de paramètres $y_1...y_J$ à l'intérieur d'un ensemble qui minimise une mesure de distorsion.

Le problème posé est le suivant : Comment déterminer un ensemble de références ou codebook. Un dictionnaire est un ensemble fini de vecteurs de référence.

Une méthode itérative est proposée pour créer un livre de codes qui minimise une distorsion sur un grand nombre de données. Elle converge vers un minimum local et non obligatoirement vers un minimum global.

L'objectif est de constituer un ensemble de $2^b = B$ vecteurs représentant un ensemble large de vecteurs rassemblés après analyse des données.

A l'initialisation, B vecteurs sont choisis arbitrairement ou à la suite d'une étude de représentativité. La première étape détermine par la méthode du plus proche voisin les groupes de vecteurs représentés par les B vecteurs initiaux.

La distorsion totale est déterminée en sommant la distorsion de chaque vecteur de la base de données lors de la répartition des classes.

La deuxième étape se propose d'améliorer le livre de codes en choisissant un meilleur représentant pour chaque classe ; exemple le centroïde de la classe évaluée à l'étape 2.

Cet algorithme conduit à une distorsion globale monotone décroissante qui converge donc vers un optimum local.

La vitesse de convergence dépend du choix des B vecteurs initiaux.

Algorithme

1) Initialisation

Fixer un nombre N de représentants du dictionnaire

un seuil $\varepsilon > 0$

Un dictionnaire initial R_0

Une séquence d'entraînement $(X(j); j=1, \dots, J)$

$D_{-1} = \text{infini}$

$m = 0$

Phase N°1

Pour le dictionnaire courant $R_m (y_i; i=1, \dots, N)$

Trouver la partition optimale $\text{Part}(R_m = (C_i; i=1, \dots, N))$

qui minimise la distorsion moyenne sur une séquence d'entraînement

soit

$X(j)$ de C_i

si $d(X(j), Y_i) \leq d(X(j), Y_j)$ pour i différent de j

Calculer la distorsion moyenne pour la répartition obtenue et le dictionnaire R_m

$$D_m = \sum_{j=1}^J \min_i d(X(j), y_i)$$

Phase N°2

Si $(D_{m-1} - D_m)/D_m \leq \varepsilon$, le dictionnaire optimal est R_m

sinon continuer

Phase N°3

Trouver le dictionnaire optimal R_{m+1} pour une répartition $\text{Part}(R_m)$ pour tout $i, (i=1, \dots, N)$, $y = \text{barycentre de } C_i$

$m=m+1$

exécuter la phase 1

Méthode de Kohonen

C'est une approche algorithmique connexioniste, basée sur les réseaux

de neurones. Cette technique consiste en un apprentissage auto-adaptatif. Le classement des vecteurs de l'ensemble d'apprentissage se fait au fur et à mesure. Les vecteurs sont présentés individuellement en entrée d'un réseau de neurones affectés respectivement par des poids de pondération. Les vecteurs appartenant à une même classe sont de plus en plus regroupés, les autres sont dispersés. La répartition des vecteurs en classes est ainsi effectuée, l'ensemble des vecteurs qui représenteront le "training set" est ainsi constitué.

L'optimisation du dictionnaire dépend principalement du nombre de vecteurs contenus dans l'ensemble d'apprentissage et de la durée nécessaire pour que le réseau de neurones effectue son apprentissage.

Cette technique est celle recueillie des avis positifs, justifiés par les possibilités qu'elle peut offrir pour le rafraîchissement dynamique des dictionnaires. Son efficacité n'égale pas encore celle de ses concurrentes; cependant, elle a beaucoup d'avenir, car elle présente une implémentation relativement simple et elle offre la possibilité d'une extension simultanée du dictionnaire. Cette dernière éventualité sera détaillée dans le chapitre suivant.

II - C - 4 Distance déterminant la distorsion entre les vecteurs

La mesure de distorsion la plus utilisée en codage d'images est l'erreur quadratique, déterminée par l'expression suivante :

$$\text{Deq}(X,Y) = \sum_{i=0}^k (x_i - y_i)^2$$

Deq (X,Y) détermine la distorsion qui sépare les vecteurs X et Y. Ce paramètre ne tient pas compte du système psycho-visuel humain; elle ne peut distinguer entre une erreur de codage visible et une autre imperceptible par l'observateur. Cependant, elle constitue un facteur qui permet de comparer l'éloignement entre les vecteurs, même si les résultats qu'elle donne sont plus ou moins grossiers.

D'autres paramètres de distorsion, qui sont des variantes de celle proposée dans le paragraphe ci-dessus, sont citées dans la littérature spécialisée; erreur quadratique moyenne normalisée, moyenne temporelle de l'erreur quadratique ou encore le critère de distorsion minimale - maximale.

Chapitre III

METHODES ADAPTATIVES DE CODAGE D'IMAGES

Dans le chapitre précédent nous avons décrit les méthodes de codage classiques MICD, TCD et la Quantification Vectorielle. Afin d'étudier les particularités respectives de chaque technique et de mieux cerner la problématique du codage, nous nous proposons de décrire dans ce chapitre quelques exemples de procédés adaptatifs de compression d'images. Le même schéma suivit au chapitre précédent sera respecté ici : méthodes prédictives, codage par transformation et enfin quantification vectorielle.

III - A Méthode prédictives

Les techniques de codage par modulation des impulsions codées différentielles exploitent les évolutions du signal image. Ces variations sont temporelles et/ou spatiales. Les techniques classiques utilisant un estimateur et un quantificateur d'erreur de prédiction fixes sont inefficaces. Elles ont montré leur incapacité à obtenir de bonnes performances de codage et à les maintenir pour différents types d'images.

Pour améliorer les performances en qualité subjective de l'image décodée et en taux de compression, plusieurs techniques font appel à l'adaptabilité. Cette dernière peut s'appliquer aux différentes étapes du codage différentiel : au niveau des prédicteurs, ou au niveau des quantificateurs de l'erreur d'estimation. Le but recherché est d'optimiser la

technique de codage en améliorant ses performances.

L'adaptabilité réside dans le fait que la technique de codage adaptative exploite un prédicteur ou un quantificateur ou les deux à la fois, s'accommodant avec les échantillons du signal à coder. Les caractéristiques des composants du codeur sont variables en fonction des évolutions et des particularités du signal image.

III - A - 1 MICDA-ER

Nous proposons ici de décrire en exemple une méthode originale de compression d'images utilisant plusieurs estimateurs et plusieurs quantificateurs scalaires.

Dans [ZHAO 89], Y. ZHAO a présenté une technique adaptative de codage d'images. Cette méthode évalue la corrélation liant des pixels d'un même bloc et celle les associant avec les pixels du bloc identique dans l'image antécédente, dans la même séquence.

En tenant compte de la fonction d'interdépendance calculée, un estimateur de la luminance est choisi. La quantification scalaire de l'écart entre la valeur réelle de luminance et celle prédite est effectuée. Le mot code transmis comporte une indication pour identifier le prédicteur ainsi que le quantificateur utilisés, le récepteur déchiffre les indications pour décoder la valeur de l'erreur d'estimation quantifiée.

L'adaptation réside dans la possibilité d'avoir le choix entre deux prédicteurs et entre deux quantificateurs. Les meilleurs sont sélectionnés en fonction des paramètres de corrélation associant les pixels se succédant sur une même image ou sur une séquence d'images.

L'estimation de la redondance est effectuée après le calcul de la fonction corrélation C_{ni} , cette dernière est liée aux trois composantes d'un vecteur test :

$$C_{ni} = F(C_{ni1} ; C_{ni2} ; C(n,n-1)_i).$$

Le bloc image (8x8) est divisé en deux sous blocs (4 x 8); les coefficients C_{ni1} et C_{ni2} sont leurs facteurs associés déterminant la corrélation intra-image. Les indices correspondent à l'instant n , à l'image i et à l'indexe du sous bloc. Les coefficients permettent d'évaluer la redondance et

l'interdépendance existant entre les pixels du même sous-bloc.

$$\Delta A(j) = A_n(i,j) - A_{n-1}(i,j)$$

$A_n(i,j)$ désigne la luminance du pixel d'ordre j appartenant au bloc i dans l'image. Δ représente l'écart entre deux pixels successifs. L'indice j varie de 0 à 63.

$C(n,n-1)_i$ coefficient de corrélation inter-image; il est estimé par comparaison de tous les pixels d'ordre j du bloc i dans l'image à l'instant n et ceux du bloc correspondant dans l'image à l'instant $n-1$. $C(n,n-1)_i$ prend l'une des valeurs suivantes (0 ; 0,5 ; 1) en fonction des dépassements de seuils comptabilisés effectués par le paramètre $\Delta A(j)$.

Un calculateur commande le choix des prédicteurs et des quantificateurs. Pour l'évaluation de la corrélation temporelle, l'organe de décision utilise un compteur qui comptabilise le nombre de dépassements de seuils prédéterminés, effectués lors du calcul des $\Delta A(j)$.

Des tests comparatifs déterminent la classe d'appartenance du bloc image en fonction du taux de dépassements calculé. La valeur du taux permet de prendre une décision en fonction de l'importance de la corrélation du bloc avec son prédécesseur. Si l'écart constaté entre les deux blocs successifs est faible, alors au décodage, la visualisation du bloc précédent est maintenue. Dans le cas contraire, l'estimation de la corrélation spatiale ou intra-image est effectuée sur chaque demi-bloc :

les facteurs d'interdépendance; C_{n1i} et C_{n2i} sont déterminés à partir du taux de dépassement constaté entre les luminances des pixels de chaque demi-bloc et sa moyenne. Si ce taux est faible le demi-bloc est considéré uniforme, seule sa moyenne est donc transmise au décodeur. Sinon, la comparaison entre les facteurs de dépendance temporelle et spatiale permet le choix entre deux prédicteurs de luminance. Le premier donne pour valeur de prédiction la luminance du pixel à la même position dans l'image précédente, le deuxième propose, lui, la moyenne des luminances de deux pixels voisins.

Le prédicteur choisi est celui qui fait l'estimation la plus proche de la valeur réelle des luminance de chaque sous-bloc d'image.

En ce qui concerne la quantification de l'erreur de prédiction, deux éventualités sont possibles :

- 1 : L'écart entre la valeur réelle de la luminance et la valeur prédite, la plus fidèle des deux possibles, est limité. Dans ce cas, le quantificateur choisi délivre des erreurs de prédiction quantifiées et codées sur deux bits.
- 2 : Une variation importante s'est produite entre les deux images successives, doublée d'une évolution spatiale conséquente des luminances. Dans cette éventualité, l'écart entre les valeurs réelle et les estimation est important, le quantificateur sélectionné alloue pour les erreurs de prédiction un mot code dont la longueur est quatre bits.

La Modulation par Impulsions Codées Différentielle Adaptative avec Estimation de la Redondance (MICDA-ER) exploite plusieurs prédicteurs qui opèrent simultanément. Elle sélectionne l'estimateur optimal, celui qui va présenter l'erreur de prédiction minimale. A la suite de ce choix, un quantificateur scalaire est déterminé pour évaluer et coder l'erreur de prédiction.

La MICDA-ER améliore les performances de la technique prédictive classique, qualité subjective des images décodées et taux de compression de l'image, en exploitant des prédicteurs et des quantificateurs dont les caractéristiques sont variables en fonction de l'évolution des échantillons du signal image à coder [ZHAO 89] (cf. Fig. N°10).

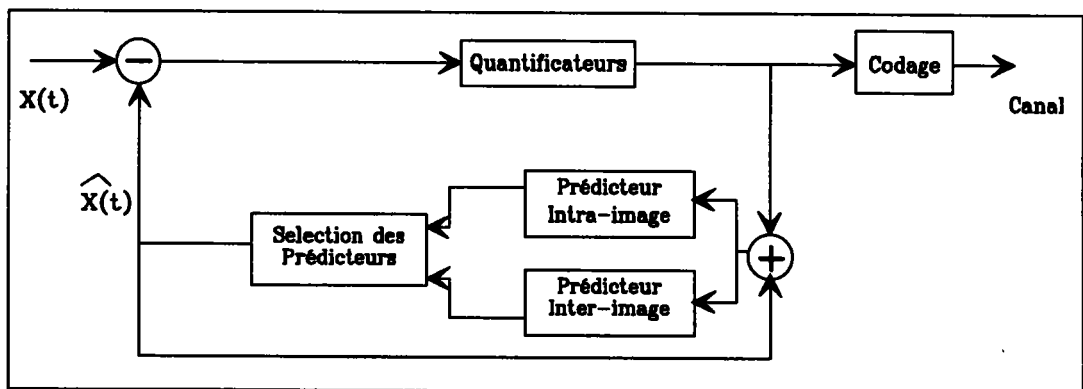


Figure N°10 Codage prédictif par estimation de la redondance

III - B Techniques adaptatives de codage par transformation TCD

L'adaptation pour les techniques de codage basées sur des transformations mathématiques peut avoir deux formes différentes. Elle peut concerner les caractéristiques de la transformation, ou la série de procédures appliquées aux coefficients transformés pour les coder.

III - B - 1 Adaptation par variation de la taille des blocs transformés

Elle est applicable à chaque fois que les images à coder se présentent sous forme de plages ayant les mêmes structures picturales. L'image est fractionnée de façon à former les plus grandes zones possibles de même structure. Ainsi, après la segmentation, des blocs de tailles variables seront formés dans les plages uniformes, ou encore dans des zones contenant des contours ou des textures particulières.

Vaissey et al proposent dans [VAIS 87] une technique de codage par transformation, avec des tailles de blocs variant de 4×4 pixels à 32×32 pixels. Pour les images dont l'évolution des luminances des pixels sont faibles sur de grandes plages, la taille 32×32 pixels leur est réservée, un taux maximum de compression est obtenu. Par contre, pour les zones très contrastées, contenant des détails indispensables à la reconstitution des images il est nécessaire d'avoir un codage adéquat. Des blocs de petite taille sont consacrés aux particularités de l'image pour respecter la fidélité et obtenir la meilleure qualité subjective de l'image décodée.

Nous constatons que la taille des médaillons à transformer est associée au contenu structurel de l'image. Cependant, cette catégorie de techniques adaptatives met en oeuvre plusieurs calculateurs exécutant la transformée ainsi qu'une puissance de traitement importante pour identifier les zones englobant des structures semblables. Sans oublier que pour identifier le type de transformation effectuée au codage, il est nécessaire de transmettre des codes supplémentaires au décodeur. Toutes ces contraintes handicapent énormément ce type de techniques adaptatives de codage. Elles posent un réel problème pour leur implémentation, ce qui compromet fatalement l'éventuelle industrialisation de codeurs - décodeurs d'images conçus à partir de procédés similaires.

III - B - 2 Stratégies adaptatives de codage de coefficients transformés

Nous avons constaté lors de l'étude préliminaire de la transformée en cosinus discrète que, compte-tenu de l'importance de la dynamique des coefficients obtenus, le passage dans le domaine transformé ne constitue pas en lui-même une compression de données. Nous avons aussi remarqué que la disposition des coefficients transformés dans la matrice finale est étroitement liée au contenu de l'image.

Toutes ces caractéristiques ont fait la faillite des méthodes classiques de codage par transformation, en contribuant à leurs mauvaises performances. De nouvelles approches algorithmiques apportent des méthodes d'accommodation aux particularités des contenus informatifs de l'image. D'autres techniques repèrent les zones de la matrice transformée contenant les coefficients pertinents.

Pour mieux cerner les opportunités offertes par les techniques adaptatives, nous présenterons celles qui nous ont semblé les plus originales, les plus robustes et les plus compétitives.

Classification basée sur des paramètres énergétiques

La répartition statistique des amplitudes des coefficients $F(u,v)$ transformés s'approche d'une distribution Gaussienne; l'énergie moyenne des éléments de la matrice décroît rapidement avec le rang de leur position. De ces constatations, découle la nécessité d'octroyer un nombre variable de bits pour le codage de chaque coefficient $F(u,v)$ en fonction de sa position dans le bloc transformé. Le nombre total d'octets alloués pour chaque bloc transformé est associé à l'activité spatiale des luminances de l'image.

Une méthode de codage TCD adaptative, présentée dans [CHEN 77], exploite des paramètres évaluant la distribution d'énergie dans la matrice transformée. Un vecteur caractéristique est calculé à partir des valeurs des coefficients transformés $F(u,v)$. Une procédure de classification est accomplie suivant une règle de décision exploitant la caractéristique du bloc transformé.

La classification d'après [CHEN 77], s'effectue en calculant le paramètre matérialisant l'énergie totale:

$$\sigma_s = \left(\sum_{u=0}^{u=k} \sum_{v=0}^{v=k} F(u,v)^2 \right) - F(0,0)^2$$

La comparaison de la valeur obtenue avec trois seuils pré-établis permet d'identifier la classe d'appartenance du bloc transformé. Les limites entre les quatre familles sont fixées par apprentissage de façon à obtenir pour chaque bloc des possibilités d'appartenance équiprobables.

Quatre matrices B_k d'allocation de bits sont calculées après plusieurs expérimentations sur un ensemble d'apprentissage. La détermination des éléments des matrices B_k se fait à partir des variances des éléments transformés $F(u,v)$:

Pour chaque classe C_k , on détermine sur l'ensemble d'apprentissage les variances des éléments transformés :

$$a_{ij} : \sigma^2_{a_{ij}(k)} \text{ pour } (i,j) = 0, \dots, N-1 \text{ sauf } (0,0) \text{ et } k = 1, \dots, 4$$

on en déduit alors les matrices B_k d'allocation des bits :

$$[B_k] = [n^*_k(i,j)]$$

avec

$$n_k(i,j) = 1/2 \log_2[\sigma^2_{a_{ij}(k)}] - \log_2(D_k)$$

pour (i,j) différent de $(0,0)$

où n^* est la troncature de n à l'entier le plus proche.

D_k est le nombre total de bits alloué au codage de la classe C_k

$$n_k(0,0) = 8.$$

Le codage adaptatif et le décodage se fait d'une façon générale selon les blocs-diagrammes (cf. Fig. N°12 et 13) .

La quantification des éléments a_{ij} porte en fait sur les éléments normalisés a_{ij}^N . Les facteurs de normalisation sont de la forme de :

$$\lambda_k(i,j) = C \cdot 2^{[n^*_k(i,j)-1]} \quad \text{avec } (i,j) \text{ différent de } (0,0)$$

Chaque élément normalisé est donc quantifié en utilisant la loi de

quantification normalisée à $2^{[n \cdot k(i,j)]}$ niveaux si le bloc est affecté à la classe k . Le codage d'un bloc nécessite d'indiquer au décodeur la classe identifiée [BARB 85]. Les quatre matrices suivantes définies par Chen et al sont dans le cas de blocs (16 x 16). Le débit moyen obtenu est 1 bit/point.

Les coefficients transformés $F(u,v)$ sont normalisés pour donner $NF(u,v)$, par des facteurs calculés à partir du nombre de bits associé au même rang dans le bloc.

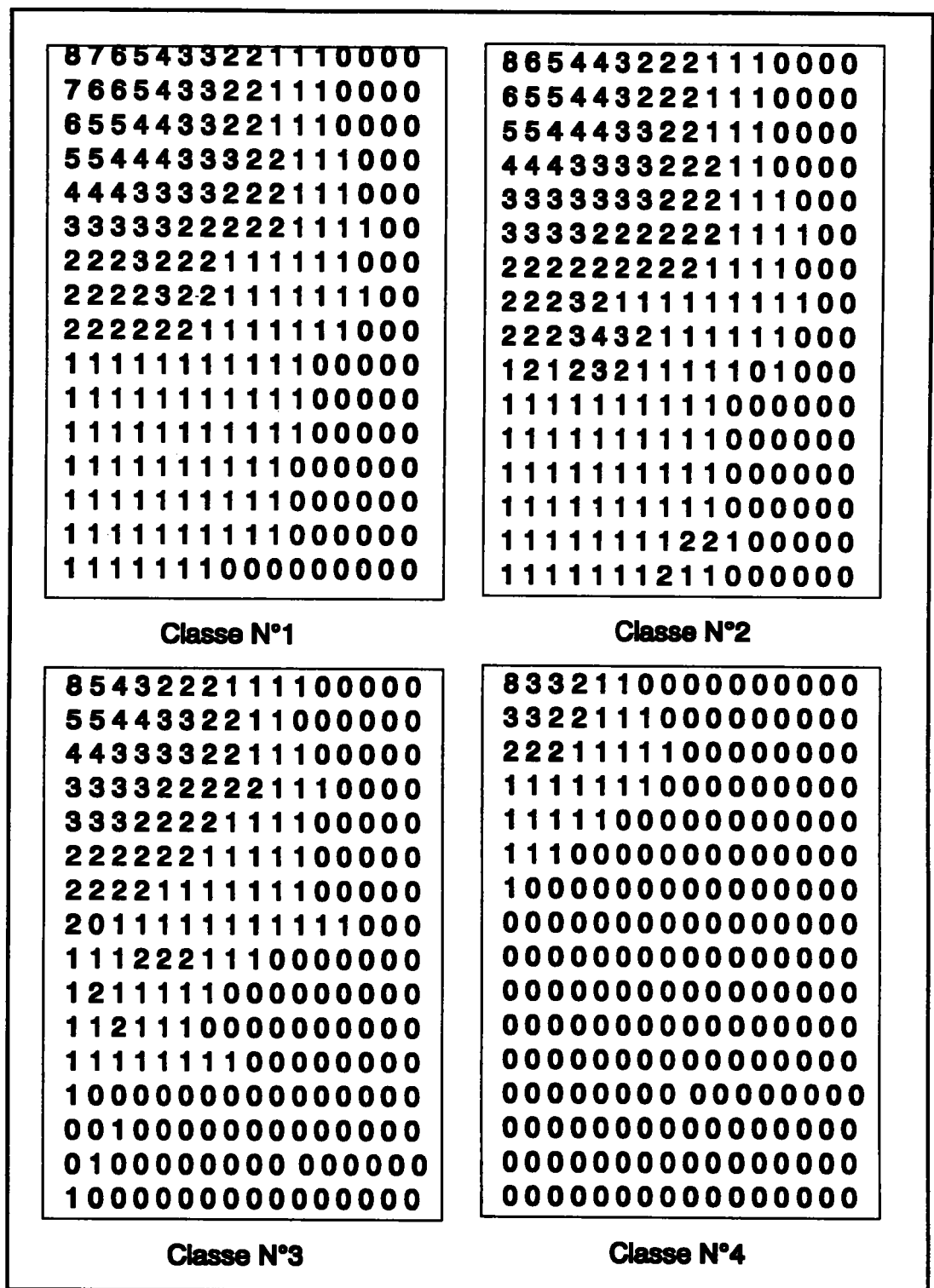


Figure N°11 Classification par matrices d'allocation de bits

A chaque classe d'énergie C_k est associé un quantificateur spécifique des valeurs normalisées $NF(u,v)$, déterminé par la loi de quantification à 2

$nk(u,v)$ niveaux. L'en-tête affectée à chaque bloc identifie sa classe énergétique C_k et par conséquent la matrice d'allocation de bits, les valeurs de normalisation ainsi que le quantificateur approprié (cf. Fig. N°11).

Les figures suivantes décrivent les différentes étapes nécessaires pour effectuer le codage adaptatif par allocation de bits. Certaines procédures sont exécutées d'une façon parallèle, cela réduit considérablement le temps total des traitements et augmente les performances du codeur-décodeur d'images.

La technique proposée par Chen et al, définit la disposition globale des coefficients TCD dans la matrice transformée. Elle ne permet pas de situer la position précise de la zone des coefficients pertinents. Définies grossièrement, les matrices d'allocation de bit ne répondent pas à tous les cas de figures éventuels dans la distribution de l'énergie dans le bloc transformé. Cette technique permet néanmoins d'obtenir un taux de compression de 8 pour une qualité de l'image subjective acceptable.

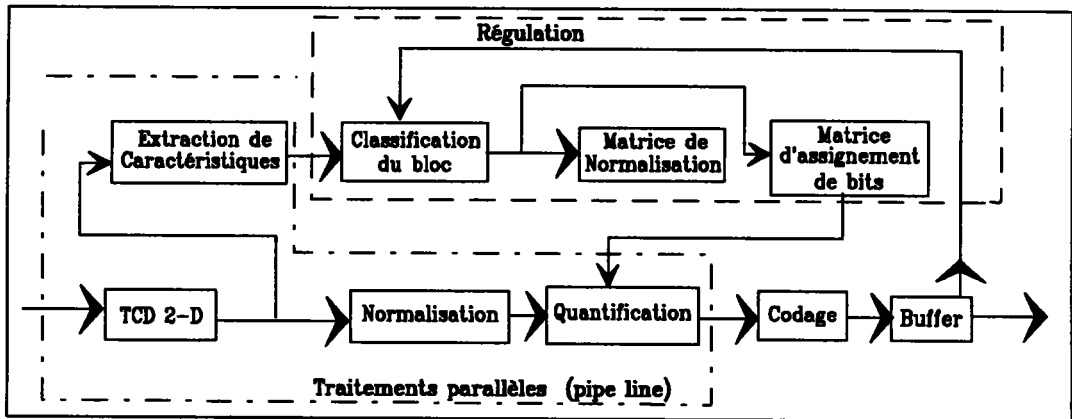


Figure N°12 Codage adaptatif intra-image

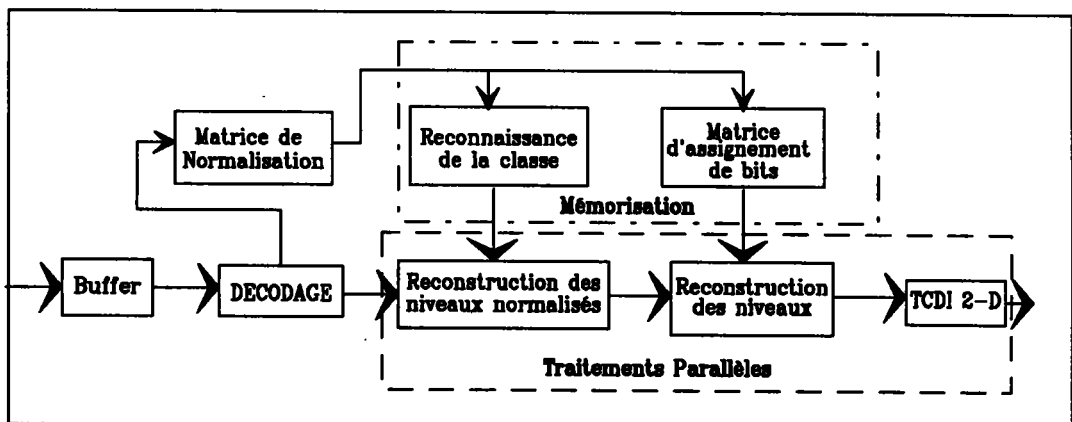


Figure N°13 Décodage par classes

Wu et al proposent dans [WU 82] une méthode basée sur une évaluation de paramètres énergétiques qui tiennent compte de la disposition directionnelle ou/et radiale des coefficients dans la matrice transformée. La classification, très sélective, est établie à la suite d'une décision qui tient compte d'un vecteur caractéristique. Cette opération provoque une répartition fine; les classes obtenues sont disjointes pour permettre d'une part d'éviter les distorsions dues au codage, d'autre part d'accommoder le nombre de classes aux situations réelles.

Un avantage de cette méthode de codage est que, non seulement elle respecte avec fidélité les particularités des blocs transformés, mais en plus le traitement mis en oeuvre pour établir le vecteur caractéristique, ainsi que les procédures de classement, restent relativement simples.

Les composantes du vecteur caractéristique de Wu et al sont ainsi définies : (*MACE*; *DIR*; *FIN*).

Mace représente l'énergie des éléments transformés dans une bande de "fréquences moyennes". L'appellation "fréquence" est souvent utilisée, bien qu'elle ne soit qu'un héritage de la transformée de Fourier.

Pour des blocs transformés de taille 16 x 16.

$$MACE = \sum_{u=2}^9 \sum_{v=2}^9 (F(u,v))^2$$

le paramètre *DIR* détermine la direction dominante de la disposition des coefficients TCD. Il est en effet sensible aux structures dans l'espace transformé dont les orientations sont horizontales ou verticales en basses et moyennes fréquences.

$$DIR = \frac{\sum_{u=0}^{12} \sum_{v=0}^{12} \text{Arctg} \left(\frac{u}{v} \right) \cdot |F(u,v)|}{\sum_{u=0}^{12} \sum_{v=0}^{12} F(u,v)}$$

avec (u,v) différent du couple (0,0)

La dernière composante du vecteur caractéristique est *FIN*. Elle évalue la diminution radiale de l'énergie sur le bloc transformé en comparant l'énergie située en basse fréquence et celle située en moyenne fréquence.

$$FIN = \frac{\frac{1}{N_1} \sum_{(u,v) \text{ de } R_1} F(u,v)^2}{\frac{1}{N_2} \sum_{(u,v) \text{ de } R_2} F(u,v)^2}$$

R_1 et R_2 sont les zones définies par les limites déterminées dans la figure suivante N°14.

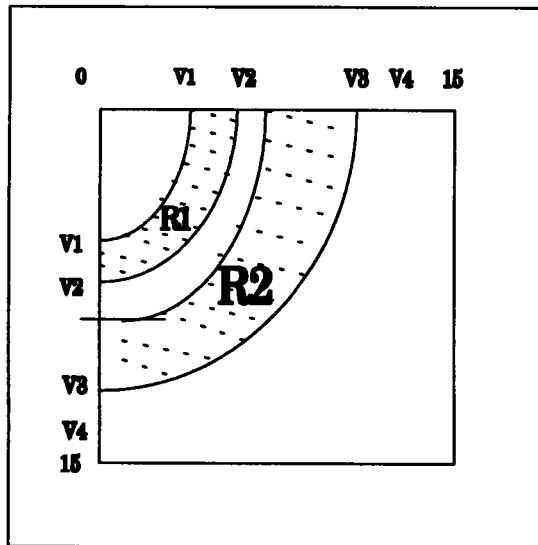


Figure N°14 Structures annulaires du spectre énergétique

Un arbre de décision est établi de façon à élaborer les différentes classes d'appartenance : Un jeu de sept seuils, quatre pour établir la sélection suivant la composante *MACE*, deux selon *DIR*, enfin le dernier seuil pour la discrimination selon la caractéristique *FIN*. Les seuils sont spécifiés à la suite de plusieurs expérimentations effectuées sur un ensemble d'apprentissage, assez représentatif pour que la classification soit fidèle à la réalité de l'espace transformé.

Dix classes sont formées. Par l'intermédiaire de *DIR* sont identifiés les groupes à structures verticales ou horizontales. Pour les classes non

directionnelles, *FIN* localise celles qui représentent des blocs images dont les textures sont fines mais contrastées. La caractéristique *MACE* et les seuils qui lui sont indexés, donnent une indication sur l'étendue des coefficients dans le domaine transformé TCD.

L'amélioration apportée par la classification décrite par Wu et al par rapport à une technique similaire à celle de Chen et al, est significative. Elle est perceptible au niveau de la qualité subjective des images restituées, puisque le codage s'adapte au contenu informatif des blocs transformés. Un autre avantage de cette méthode de classement est qu'elle améliore le taux compression de 20% à 30%.

III - C Approches adaptatives pour la quantification vectorielle

Rejoignant pour cela le codage par transformation, le codage par quantification vectorielle nécessite avant tout un découpage de l'image. Les blocs obtenus rassemblent un nombre de pixels susceptibles d'être interdépendants. La corrélation inter-blocs est ignorée en faveur de la restitution des particularités contenues dans un médaillon.

Le taux de compression ainsi que la qualité subjective des images reconstituées sont étroitement liés à la taille des blocs exploités. Pour des petites tailles de blocs, les détails, même locaux, sont pris en compte lors du codage. Cependant nous obtenons une compression importante pour des images non contrastées découpées en blocs de grande taille. Une recherche de compromis, bonne qualité visuelle et bon taux de compression, est envisageable sous forme d'adaptation de la taille des blocs pour la quantification vectorielle.

III - C - 1 Accommodation des tailles de bloc aux images codées

La diversité des motifs composant l'image est une des particularités fondamentales de cette dernière. Cette affirmation est justifiée par la richesse des caractéristiques des symboles visibles; leurs formes, leurs détails, les variations de leurs niveaux de gris ainsi que leur tolérance visuelle aux erreurs de codage. Il est donc indispensable d'accommoder la technique de compression d'images à leurs contenus.

Dans la quantification hiérarchique [NASR 85], des blocs de tailles: 2×2 , 4×4 , 8×8 et 16×16 sont utilisés, à la suite d'une segmentation spatiale de l'image en zones de structures similaires. Cette répartition différencie les plages uniformes des zones contrastées contenant des détails de taille variable primordiaux pour la compréhension de l'image décodée.

Le codeur élaboré par Nasrabadi et al, forme un arbre binaire dont les feuilles sont les zones de l'image répondant à un critère particulier. Un découpage des blocs 32×32 en quatre est réalisé. Par la suite, une identification, en fonction des structures contenues dans chaque sous-bloc 16×16 , est tentée. Les sous-blocs pour lesquels elle est réussie ont leurs adresses classées dans l'arbre, ils sont codés en fonction du résultat de l'identification. Pour les sous-blocs dont le contenu est plus complexe, un second découpage en quatre est effectué. Ainsi successivement, la reconnaissance des particularités se poursuit jusqu'à l'obtention de blocs 2×2 incluant des détails non classés.

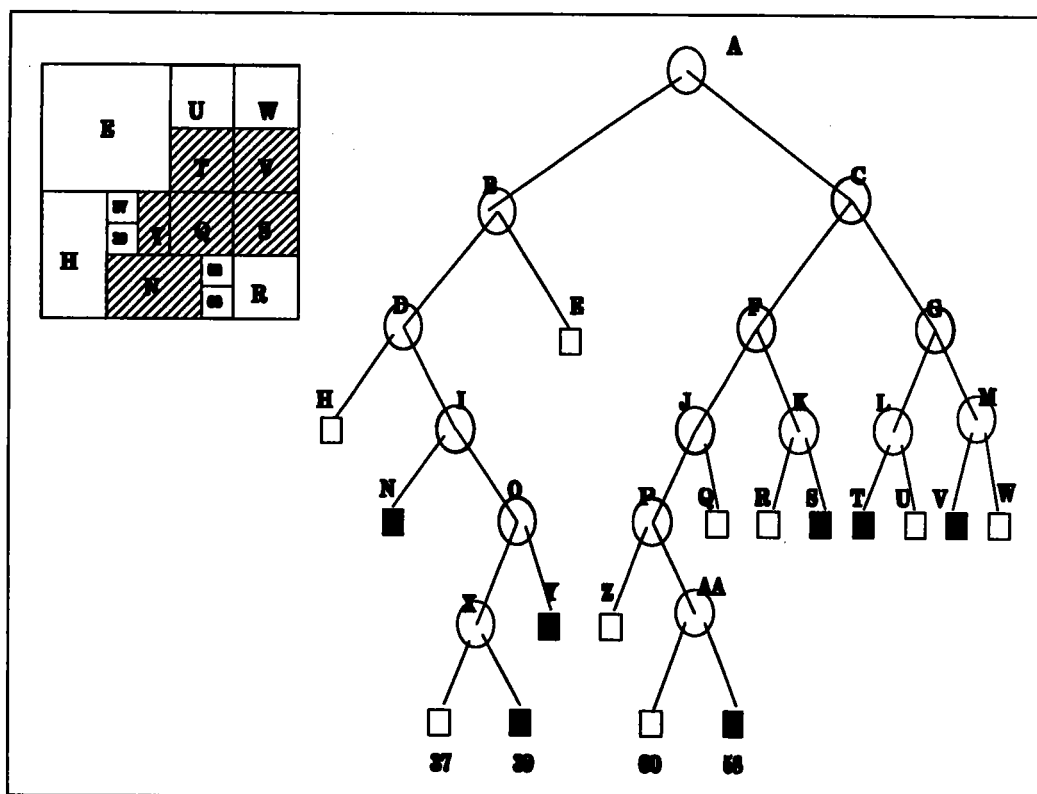


Figure N°15 Reproduction de l'image par arborescence

Les sous-blocs de pixels, dont les tailles sont égales à 2×2 et 4×4 , sont codés par un quantificateur vectoriel. Un codage par transformation est appliqué aux autres sous-blocs. L'information de position des blocs est délivrée par l'arbre binaire, elle est aussi transmise au décodeur ; la

reconstitution de l'image en dépend.

Cette méthode de codage adaptatif présente des taux de compression avantageux, tout en préservant une bonne qualité subjective de l'image restituée. En effet, les détails et les petites évolutions de luminances sont codés avec fidélité par un quantificateur vectoriel adéquat.

III - C - 2 Quantification vectorielle multi-étages

Une possibilité d'adaptation en codage d'images est d'appliquer plusieurs quantifications vectorielles successives. Cette éventualité offre au quantificateur la capacité de coder au mieux le signal image. En effet à la suite de la première quantification du vecteur incident, une seconde est réalisée par un quantificateur sélectionné d'après l'erreur de quantification. Cette même procédure peut se répéter N fois selon la précision de codage désirée.

L'étage initial de quantification utilise un dictionnaire de taille réduite formé de vecteurs pixels. Le suivant manipule, lui, des vecteurs erreurs de quantification. Ainsi d'étage en étage, l'écart entre le vecteur initial et son représentant s'amenuise. Le code final est la concaténation de toutes les adresses fournies à la suite des passages par les différents étages. Au décodage, l'image est restituée après la sommation de tous les vecteurs indexés lors des quantifications successives.

Un des avantages de la technique de quantification multi-étages est d'effectuer le codage en parcourant d'une façon sélective l'ensemble des dictionnaires stockés. Cette propriété justifie le bon rapport (qualité des images restituées / temps de calcul), meilleur que celui obtenu par un quantificateur à recherche exhaustive muni d'un dictionnaire de taille similaire. Cependant, la formation des dictionnaires reste un problème complexe; en effet si le premier dictionnaire est constitué par des vecteurs pixels, les suivants le sont par des vecteurs erreurs de quantifications. L'absence de caractéristiques connues ou d'erreurs de référence rend la conception des dictionnaires hasardeuse.

III - C - 3 Quantification vectorielle à dictionnaire rafraîchi

Pour les applications de transmission d'images animées, l'utilisation de

dictionnaires de grande taille est à bannir. En effet, le temps de calcul ainsi que la complexité des traitements sont des contraintes très sélectives pour ce type d'utilisations.

Dans [GOLD 86], est proposée une méthode de quantification vectorielle dont le dictionnaire initial est de petite taille. Conçu à la suite d'une estimation statistique des caractéristiques de l'image, ce dictionnaire est la concaténation de plusieurs sous dictionnaires locaux représentant différentes zones dans l'image. Ils sont transmis au décodeur au début de la communication. Un module de classification en entrée du codeur identifie la cellule d'appartenance de chaque vecteur. La quantification vectorielle par le sous-dictionnaire approprié est optimale si l'erreur de quantification est inférieure à une distorsion maximale prédéfinie. Si cette valeur de tolérance est dépassée, le dictionnaire initial est considéré insuffisant, son extension au vecteur incident est envisagée.

Goldberg et al proposent plusieurs techniques de restructuration du dictionnaire rafraîchi; identifier de nouvelles cellules englobant les nouveaux vecteurs ou redéfinir les représentants des cellules existantes en les remplaçant par la moyenne globale des vecteurs.

Parmi les performances annoncées, le taux moyen est de 1 à 1.5 bit/pixel pour des vecteurs formés par des blocs de 2 x 2 pixels.

La technique présentée ici est intéressante lorsque la scène à traiter comporte des zones de caractéristiques distinctes. Le recouvrement des sous-dictionnaires est, en effet, un problème non négligeable et lourd de conséquences; une représentativité médiocre des sous-dictionnaires initiaux nécessite un rafraîchissement important et donc une charge lourde de calcul pour le codeur.

D'autres problèmes de la quantification vectorielle à rafraîchissement du dictionnaire sont identifiées; monopolisation du canal par la transmission des sous-dictionnaires restructurés, complexité du classement simultané des sous-dictionnaires.

Chapitre IV

TECHNIQUES DE CODAGE PAR CLASSIFICATION VISUELLE

L'adaptabilité offerte par les méthodes classiques, citées dans le chapitre précédent, ne présente pas les performances espérées. Pour améliorer le taux de compression des images codées tout en gardant une qualité subjective tolérable à la restitution, une approche visuelle s'impose. Elle se traduit par une évaluation des échantillons du signal image et une quantification de leurs variations. Ces opérations permettent d'estimer l'impact visuel des différentes suites d'échantillons. Il est donc nécessaire d'inclure dans le codeur un module de traitement permettant d'identifier le signal image, de reconnaître ses composantes et leurs particularités perceptibles. De cette manière, à la transmission ne sont maintenues que les informations pertinentes indispensables à une restitution acceptable des images.

Pour chacune des trois techniques classiques, des approches visuelles sont citées dans la littérature. Mais celle qui se prête le mieux à ce genre de méthode est la quantification vectorielle. En effet, son aptitude à traiter globalement un ensemble d'échantillons et sa manière d'effectuer une recherche des plus proches voisins ou ressemblants parmi un ensemble de blocs d'éléments picturaux, la désignent logiquement à présenter le meilleur perfectionnement à la suite de l'application d'une classification visuelle.

Nous consacrerons donc le premier paragraphe de ce chapitre à la quantification vectorielle, donnant ainsi une importance moindre aux autres techniques.

IV - A Quantification vectorielle

Gresho et Ramamurthi ont été les précurseurs de l'utilisation de l'approche visuelle en quantification vectorielle [GERS 82] [RAMA 84].

Pour neutraliser les distorsions qui surviennent au niveau des contours sur les images restituées, une multitude de dictionnaires adéquats sont exploités. Ces quantificateurs vectoriels visuellement adaptatifs sont conçus pour améliorer le rendu des images codées, tout en maintenant un taux de compression élevé.

La technique décrite par Gersho et Ramamurthi identifie la ou les caractéristiques visuelles attachées au bloc d'image incident. Les principales particularités recherchées sont les contours et leur disposition dans l'image. D'autres méthodes, plus fines dans la classification, étendent les caractéristiques aux textures.

La méthode utilisée par Gersho, nécessite la formation pour chaque bloc (4 x 4) d'une table de gradients horizontaux et une autre de gradients verticaux. Ces tables sont calculées d'une façon locale; pour des pixels adjacents. Si la valeur du gradient dépasse celle de la moyenne affectée d'un coefficient pondérateur, le signe du gradient est retenu sur la table dans la direction qui lui correspond; si ce n'est pas le cas, c'est un zéro. Les tables matérialisant un contour sont reconnaissable par le nombre de signes (au moins trois) et par la disposition de ces derniers.

++++	+++ -	+++ -
++++	+++ -	++ - -
- - - -	+++ -	+ - - -
- - - -	+++ -	- - - -
++++	++ - -	++++
++++	++ - -	+++ -
++++	++ - -	++ - -
- - - -	++ - -	+ - - -

Une normalisation spécifique est appliquée à chaque type de bloc, de façon à faciliter la recherche de son plus proche vecteur voisin dans le dictionnaire qui correspond à chaque classe visuelle. Le traitement affecté au

classement entraîne une charge supplémentaire pour le codeur. Cependant, le gain qu'apporte cette répartition au niveau de la quantification vectorielle est significatif.

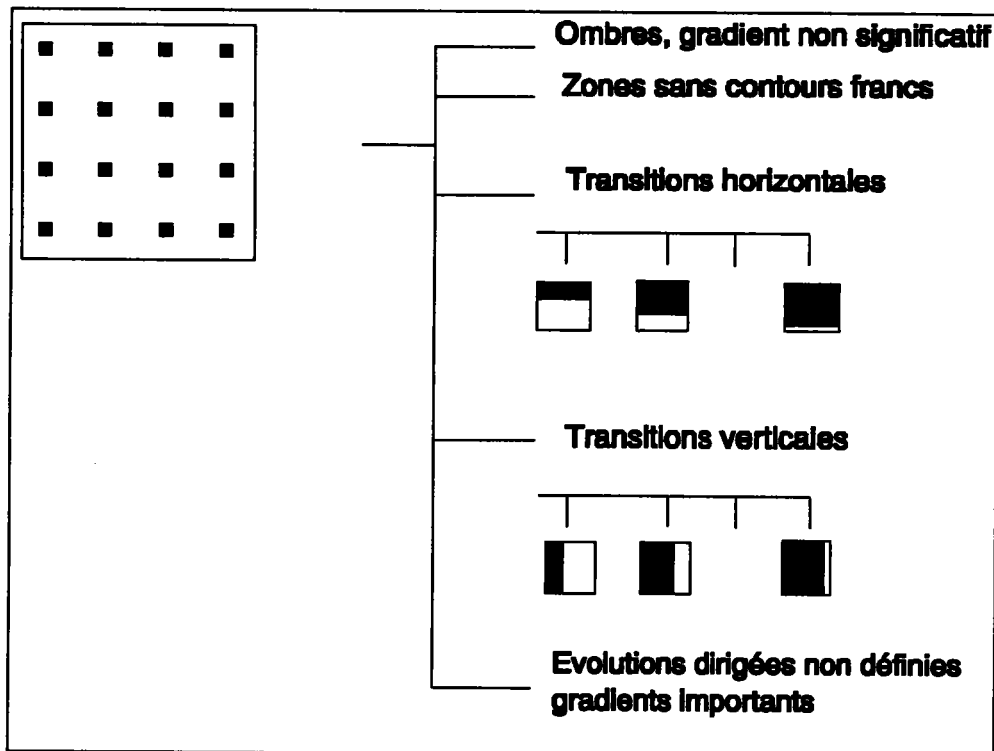


Figure N°16 classification visuelle pour la quantification vectorielle

Avec une approche algorithmique similaire à celles présentées dans [GERS 82] [RAMA 84], Y. Zhao propose un quantificateur vectoriel basé sur une identification des structures contenues dans l'image. Le procédé consiste à reconnaître plusieurs tendances de variation des luminances en se basant sur un vecteur de critères visuels.

L'évolution des luminances dans une image prend plusieurs formes : variations brutales, évolutives ou insignifiantes pour l'observateur. Ces évolutions traduisent les formes, les objets et les structures composant l'image. Leur localisation et leur exploitation pour le codage permet d'optimiser toutes les procédures de traitement, en maintenant une bonne qualité subjective pour les images restituées au décodage.

La quantification vectorielle exploite plusieurs sous-dictionnaires représentant les différentes familles visuelles; blocs uniformes, blocs contours et blocs composés de variations évolutives de niveaux de gris. Des paramètres sont identifiés pour représenter les particularités déterminant chacune des familles de blocs.

Les images sont découpées de façon contiguë, sans recouvrement, en blocs carrés de quatre lignes quatre colonnes. L'avantage offert par cette taille de bloc est la possibilité de localiser de petits détails dans l'image et de les prendre en compte au codage. Pour chacun des blocs (4 x 4), des gradients directionnels normalisés sont calculés localement. Cette dernière étape nécessite l'extraction des moyennes des luminances affectées aux couples de pixels adjacents dans le même bloc ainsi que les distances absolues qui les séparent:

$$M_v[l,c] = \frac{A[l,c] + A[l+1,c]}{2} .$$

M_v moyenne locale de deux luminances successives sur une même colonne.

$$l,c=0,1,2,3.$$

$$D_v[l,c] = \frac{A[l,c] - A[l+1,c]}{M_v[l,c]}$$

$$l = 0,1,2 \quad c = 0,1,2,3.$$

$D_v[l,c]$ gradient vertical local.

La même opération est réalisée pour les variations horizontales.

$$M_h[l,c] = \frac{A[l,c] + A[l,c+1]}{2} .$$

M_h moyenne locale de deux luminances successives sur une même ligne.

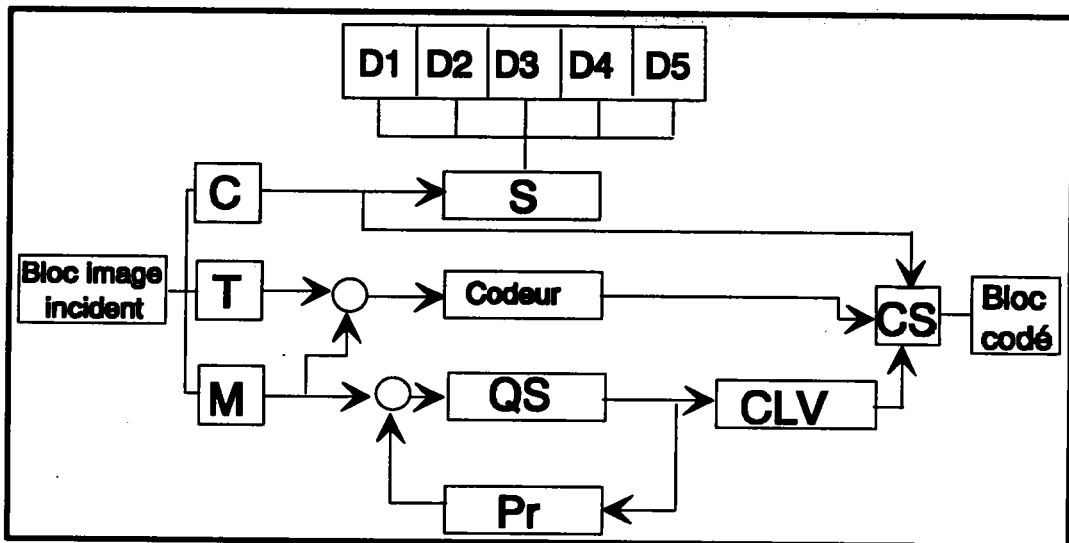
$$D_h[l,c] = \frac{A[l,c] - A[l,c+1]}{M_h[l,c]}$$

$$l = 0,1,2,3 \quad c = 0,1,2.$$

$D_h[l,c]$ gradient horizontal local.

La quantification de ces gradients aide à l'évaluation des variations de luminance. Leur comparaison avec des seuils prédéfinis contribue à

l'identification de la famille visuelle d'appartenance du bloc image. Les seuils sont déterminés à la suite de plusieurs essais subjectifs. Pour cette opération, la sensibilité des yeux pour les différentes variations de luminance est prise en compte, ainsi que l'élimination des effets provoqués par les bruits.



C : Classification visuelle de l'image.
T : Retard.
M : Calcul de la moyenne.
Pr : Prédicteur.
QS : Quantificateur scalaire 4 bits.
CLV : Codage à longueur variable.
CS : Codeur de sortie.
S : Sélecteurs de sous-dictionnaire.
Di : les différents sous-dictionnaires

Figure N°17 schéma synoptique du codage par quantification vectorielle visuelle adaptative.

Un arbre de sélection détermine la classe d'appartenance des blocs en entrée du codeur d'images. La fonction de décision est caractérisée par un vecteur de dimension deux. Elle tire parti des cumuls de dépassements des seuils affectés à chacun des deux gradients et comptabilisés par deux compteurs incrémentaux. En effet, à chaque fois que le seuil est dépassé par les valeurs caractéristiques, le compteur correspondant est incrémenté. Le total des dépassements détermine la classe visuelle.

Les classes visuelles suivantes sont identifiées :
- bloc de luminances uniformes.

- bloc à transition verticale.
- bloc à transition horizontale.
- bloc à transition simple de direction indéfinie
- bloc à transitions multiples de directions indéfinies.

L'organe de décision du codeur réside dans le module de classification visuelle, il sélectionne le sous-dictionnaire par l'intermédiaire duquel la quantification vectorielle est appliquée au bloc d'écart de luminances en entrée.

Pour chaque famille un sous-dictionnaire est conçu à partir de plusieurs images de référence. Le nombre de vecteurs contenus dans les différents sous-dictionnaires n'est pas identique; en effet, la probabilité d'appartenance aux différentes familles est étroitement liée aux structures qui composent l'image, elle dépend aussi des valeurs pré-établies des seuils de sélection. Les dictionnaires matérialisant les transitions verticales et horizontales restituent une bonne qualité d'image à partir d'un nombre limité de vecteurs : 64. Les dictionnaires représentant les transitions multiples de directions indéfinies ne donnent une qualité subjective acceptable qu'à partir de 256 vecteurs. Malgré, la faible probabilité d'apparition de cette catégorie de blocs le nombre de vecteurs nécessaire est important, à cause de la diversité des évolutions multiples non orientées dans l'image.

L'action de quantification est appliquée sur des vecteurs d'écart de luminance par rapport à la moyenne du bloc. Ce réajustement des composantes permet d'étendre le domaine de représentativité des dictionnaires.

$$M = \frac{1}{N} \sum_{i=0}^7 \sum_{j=0}^7 A(i,j) \quad (1).$$

M : La moyenne du médaillon:

Le vecteur à quantifier a pour composantes:

$$E(i,j) = A(i,j) - M. \quad \text{avec } i,j = 1,2,3,4.$$

La moyenne des luminances du bloc est traitée séparément par codeur prédictif différentiel. Le prédicteur utilisé est une fonction des valeurs moyennes attachées aux deux blocs précédents, le premier se trouve sur la même ligne que le bloc à coder et le second sur la même colonne que le bloc à coder. A la sortie du quantificateur scalaire nous obtenons des codes de

longueurs variables.

Chaque bloc est donc représenté par un mot code d'identification du sous-dictionnaire - sur 3 bits -, par l'adresse du vecteur écart le plus proche - sur 6 à 8 bits - et enfin par le code de sa moyenne qui vaut en moyenne 0,7 bits.

Le taux moyen avancé par Y. Zhao et de 0,61 bit/pixel. Cette compression maintient néanmoins un taux d'erreurs visible. En effet, des distorsions significatives sont perceptibles au niveau des contours. Toutefois, l'amélioration apportée par cette méthode de codage par classification visuelle est très marquante par rapport aux techniques classiques de quantification vectorielle.

Les méthodes de classification décrites auparavant nécessitent un charge de prétraitements très coûteuse pour l'équipement de codage d'images. Cette charge réside dans le calcul de plusieurs gradients locaux pour chaque couple de pixels. Pour contourner cette étape, A. Davignon a optimisé la classification visuelle [DAVI 89]. Une répartition de l'espace image en familles de transitions primitives orientées est effectuée. Elle est fondée sur le calcul de la moyenne du bloc image et de l'écart-type présenté par les luminances du bloc par rapport à cette moyenne.

L'activité des luminances est quantifiée à l'aide de seuils de décision. Si l'écart-type est supérieur à une certaine valeur, le bloc est déclaré contenant un contour. Cette opération permet de reconnaître les blocs uniformes. Une représentation binaire est choisie pour les médaillons actifs. A la position des pixels dont la luminance dépasse la valeur moyenne, la valeur "1" est affectée; pour les autres positions la valeur "0" leur est réservée.

Cette nouvelle description des blocs permet de différencier huit classes de transitions primaires orientées. Elles sont définies pour représenter les possibilités de contours horizontaux verticaux et diagonaux. Ces derniers sont appelés par l'auteur Blocs Binaires de Contours Fondamentaux : BBCF.

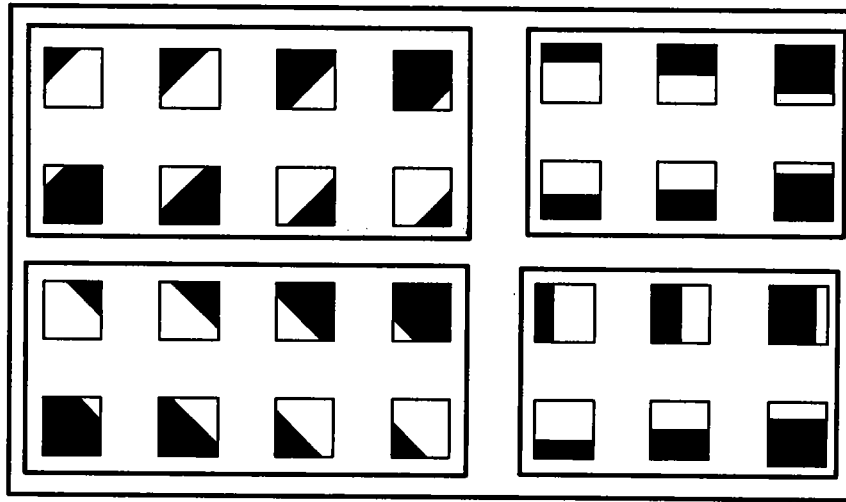


Figure N°18 Classification visuelles

Signalons que ces familles ne font pas une description exhaustive des transitions, mais elles y contribuent largement [DAVI 89]. Par une comparaison logique entre la reproduction binaire du bloc incident et les différents BBCF, le groupe d'appartenance du médaillon en entrée est déterminé. Cette opération est une quantification vectorielle manipulant une distance particulière élaborée par A. Davignon. Pour deux vecteurs a et b , la distance $d(a,b)$ les séparant est définie par la relation :

$$d(a,b) = \sum_{i=0}^7 Ci(a) \text{ EXOR } Ci(b)$$

Ci désigne la fonction qui retourne la i -ième coordonnée binaire d'un mot ou d'un vecteur. EXOR l'opérateur logique OU exclusif. Toutes les classes sont parcourues pour chaque bloc recherchant le BBCF qui lui est le plus proche. Ce dernier sera la représentation la plus convenable pour reconstituer l'image. La discrimination se fait en tenant compte d'un seuil maximum de distorsion égal à 3. Si la distance minimale séparant un bloc incident et son plus proche voisin BBCF dépasse ce seuil, le bloc en question est classé dans les "détails aléatoires", il sera codé avec une autre technique [DAVI 89].

Une autre particularité de la technique d'André Davignon est de segmenter l'image en blocs (8x8). Des tests d'identification des zones uniformes déterminent si cette caractéristique s'étend sur tout le bloc (8x8) ou si elle est seulement locale, ne concernant qu'une partie des médaillons (4 x 4) le composant. Cette nouvelle adaptation apporte un gain en débit; mais aussi une complexité supplémentaire pour l'équipement de codage-décodage.

IV - B Codage par transformé en cosinus discrète

Les approches adaptatives utilisant la TCD tirent parti de la répartition statistique de l'énergie dans le domaine transformé. Une classification détermine l'allocation des bits nécessaires pour coder le bloc transformé en fonction de l'étendu de son "spectre" énergétique. Toutefois, dans un besoin d'amélioration des performances des solutions s'appuyant sur une approche visuelle sont proposées.

Selon Lohsheller [LOHS 84] la classification visuelle est fondée sur quatre caractéristiques, grâce auxquelles des structures orientées sont détectées sur l'image. Les paramètres sont calculés à partir des luminances des pixels de l'image.

- La différence quadratique moyenne horizontale :

$$D_h = \frac{1}{N(N-1)} \sum_{m=0}^{N-1} \sum_{n=0}^{N-2} (x_{m,n} - x_{m,n+1})^2$$

avec $x_{m,n}$ est la luminance du pixel situé à la position (n,m) dans le bloc image.

- La différence quadratique moyenne verticale :

$$D_v = \frac{1}{N(N-1)} \sum_{m=0}^{N-2} \sum_{n=0}^{N-1} (x_{m,n} - x_{m+1,n})^2$$

- La différence horizontale maximale :

$$H_e = \text{MAX}_{m,n} | x_{m,n} - x_{m,n+1} |$$

- La différence verticale maximale:

$$V_e = \text{MAX}_{m,n} | x_{m,n} - x_{m+1,n} |$$

Une répartition grossière est élaborée à partir des deux premières caractéristiques D_h et D_v . A la suite de cette opération, cinq familles sont définies. Les frontières qui les délimitent sont déterminées par un vecteur de

paramètres fixes (T_1, T_2, T_3). Le paramètre T_3 permet de différencier les structures verticales, horizontales ou diagonales.

Suivant une règle discriminatoire, l'appartenance visuelle du bloc en entrée est confirmée. Cependant, cette préclassification est insuffisante, car la variation structurelle des images ne se limite pas aux classes primaires décrites. Cette constatation se traduit par le fait que les paramètres de sélection ne permettent pas de distinguer entre les blocs de contours multiples et les blocs représentant de larges plages de textures, ces deux types de blocs génèrent des vecteurs spécifiques similaires.

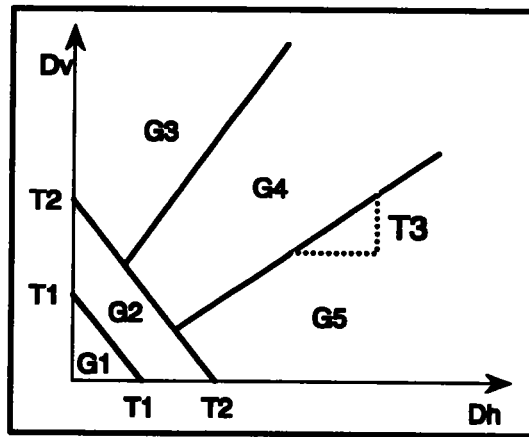


Figure N°19 Règles de pré-classification des blocs en 5 classes

Les caractéristiques H_e et H_v sont utilisées alors pour affiner la répartition. Les zones G_3 , G_4 et G_5 sont subdivisées pour former six nouvelles sous-classes. Le critère de sélection est un paramètre T_4 , appliqué à H_e et H_v . Le vecteur test final (T_1, T_2, T_3, T_4) est fixé de façon à obtenir une équirépartition des classes [LOHS 84-1].

La version améliorée du codage par classification de Lohsheller est fondée sur l'élaboration de seuils adaptatifs permettant la discrimination de certains coefficients transformés.

Par une suite de tests de visibilité des actions provoquées par les coefficients, une fonction $\gamma_{i,j}$ est déterminée [LOHS 84-2]. Elle identifie le seuil différentiel de visibilité pour chacune des fréquences de position (i,j) du bloc transformé. Le rapport de cette valeur avec l'écart type correspondant à la même position dans le bloc image fournit une fonction $W_{ai,j}$:

$$W_{ai,j} = 20 \log_{10} \frac{\gamma_{i,j}}{a_{i,j}}$$

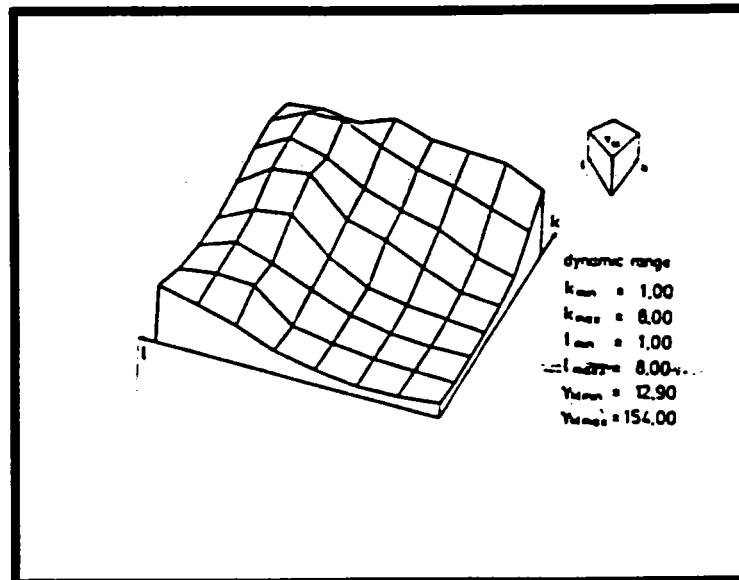


Figure N°20 Valeurs mesurées des seuils différentiels de visibilité $\gamma_{k,l}$ dans l'espace transformé (taille de bloc : 8x8)

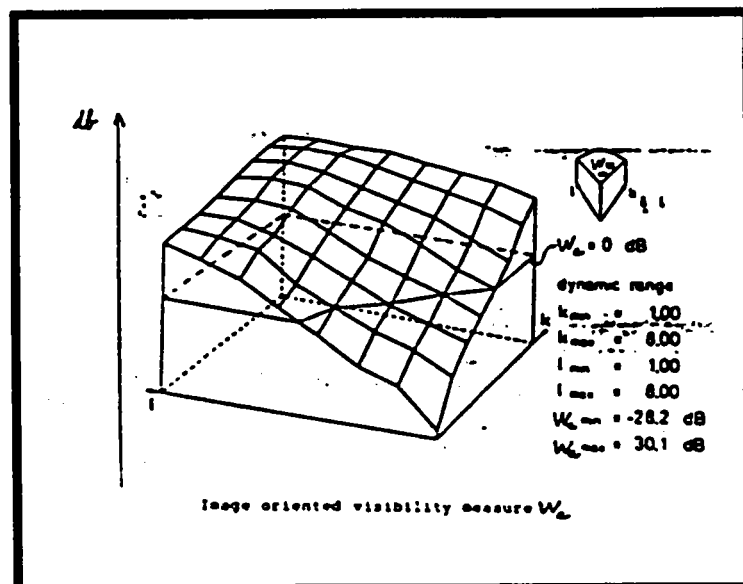


Figure N°21 Rapports normalisés des seuils différentiels de visibilité dans l'espace transformé ($W_a = \gamma / \sigma_a$ en db)

Cette fonction détermine l'impact visuel de chaque coefficient de la matrice transformée. Elle met en valeur tous les éléments dont l'amplitude $a_{i,j}$ est

inférieure à leur seuil de visibilité, nous obtenons pour ce type de coefficients un rapport $W_{ai,j}$ égal à 0 db.

Un quantificateur scalaire élabore les codes des coefficients en fonction de leur classe de visibilité et en fonction de l'importance du nombre de bits qui leurs sont alloués.

Pour l'ensemble des huit classes, le nombre de coefficients à coder varie de 1 à 16; cela nécessite un nombre total de bits variant de 10 à 67. Le coefficient matérialisant la composante continue $F(0,0)$ est traité à part. La longueur du mot code qui est réservé à $F(0,0)$ est différente de celles associées aux autres coefficients. En effet, pour ces derniers la taille du mot code varie entre 2 bits et 7 bits.

Les performances annoncées pour cette méthode de codage sont bonnes en termes de taux de compression à qualité subjective donnée : sur des images de type portrait, on obtient un débit moyen de 0.32 bit/pixel dont 0.274 bit/pixel pour le codage des coefficients.

IV - C Codage Différentiel

Cette technique a permis de développer des codeurs décodeurs avec de bonnes performances. Le fait de traiter séparément les éléments picturaux donne au codage prédictif une robustesse au niveau de la restitution des images. Mais dès que nous aspirons à obtenir un taux de compression important les erreurs de quantification prennent des proportions importantes et intolérables surtout autour des contours et des zones contrastées. Une adaptation aux structures composant l'image serait bénéfique pour les performances du codage.

L'analyse des dégradations introduites par le codage MICD et de leurs seuils de visibilité a poussé Dewitte et al [DEWI 81] à adapter la loi de quantification de l'erreur de prédiction aux particularités de l'image, et ce afin de gagner en performance de codage.

Afin de pouvoir choisir un prédicteur adéquat, une étude paramétrique définit l'appartenance de la zone image. Elle est fondée sur l'évaluation de la dynamique de l'évolution des luminances des pixels. Elle permet la distinction entre plusieurs catégories de blocs :

- Zones quasi-uniformes, où la dynamique locale (différence entre 2

points adjacents) est inférieure à 2% de la dynamique totale.

- Zones de détails peu contrastés, où la dynamique locale peut atteindre 5% de la dynamique totale. Dans ce cas comme dans le précédent, si l'erreur de prédiction est faible, la visibilité du bruit de quantification est élevée.
- Zones de détails contrastés, où la dynamique locale comme le gradient peuvent atteindre 15% de la dynamique totale. Si la visibilité de défauts y est plus réduite, elle est par contre difficile à modéliser, du fait du nombre de paramètres qui interviennent dans la description des textures.
- Zones de contours, dans ces plages le contraste des luminances est élevé (gradient supérieur à 15% de la dynamique totale) , mais la structure est beaucoup plus facilement identifiable : en effet, la largeur de transition est de 1 à 2 points d'image. Les défauts pouvant affecter l'image sont plus facilement modélisables [KRET 83].

Une étude statistique pour évaluer le nombre des transitions dans les images de télévision ainsi que leur impact sur le taux d'erreurs de prédiction a été menée. Elle montre une prédominance pour les évolutions primaires (verticales et horizontales). Mais le résultat le plus probant est la constatation suivante : les points qui présentent une erreur de prédiction dépassant 50 ou 60, dans le cas de l'utilisation d'un prédicteur fixe, correspondent pour leur majorité à des points appartenant aux contours. Cela signifie que si la prédiction de ces points particuliers pouvait être améliorée, la dynamique du quantificateur, c'est à dire la différence entre ces niveaux extrêmes, serait réduite et les performances du codeur en bénéficieraient.

Les constatations précédentes justifient tout à fait une adaptation de la méthode prédictive par un procédé basé sur l'extraction des particularités visuelles de l'image. Cette adaptation se traduit par un choix de prédicteur suivant l'évolution de luminances détectée.

Une projection spatiale décrit la direction du contour auquel appartient le pixel à coder. Cette direction est définie à la suite de la minimisation des erreurs de prédiction découlant de plusieurs prédicteurs (cf. Fig. N°24). En tenant compte des éléments déjà disponibles dans le récepteur, on affecte l'orientation du contour déterminée au codage et donc l'estimateur spatial adéquat à partir duquel la valeur de la luminance sera

reconstituée.

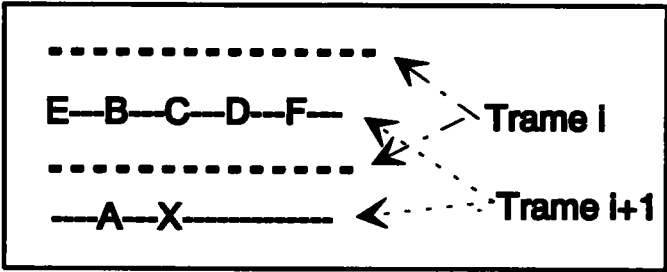


Figure N°22 Point utilisés par les prédicteurs

Orientations des contours	Prédicteurs I	prédicteurs II
153.4°	$\frac{A+B}{2}$	$\frac{2A + C}{3}$
135°	E	$\frac{2A + C}{3}$
116.6°	B	$\frac{A + 2C}{3}$
90°	C	$\frac{A + 2C}{3}$
63.4°	$\frac{C + D}{2}$	$\frac{2A + C + D}{4}$
45°	D	$\frac{A + D + F}{4}$
26.6	F	$\frac{2A + D + F}{4}$

Figure N°24 Exemples de prédicteurs adaptatifs

La diversité des cas envisageables fait que cette technique est très sensible aux erreurs de transmission. Elles produisent une désynchronisation entre l'émetteur et le récepteur. Ce genre d'erreur se propage sur le reste de l'image et affecte le calcul de prédiction, le choix du prédicteur ou celui du quantificateur.

Pour augmenter la stabilité des systèmes prédictifs adaptatifs face aux différentes erreurs auxquelles ils sont exposés, le rafraîchissement de l'image s'impose. Cette contrainte permet de limiter l'étendue des pixels affectés par les erreurs et de remplacer des valeurs erronées sur l'image. Au détriment du débit, le rafraîchissement de l'image renforce la robustesse du système de codage en comblant une de ses principales lacunes, la répercussion des erreurs sur les valeurs codées successivement.

Chapitre V

METHODES HYBRIDES DE CODAGE

L'association de plusieurs techniques de compression permet d'accumuler les performances du codage d'images. Nous nous proposons ici de traiter une méthode hybride qui nous a semblé intéressante du point de vue algorithmique est qui apporte une nouvelle approche du problème de codage sans cumuler la complexité des techniques utilisées.

V - A La quantification vectorielle appliquée dans le domaine transformé

L'utilisation de cette technique découle directement de la constatation d'après laquelle les coefficients obtenus à la suite d'une transformée mathématique ne sont pas complètement décorrélés. En effet, entre les coefficients persiste une fonction d'interdépendance. Cette corrélation résiduelle est due principalement à la matrice image initiale et aux évolutions des luminances qu'elle présente.

La distribution énergétique du spectre est variable en fonction du contenu de l'image, cela se répercute directement sur l'importance des représentativités des différents coefficients. De plus, cet état de fait est complètement quantifiable, en effet, un module de traitement classifie les blocs transformés selon des paramètres énergétiques. De cette manière sont identifiés les éléments nécessaires au codage.

Plusieurs auteurs ont développé des codeurs à base de quantification vectorielle des coefficients transformés. Dans [NASR 88], Nasrabadi et al cite plusieurs cas de figure, présentant chacun des approches particulières.

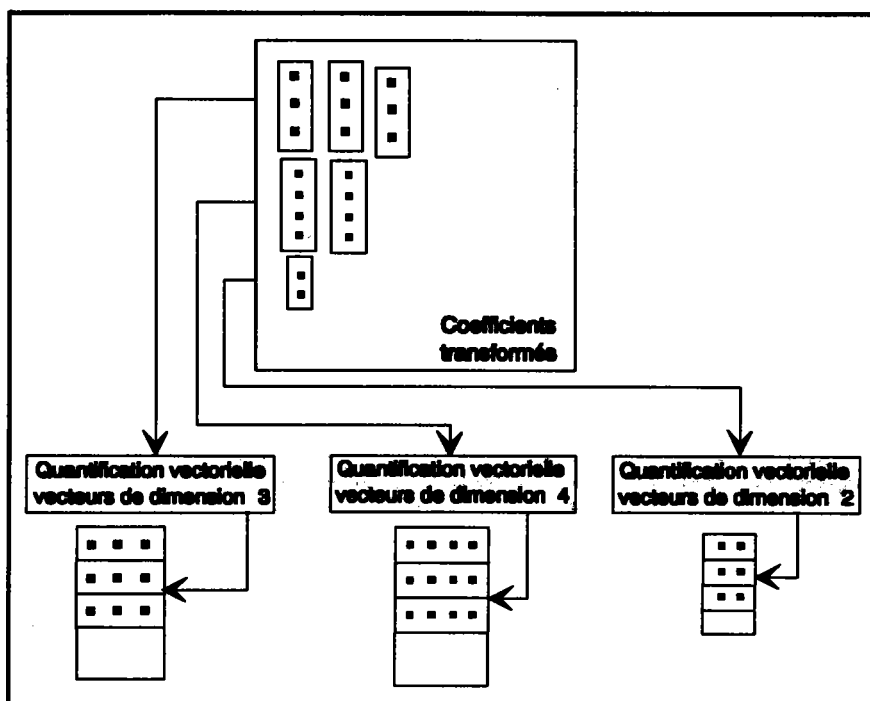


Figure N°25 Quantification vectorielle sur l'espace transformé
Multi-classes et dictionnaires

Nasrabadi [NASR 84], quant à lui, aborde le problème par des adaptations aux différentes classes énergétiques de la taille et de la position des vecteurs coefficients. Après avoir écarté les coefficients "hautes fréquences", l'adaptation est accomplie par la sélection de plusieurs vecteurs de dimensions différentes et ce, en correspondance avec les distributions statistiques dans la direction verticale de la matrice transformée. La taille des vecteurs est déterminée par un calcul préalable de fonction de corrélation entre les éléments d'un même vecteur. Un test permet de vérifier si le résultat est suffisant pour réaliser le codage du vecteur avec la dimension qui lui est affectée. La quantification vectorielle est appliquée à l'aide de dictionnaires formés de vecteurs de différentes tailles.

Le débit moyen obtenu se décompose en :

$$R = A_1 C_1 \log_2 \frac{M_1}{K_1} + \dots + A_i C_i \log_2 \frac{M_i}{K_i}$$

avec

- A_i Le nombre de fois où le dictionnaire i est utilisé.

- M_i La taille du dictionnaire.

- K_i La dimension de ses vecteurs.
- C_i Le pourcentage des coefficients retenus.

L'approche de Nasrabadi met en oeuvre un dispositif conséquent en temps de calcul pour réaliser la classification des blocs et la détermination de la dimension maximale de chaque vecteur.

V - B Codage hybride; adaptation visuelle

V - B - 1 Quantification vectorielle dans l'espace transformé

Pour des raisons d'amélioration des performances surtout de la qualité subjective, J. P. Marescq a étudié une technique de quantification vectorielle dans l'espace transformé, fondée sur l'exploitation des structures perceptibles sur l'image télévisuelle [MARE 86].

Une des particularités de cette méthode est qu'elle effectue un pré-classement suivant les propriétés visuelles des blocs d'image (4x4). Des dictionnaires adaptés à chaque famille sont constitués par la technique Linde Buzo Gray (LBG).

Une autre particularité de la méthode de Marescq est d'évaluer l'énergie totale de chaque bloc transformé. Cette entité intervient directement sur les tests de classements. Elle est calculée à partir des valeurs de tous les coefficients transformés sauf du représentant de la moyenne de la matrice (4 x 4) d'image :

$$A_{tot} = \sum_{i=1}^N (F_i^2) - F_1^2$$

F_i : Coefficients transformés.

F_1 : Coefficient de moyenne.

N : Nombre total de coefficients (dans ce cas:16).

Rappelons que pour garder une parfaite qualité visuelle des images reconstituées, le coefficient F_1 est transmis indépendamment des autres. En effet, 8 bits sont alloués pour le codage de F_1 , il présente une dynamique de 0 à 255.

L'activité A_{tot} , représente l'énergie totale véhiculée par les coefficients transformés. Ce terme donne une idée de l'activité spatiale des blocs, en quantifiant le taux de variation des luminances des pixels les composant.

Toutefois, si A_{tot} permet de répartir grossièrement les blocs dans des catégories d'activités différentes, il n'est pas suffisant pour distinguer des structures complexes : contours, textures etc.

Les éléments de la matrice obtenue après la transformée en cosinus discrète ont une signification semblable à ceux obtenus après une transformée de Hadamard. Ils sont des facteurs de pondération pour des transitions primaires ou des compositions de transitions.

La représentativité visuelle de la matrice transformée est partiellement identifiée. Quelques correspondances entre l'espace image et l'espace transformé sont exploitées. Cela se traduit principalement par une série de constatations :

- La première ligne, respectivement première colonne de la matrice transformée, en excluant l'élément qui correspond à la moyenne F_1 , reproduisent dans l'image essentiellement les contours verticaux, respectivement horizontaux.
- si la diagonale de la matrice transformée est importante, le coefficient F_1 toujours écarté, le bloc de pixel correspondant représente les médaillons d'image présentant une symétrie centrale.
- D'autres "tendances d'évolutions" sont identifiées à partir d'activités partielles.

Un vecteur de caractéristiques de perception est défini à partir d'activités partielles ($A_{diag}, A_{sup}, A_{inf}, A_{lig}, A_{col}$). Il est formé par la somme des carrés des coefficients localisés dans une zone de la matrice transformée et liés à une structure visuelle (cf. Fig.).

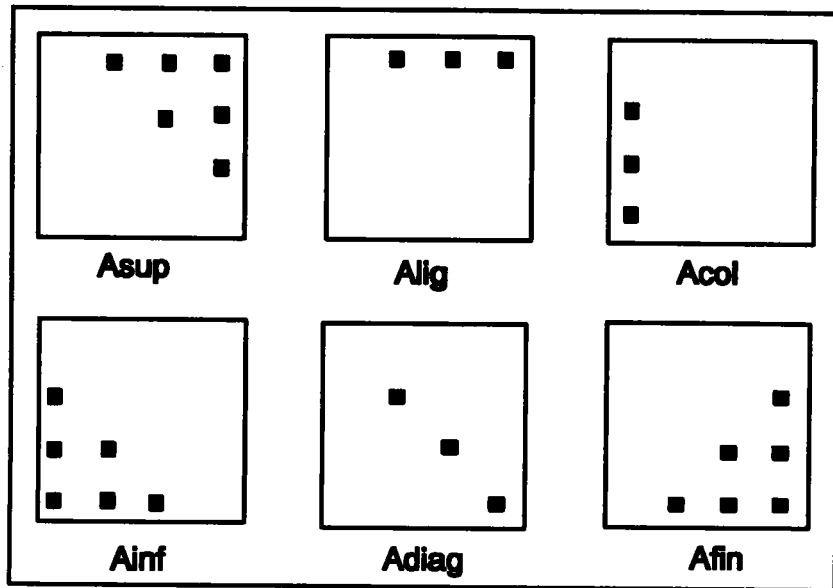


Figure N°26 Choix des coefficients transformés en fonction des classes

Un arbre de sélection, dont la fonction de décision exploite les activités spatiales partielles, est mis en oeuvre. Des paramètres caractéristiques λ sont définis pour aboutir à la répartition finale de l'ensemble des blocs de l'image par comparaison à des seuils:

$$\lambda_{diag} = \frac{A_{diag}}{A_{tot}}$$

$$\lambda_{sup} = \frac{A_{sup}}{(A_{tot} - A_{diag})}$$

$$\lambda_{inf} = \frac{A_{snf}}{(A_{tot} - A_{diag})}$$

$$\lambda_{lig} = \frac{A_{lig}}{A_{sup}}$$

$$\lambda_{col} = \frac{A_{col}}{A_{sup}}$$

Les blocs homogènes sont déterminés par une comparaison entre A_{tot} et un seuil variable défini par la loi de Weber transposée dans le domaine transformé [MARE 86]. Cette loi décrit la sensibilité visuelle par rapport aux variations de luminances. A_{tot} est aussi utilisé pour classer les blocs image suivant l'activité spatiale totale; blocs à variations faibles, blocs à évolutions brutales etc.

Les structures visuelles, quant à elles, sont identifiées par cette méthode à la suite de la localisation des coefficients transformés les plus significatifs. Cette approche discrimine les zones non représentatives. L'importance de l'activité "fréquentielle" d'une zone donnée du bloc transformé détermine la direction dominante du gradient de luminance du bloc image. Puisqu'à gauche de la première diagonale du bloc transformé (respectivement à sa droite) sont identifiés des coefficients représentant les blocs dont le gradient de luminance est vertical (respectivement horizontal).

L'arbre de sélection est conçu après une série de tests sur des images de référence présentant des composantes visuelles variées. Le vecteur caractéristique ($A_{diag}, A_{sup}, A_{inf}, A_{lig}, A_{col}$) et l'ensemble des seuils qui lui sont associés permettent une classification dite visuelle. Cette appellation est justifiée, puisque les classes obtenues se différencient par les particularités perceptibles : blocs homogènes, contours orientés etc..

Une variante de la méthode Marescq de classification est citée par C. Labit [LABI 88]. Elle introduit la reconnaissance des blocs image présentant des textures par un nouveau critère caractéristique A_{fin} en regroupant les carrés des coefficients "hautes fréquences".

$$\lambda_{fin} = \frac{A_{fin}}{A_{tot}}$$

Une fois les différentes zones de coefficients pertinents fixées, un quantificateur adaptatif traite les vecteurs formés à partir des éléments non écartés. Par un parcours en zigzag les composantes des vecteurs respectent la croissance du rang des "fréquences spatiales" qu'elle représentent.

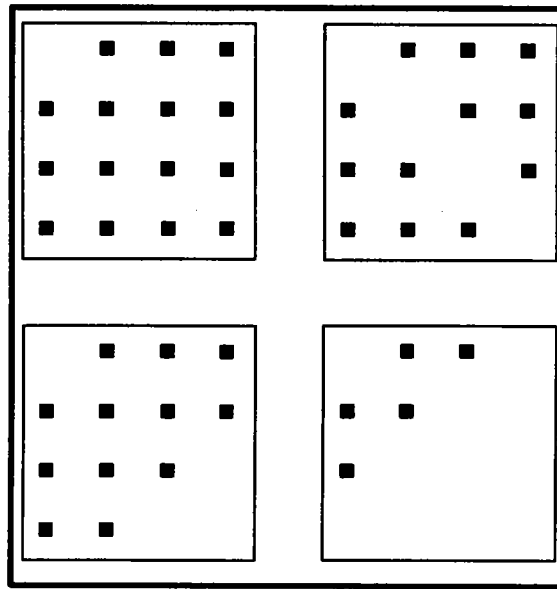


Figure N°27 Echantillonnage adaptatif des coefficients transformés selon les classes

Le quantificateur vectoriel exploite plusieurs dictionnaires correspondants aux 12 classes. Chacun d'eux est formé de vecteurs de dimension variant entre 5 pour les blocs homogènes et 15 pour les blocs images dépourvus de particularités visuelles.

La mesure de distorsion retenue est la distorsion vectorielle. Elle est appliquée aux vecteurs incidents pour trouver leurs représentant les plus proches dans les sous-dictionnaires. Ces derniers sont initialement constitués par l'algorithme LGB d'optimisation et de classement.

Les performances de cette technique de codage par classification visuelle ont été mesurées par le nombre moyen de composantes prises en compte pour le codage:

$$n_c = \sum_{i=1}^{\text{nb de classes}} p_i \dim \text{vect}_i$$

avec :

- $\dim \text{vect}_i$: dimension du vecteur de composantes de la classe i .
- p_i : probabilité d'occurrence de la classe i .

Le schéma de codage que nous avons décrit a été conçu pour les images télévisuelles. La principale contrainte requise était donc une parfaite qualité subjective des images décodées. Cette obligation fait que le nombre de

coefficients écartés reste très limité, cela agit directement sur la taille des vecteurs, donc, sur l'optimisation de la représentativité du dictionnaire.

Le quantificateur vectoriel reste assez rigide puisque son adaptation ne concerne que des caractéristiques limitées. En effet, pour répondre à la grande variété des structures visuelles composant les images, la taille du dictionnaire doit être assez conséquente. L'exemple présenté par C. Labit décrit un dictionnaire qui peut atteindre une taille de 4096 vecteurs.

Une solution est effectivement envisageable, il s'agit du rafraîchissement temporel du dictionnaire. Cela se traduit par un enrichissement simultané du dictionnaire avec de nouveaux vecteurs. Ceux dont la distorsion minimale par rapport aux vecteurs du dictionnaire dépasse le seuil de tolérance prédéfini, sont affectés au dictionnaire. Cette technique apporte une amélioration importante, mais elle apporte aussi une complexité accrue pour la restructuration des dictionnaires. En effet, la disposition des vecteurs dans le dictionnaire tient compte de leurs taux d'apparition pendant le codage. Les vecteurs les plus utilisés sont classés au début du dictionnaire de façon à éviter le parcours complet dudit dictionnaire.

V - B - 2 Quantification vectorielle dans l'espace image

Dans la famille des techniques hybrides nous nous proposons de détailler la méthode présentée par Y. Buhler dans [BUHL 86]. Les principaux avantages de cette technique résident dans l'utilisation d'un codage hiérarchique exploitant la transformée en cosinus discrète doublée d'une quantification vectorielle basée sur les particularités visuelles contenues dans l'image.

Classification visuelle

Le codage s'effectue sur plusieurs étapes. La première étant le fractionnement de l'image en blocs de tailles différentes, en se basant sur l'analyse du contenu de l'image. Ce découpage s'effectue à la suite d'une classification en fonction de l'activité de l'image. Il permet de former à partir d'images 256x256 des blocs de tailles 32x32, 16x16, 8x8 et 4x4 pixels. Le fractionnement et la restitution de l'image s'effectuent de gauche à droite et de haut en bas suivant l'algorithme "quadtree" qui détermine suivant la position et la taille d'un sous bloc donné la position et la taille du bloc suivant.

Une première répartition grossière détermine des caractéristiques

visuelles contenues dans l'image. Elle exploite les significations des histogrammes de luminance de plusieurs tailles de bloc. Une classification visuelle détermine les familles suivantes :

- Blocs de type 1; blocs uniformes ou quasi-uniformes, le contenu est composé essentiellement de plans formés de pixels dont la luminance varie peu. A la rencontre d'un bloc de ce type l'algorithme de fractionnement cesse la division pour obtenir des blocs de petites tailles. La distribution dans l'histogramme de l'ensemble des valeurs de luminance sur le nombre de pixels est restreinte. Cela se traduit dans l'histogramme par un pic de base étroite.
- Blocs de type 2; zones d'ombrages doux, elles renferment des plages de pixels dont la luminance varie spatialement d'une façon progressive. La disposition des pixels dans l'histogramme est caractéristique, elle est différente des dispositions correspondant aux autres classes visuelles. Pour le codage de ce type de bloc Buhler et al. appliquent la transformée en cosinus discrète TCD. La quantification des "fréquences basses" est suffisante pour la reconstitution du bloc image.
- Blocs de type 3 formés de textures fines. Ce sont des régions contenant beaucoup de détails. La classification par histogramme n'étant pas possible, car leurs caractéristiques sont semblables à celles des blocs uniformes ou quasi-uniformes. Une nouvelle caractéristique de répartition permet de différencier les deux familles. Elle quantifie les coefficients "hautes fréquences", obtenus après application de la TCD, à l'aide d'un paramètre *EHF* énergie hautes fréquences.
- Blocs de type 4 contenant des textures fortes. Les détails contenus dans ces zones sont pertinents et d'une grande importance visuelle. Ces zones sont codées par des blocs de taille 4x4 de façon à restituer leurs contenus.
- Bloc de type 5, ils représentent les contours nets. Ces parties de l'image sont indispensables à sa compréhension. Ces zones sont facilement reconnaissables par leurs histogrammes. Les amplitudes élevées et l'étendue spatiale importante sont autant de caractéristiques qui les différencient des autres structures visuelles. Pour leur codage, une grande fidélité est nécessaire. Un fractionnement maximal leur sera appliqué de façon à obtenir des blocs de 4x4. La quantification vectorielle adaptative permet de les coder à l'aide de dictionnaires appropriés.

Résumé

Le tableau présenté ci-dessous permet de comprendre la classification visuelle utilisée par Buhler [BUHL 86], ainsi que l'efficacité des paramètres caractéristiques qu'il exploite. En effet, les caractéristiques utilisées sont complémentaires. Dans le cas où deux familles se recouvrent partiellement et que l'étude de l'histogramme ne permet pas la détermination de l'appartenance d'un bloc image donné le recours au second paramètre *EHF* est dans ce cas indispensable.

Type de zone	Energie Hautes Fréquences	Histogramme des luminances
1	Très faible à faible	Plage étroite (pic)
2	Faible	Plage moyenne
3	Forte	Plage étroite
4	Forte	Plage très étendue
5	Assez forte à forte	Plage étendue

Tableau N°1 Identification des structures visuelles par les caractéristiques élaborées pour le codage.

Les caractéristique descriptives d'identification des familles visuelles sont les histogrammes, l'énergie des hautes fréquences et enfin les aspects visuels de l'image identifiés par l'étude des évolutions de luminance. Cette dernière caractéristique est utilisée pour les blocs 4x4 codés par quantification vectorielle avec adaptation visuelle.

L'histogramme des luminances est un outil pertinent pour l'analyse de l'image et de son traitement. Les histogrammes présentés ici sont normalisés par le nombre des pixels de la zone, de façon à permettre la comparaison de leurs signification au delà de la taille des blocs image correspondants. La valeur représentée dans les histogrammes pour une luminance donnée l sera notée $h(l)$ limitée par : $0 \leq h(l) \leq 1$.

Le critère de sélection *Act* extrait des histogrammes est déterminé par :

$$Act = - \frac{1}{M} \sum_{l=0}^{L-1} \log_2[h_n(l)].$$

$h_n(l)$ différent de 0.

$$h_n(l) = h(l) \cdot f_n(l - L) \quad .$$

Avec M désignant le nombre total des pixels de la zone. $h(l)$ désigne la valeur de l'histogramme pour la luminance l et $h_n(l)$ représente le nouvel histogramme pondéré par la fonction f_n de l'écart de la luminance l avec la luminance moyenne de la zone L . Afin que la contribution à l'activité Act augmente avec écart $(l - L)$, la fonction f_n a été choisie décroissante. Deux fonction ont été testées par Buhler : l'une agit comme l'inverse du carré de l'écart et l'autre comme une fonction gaussienne du même écart.

Le second critère de sélection est fondé sur le calcul de l'énergie des coefficients hautes fréquences obtenus à la suite de l'application de la transformée en cosinus discrète :

$$EHF = 1000 \cdot \frac{\text{Energie des coefficients H.F}}{\text{Energie totale}}$$

L'énergie des coefficients H.F. est la somme des valeurs des coefficients du plan TCD rejetés lors du codage par transformée. L'énergie totale est la somme des coefficients transformés, ce qui inclut l'élément (0,0) qui correspond à la moyenne de la luminance des pixels du bloc.

Codage par transformée en cosinus discrète

Il est appliqué à chaque fois qu'un bloc de taille 16x16 ou 8x8 est déterminé comme moyennement actif. Sachant que pour cette catégorie de bloc image, la matrice transformée en TCD présente des coefficients pertinents dans la zone des basses fréquences; un codage scalaire de ces coefficients est effectué. Il exploite des matrices d'allocation de bits similaires à celles décrites dans le chapitre des techniques adaptatives de codage par transformation TCD.

Codage par quantification vectorielle

Le fractionnement appliqué à l'image permet l'identification des zones refermant des pixels dont l'évolution de luminance est particulière. Ce morcellement aboutit à l'identification de bloc 4x4 formés de détails pertinents; contours nets ou textures fortes. Ce type de blocs est codé d'une façon personnalisée par un quantificateur vectoriel. Ces zones renferment la plus grande partie de l'information visuelle contenue dans l'image et leur codage doit être d'une grande fidélité.

L'expérience a montré que l'utilisation d'un dictionnaire non adaptatif donne des performances médiocres pour le codage. En effet, il est nécessaire d'utiliser des dictionnaires de plusieurs milliers de vecteurs pour avoir finalement des images décodées d'une très médiocre qualité subjective.

Pour coder efficacement les images en utilisant la quantification vectorielle, il est essentiel d'exploiter l'adaptation visuelle. Introduite par Ramamurthi et Gersho [RAMA 84] cette approche algorithmique est utilisée ici par Buhler [BUHL 86].

Le quantificateur vectoriel manipule des vecteurs de dimension 16 correspondant aux blocs de pixels 4x4. Une présélection détermine l'activité du bloc image traité et permet de définir le contour qu'il renferme. Les directions des contours retenues sont les directions verticales, horizontales et diagonales. Pour les deux premières familles de contours 14 dictionnaires ou livres de codes leur sont réservés, pour la famille des transitions diagonales 17 livres de codes représentent les différentes possibilités d'évolutions de la luminance avec un gradient dirigé en diagonale. Deux autres dictionnaires sont réservés aux blocs uniformes ou d'activité moyenne.

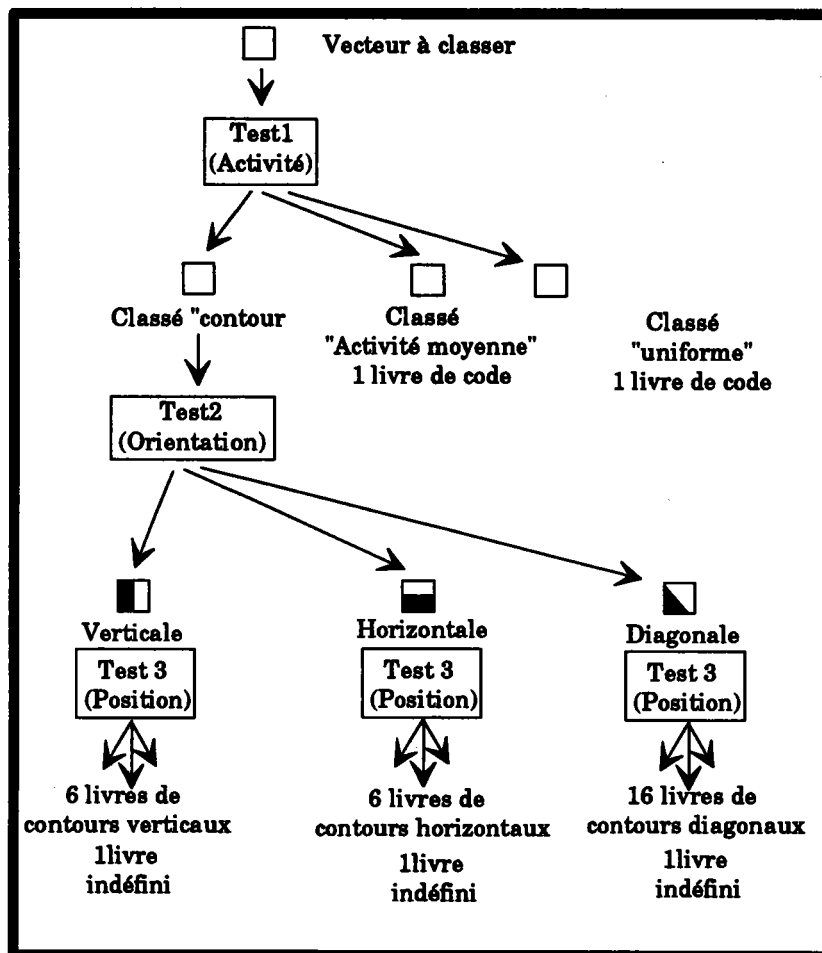


Figure N°27 arbre de décision employé pour la classification des vecteurs (33 livres de code au total).

L'algorithme de classement des blocs image est fondé sur le calcul de deux matrices de gradients $H(i,j)$ et $V(i,j)$ déterminant les directions d'évolution de la luminance horizontalement et verticalement. Les éléments de chaque vecteur ou bloc image sont notés $a(i,j)$. L'expression des deux gradients sont :

$$H(i,j) = \Phi(a(i+1,j) - a(i,j)) \quad ; i=1,..,3 \quad ; j=1,..,4.$$

$$V(i,j) = \Phi(a(i,j+1) - a(i,j)) \quad ; i=1,..,4 \quad ; j=1,..,3.$$

Avec $\Phi(x)$ fonction impaire de pondération. Sa valeur dépend de la hauteur de la transition. Elle est définie par le tableau suivant :

Hauteur de la transition x	$\Phi(x)$
$0 \leq x < 6$	0
$6 \leq x < 10$	1
$10 \leq x < 15$	2
$15 \leq x < 20$	4
$20 \leq x < 30$	6
$x > 30$	8

Tableau N°2 Valeurs de $\Phi(x)$ en fonction de la hauteur x de la transition.

Φ étant une fonction impaire, seules les valeurs positives sont données ici.

Pour déterminer la famille visuelle d'appartenance Buhler a exploité les gradients $H(i,j)$ et $V(i,j)$ pour différencier les évolutions de luminance dans chaque bloc image.

Des paramètres A_h A_v d'activité orientée sont calculés. leurs comparaison détermine la famille visuelle d'appartenance du bloc. Un arbre de décision discriminatoire réalise cette tâche.

$$A_h = \sum_{j=1}^4 \sum_{i=1}^3 |H(i,j)|$$

$$A_v = \sum_{j=1}^3 \sum_{i=1}^4 |V(i,j)|$$

Si $A_h A_v$ sont tous deux inférieurs à un seuil ϕ_1 , le bloc image est codé par le dictionnaire composé de blocs uniformes.

La somme des activités orientées détermine l'appartenance à la famille de blocs image d'activité moyenne. Si cette somme ne dépasse pas un seuil prédéfini ϕ_2 , le bloc incident est déclaré d'activité uniforme.

Si une des activités orientées $A_h A_v$ est supérieur au double de l'autre le bloc est considéré comme contenant un contour orienté, son codage est effectué par le livre de codes approprié.

Chaque dictionnaire réservé à des évolutions de luminance orientées renferme plusieurs sous dictionnaires. Chacun de ces livres de codes est adapté à la position de la transition contenue dans le bloc image. Dans le cas des dictionnaires de contours verticaux et horizontaux trois positions sont représentées de deux façons, une fois par des gradients positifs la seconde par des gradients négatifs. Le livre de codes correspondant aux évolution diagonales renferme des représentations de huit positions de contours.

Dans ce chapitre nous avons décrit quelques techniques exploitant les performances de la transformée en cosinus discrète et celles de la quantification vectorielle.

La méthode de codage que nous proposons est fondée sur ce même principe. Cependant, pour augmenter ses performances, nous nous sommes intéressés à la signification des coefficients transformés au point de vue visuelle.