

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE  
MINISTRE DE L'ENSEIGNEMENT SUPERIEURE ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE DE JIJEL



*Faculté des sciences exactes et de l'informatique*  
*Département d'informatique*

*Métriques pour l'évaluation automatique de la qualité  
des séquences vidéos numériques*

**Réalisé par :**

- **Bouldjadj Sara**
- **Boutaya Rabiha**

**Proposé par :**

**Mr. Mahrouk Zahir**

*Année 2015- 2016*

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE  
MINISTRE DE L'ENSEIGNEMENT SUPERIEURE ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE DE JIJEL



01  
02

*Faculté des sciences exactes et de l'informatique*  
*Département d'informatique*

*Métriques pour l'évaluation automatique de la qualité  
des séquences vidéo numériques*

**Réalisé par :**

- Bouldjadj Sara
- Boutaya Rabiha

**Proposé par :**

**Mr. Mahrouk Zahir**

*Année 2015- 2016*



# Remerciements

*Le grand merci s'adresse au **Bon Dieu Allah** le tout puissant, de nous avoir accordé santé, courage, volonté et patience pour mener ce travail à terme.*

*Nous exprimons d'abord nos profonds remerciements, notre vive reconnaissance et notre sincère gratitude à notre encadreur*

***Mr. Mahrouk Zahir** qui fait l'honneur d'avoir guidé et assisté tout au long de ce travail ; nous les remerciant pour tous leurs conseils, leurs disponibilité, leurs sérieux dans le travail et pour leur compréhension et leur patience.*

*Nos sincères remerciement aux membres de jury d' avoir bien voulu évaluer ce mémoire.*

*Nos sentiments de reconnaissance et nos remerciements vont également à tous toute la promotion «**Intelligence artificiel** »année 2016*

*Enfin, nous n'oublions pas de remercier tous ceux et toutes celles qui nous aidé ou encouragé, à quelque titre ou degré que ce soit et ceux qui ont contribué de près ou de loin à réaliser, à entreprendre et achever ce travail.*

# Dédicaces

*A l'aide de dieu tout puissant, qui m'a tracé le chemin de ma vie, J'ai pu réaliser ce travail que je dédie à :*

✦ *Mes très chers parents pour leurs sacrifices, soutien et amour. Je leurs serai éternellement reconnaissante.*

✦ *Mes chers frères et mes sœurs qui ont toujours été la pour moi.*

✦ *Mes cousins et cousines ;*

✦ *Tous mes proches;*

✦ *Tous mes amis (es);*

✦ *Tous nos enseignants qui nous ont enseigné du primaire jusqu'à ce jour;*

✦ *Tous ce qui m'a aidé de près ou de loin dans la réalisation de ce travail;*

✦ *Mes collègues et à toute la promotion intelligence artificiel année 2016*

*SARA*



# Dédicaces

*Je dédie ce travail à :*

- ✦ Celle qui m'ont soutenu pendant de longues années d'études aussi bien moralement que financièrement ; mes parents;*
- ✦ A celle qui m'a indiqué la bonne voie en me rappelant que la volonté fait toujours les grands ... à mon marie ;*
  - ✦ Mes chers frères, mes sœurs ;*
  - ✦ Toute ma famille sans exception ;*
  - ✦ Mes cousins et cousines ;*
  - ✦ Tous ce qui m'a aidé de près ou de loin dans la réalisation de ce travail ;*
  
- ✦ Mes collègues et à toute la promotion intelligence artificiel année 2016*

**RABIHA**



**Chapitre I : vidéo numérique**

Introduction générale.....	1
I.1 Introduction .....	3
I.2 Vision humaine .....	3
I.3 Principe de base .....	4
I.3.1 Image numérique .....	4
I.3.2 vidéo .....	4
I.3.2.1 Vidéo analogique .....	5
I.3.2.2 Vidéo numérique.....	6
I.3.2.3 Vidéo entrelacée.....	6
I.3.2.4 Vidéo progressive.....	6
I.3.3 Signale .....	7
I.3.4 Macroblock .....	8
I.3.5 Block.....	8
I.3.6 Débit.....	8
I.4 Couleur .....	8
I.4.1 Définition .....	8
I.4.2 Espace couleur .....	8
I.4.2 .1 Espace RGB .....	9
I.4.2 .2 Espace YCbCr .....	9
I.4.2 .3 Espace HSV .....	11
I.5 Système de la vidéo numérique .....	12
I.6 Structure d'un fichier vidéo .....	13
I.7 Les données d'une vidéo.....	13
I.8 Les formats de la vidéo numérique .....	14
I.9 Numérisation.....	15
I.10 Conclusion .....	16

**Chapitre II: traitement effectué sur la vidéo**

II.1 Introduction .....	17
II.2 Compression .....	17
II.2.1 Types de redondance .....	18
II.2.1.1 Redondance spatiale .....	18
II.2.1. 2 Redondance temporelle .....	18

II.2.2 Définition de codeur .....	18
II.2.3 Transformation en cosinus Discrète .....	19
II.2.4 Schéma de base de compression.....	19
II.3 Norme de compression .....	22
II.3.1 MPEG2 .....	22
II.3.1.1 Définition.....	22
II.3.1.2 Schéma de l'encodeur MPEG.....	22
II.3.1.3 Schéma de décodeur MPEG .....	23
II.3.1.4 Caractéristiques de la norme MPEG2.....	24
II.3. 2 H.264.....	25
II.3.2.1 Définition.....	25
II.3.2.2 Schéma de l'encodeur H.264 /AVC	
II.3.2.3 Schéma de décodeur H.264 /AVC.....	27
II.3.2.4 Caractéristiques de la norme H.264/AVC.....	27
II.3. 3 Codeur Erreur de transmission IP.....	32
II.3.3.1 Protocole IP .....	32
II.3.3.2 Description générale du modèle TCP/IP .....	33
II.3.3.3 Système de transmission de la vidéo par paquets.....	33
II.3.3.4 Caractéristiques de la transmission.....	34
II.3. 4 Codeur Erreur de transmission par wifi.....	35
II.3.4.1 Réseaux sans fils.....	35
II.3.4.2 Présentation générale de la norme WiFi (802.11).....	35
II.3.4.3 Diagramme émetteur / récepteur .....	36
II.3.4.4 Erreurs de transmission par un réseau sans-fil .....	36
II.4 Distorsions liées à la compression vidéo numérique.....	37
II.5 Conclusion .....	38

### Chapitre III : Métriques d'évaluation de qualité des vidéos

III.1 Introduction .....	39
III.2 Métriques subjectifs d'évaluation de qualité de la vidéo .....	39
III.2.1 Méthodes comparatives.....	39
III.2.2 Méthodes à simple stimulus .....	40
III.2.3 Méthodes à double stimuli.....	41
III.2.4 Méthodes à stimuli multiple.....	43

III.3 Avantages et inconvénients des métriques subjectifs .....	43
III.4 Métriques objective d'évaluation de qualité de la vidéo.....	44
III.4.1 Les métriques avec Références Complètes (RC).....	44
III.4.1.1 PSNR (Peak Signal to Noise Ratio) .....	45
III.4.1.2 SSIM (Structural Similarity Index).....	45
III.4.2 Les métriques avec Références Réduites (RR).....	47
III.4.2.1 Métrique de qualité vidéo VQM .....	47
III.4.3 Les métriques Sans Référence (SR).....	49
III.4.3.1 Métrique de qualité vidéo BLIIND .....	50
III.4.3.1.1 Modèle de la cohérence du mouvement.....	50
III.4.3.1.2 Modèle NVS (natural video statistics) .....	52
III.5 Indicateurs de performance de critères objectifs de qualité.....	52
III.5.1 Coefficient de corrélation linéaire : indicateur de précision .....	53
III.5.2 Coefficient de corrélation de rang : indicateur de monotonie.....	53
III.5.3 Outlier ratio : indicateur de cohérence.....	54
III.5.4 Erreur de prédiction de la qualité.....	54
III.6 Conclusion.....	55

## Chapitre IV : Expérimentation & résultats

IV.1 Introduction .....	57
IV.2 Problématique .....	57
IV.3 Environnement du travail.....	57
IV.3.1 Environnement matériel.....	57
IV.3.2 Environnement logiciel .....	58
IV.3.3 description de la Base de données LIVE .....	58
IV.3.3 .1 Nommage de fichier.....	59
IV.3.3 .2 Format de données subjectives .....	60
IV.4 Scénario de l'application.....	60
IV.5 Visualisation des vidéos YUV .....	61
IV.5.1 Algorithme de lecture d'une vidéo YUV .....	62
IV.6 Les principales fonctions implémentées .....	62
IV.7 Evaluation de la qualité de la vidéo dans le domaine spatial.....	63
IV.7.1 Implémentation de la métrique PSNR .....	63
IV.7.2 Implémentation de la métrique SSIM .....	64



---

IV.7.3 Implémentation de la métrique SSIM_diffé.....	64
IV.8 Evaluation de la qualité de la vidéo dans le domaine fréquentiel .....	66
IV.8.1 Implémentation de la métrique PSNR_DCT .....	66
IV.9 Comparaison entre les métriques d'évaluation de la qualité des vidéos .....	67
IV.10 Présentation de l'application.....	67
IV.11 Résultats expérimentaux .....	69
IV.11.1 Résultats de PSNR.....	70
IV.11.2 Résultats de SSIM .....	70
IV.11.3 Résultats de SSIM_différ .....	71
IV.11.4 Résultats de PSNR_TCD.....	71
IV.11.5 Résultats de SROOC de (PSNR, SSIM, SSIM_différ et PSNR_TCD).....	72
IV.12 Conclusion .....	74

## Chapitre I : vidéo numérique

<b>Figure I.1</b>	Vision humaine.....	3
<b>Figure I.2</b>	Elément d'une image (le pixel) .....	4
<b>Figure I.3</b>	Principe de balayage utilisé pour la vidéo .....	5
<b>Figure I.4</b>	Vidéo entrelacée et Vidéo progressive .....	7
<b>Figure I.5</b>	Signaux vidéo .....	7
<b>Figure I.6</b>	Spectre de la lumière blanche.....	8
<b>Figure I.7</b>	Maxwel.....	9
<b>Figure I.8</b>	Exemple de RGB.....	9
<b>Figure I.9</b>	Décomposition de l'image en composantes: luminance(Y) et chrominan (UetV). 10	
<b>Figure I.10</b>	Structure des macroblochs (a) 4 : 2 : 0, (b) 4 : 2 : 2 et (c) 4 : 4 : 4 .....	11
<b>Figure 1.11</b>	Représentation de l'espace couleur HSI .....	12
<b>Figure 1.12</b>	Système de vidéo numérique : capture, traitement et affichage .....	13
<b>Figure 1.13</b>	Hierarchie des données dans le flux vidéo.....	14
<b>Figure I .14</b>	convertisseur analogique-numérique (CAN).....	15
<b>Figure I.15</b>	schéma de numérisation .....	15

## Chapitre II : traitement effectuer sur la vidéo

<b>Figure II.1</b>	Schéma du codage INTRA .....	20
<b>Figure II.2</b>	Séquence type avec des images I, B et P .....	21
<b>Figure II.3</b>	Schéma du codage INTER prédiction .....	21
<b>Figure II.4</b>	: Encodeur MPEG-2 .....	22
<b>Figure II.5</b>	Décodeur MPEG-2 .....	23
<b>Figure II.6</b>	Exemple d'ordre de transmission et de réception des trames I, P et B .....	25
<b>Figure II.7</b>	Exemple de codeur MPEG2 .....	25
<b>Figure II.8</b>	Schéma bloc de d'encodeur H.264 .....	26
<b>Figure II.9</b>	Prédiction Intra.....	26
<b>Figure II.10</b>	Prédiction Inter .....	26
<b>Figure II.11</b>	Structure du décodeur H.264/AVC.....	27
<b>Figure II.12</b>	Mode d'Intra prédiction 16*16 .....	29
<b>Figure II.13</b>	Mode d'Intra prédiction 4*4 .....	30
<b>Figure II.14</b>	Exemple d'intra-prédiction du H.264 .....	31

<b>Figure II.15</b>	Partitions de macroblocs et de sous macroblocs.....	31
<b>Figure II.16</b>	Exemple de la compensation de mouvement basée sur les blocs .....	32
<b>Figure II.17</b>	Exemples de prédiction multi-références.....	32
<b>Figure II.18</b>	Image compressée.....	32
<b>Figure II.19</b>	Image filtrée.....	32
<b>Figure II.20</b>	Système de transmission de la vidéo par paquets .....	33
<b>Figure II.21</b>	Principe de la commutation de paquets .....	34
<b>Figure II.22</b>	Diagramme émetteur / récepteur.....	36
<b>Figure II.23</b>	Effet de fleu.....	38
<b>Figure II.24</b>	Effet de block.....	38
<b>Figure II.25</b>	Effet de ringing .....	38
<b>Figure II.26</b>	Effet de mosquito.....	38

### Chapitre III Métriques d'évaluation de qualité des vidéos

<b>Figure III.1</b>	Echelle comparative de l'I.T.U .....	40
<b>Figure III.2</b>	Exemple d'échelle de notation .....	41
<b>Figure III.3</b>	Illustration de la méthode DSCQS .....	42
<b>Figure III.4</b>	Illustration de la méthode DSIS sur deux présentations consécutives .....	42
<b>Figure III.5</b>	Métriques objective d'évaluation de qualité de la vidéo .....	44
<b>Figure III.6</b>	Diagramme de fonctionnement des métriques RC .....	45
<b>Figure III.7</b>	Diagramme de fonctionnement de SSIM .....	46
<b>Figure III.8</b>	Diagramme de fonctionnement des métriques RR.....	47
<b>Figure III.9</b>	Structure globale du critère de qualité vidéo VQM [Wolf et Pinson, 2002].....	48
<b>Figure III.10</b>	Tube d'extension spatio-temporelle ( $\Delta h$ , $\Delta v$ , $\Delta t$ ) utilisé dans VQM [Wolf et Pinson, 2002] .....	49
<b>Figure.III.11</b>	diagramme de fonctionnement des métriques NR .....	49
<b>Figure III.12</b>	Structure de la métrique BLIIND.....	50

### Chapitre IV : Expérimentation & résultats

<b>Figure VI.1:</b>	lecture YUV Toolkit .....	61
<b>Figure IV.1:</b>	Algorithme lecture de fichier YUV.....	62
<b>Figure IV.2 :</b>	L'interface principale de l'application .....	68



<b>Figure IV.3</b> : La fenêtre Entrer à l'application .....	68
<b>Figure IV.4</b> : PSNR des vidéos .....	73
<b>Figure IV.4</b> : SSIM des vidéos.....	73
<b>Figure IV.4</b> : SSIM_def des vidéos.....	73
<b>Figure IV.4</b> : PSNR_dct des vidéos .....	73
<b>Figure VI.5</b> : SROOC de (PSNR, SSIM, SSIM_différ et PSNR_TCD).....	74

## Chapitre I : vidéo numérique

<b>Table I.1:</b> formats vidéo numérique .....	15
---	----

## Chapitre 2 : traitement effectuer sur la vidéo

<b>Table II.1:</b> Caractéristiques de la norme MPEG2 .....	25
<b>Table II.2:</b> Caractéristiques de la norme H.264 .....	29

## Chapitre III Métriques d'évaluation de qualité des vidéos

<b>Table III.1:</b> Comparaison entre les métriques Full-Reference et Reduced-Reference et Non-Reference par le SROCC. ....	55
<b>Table III.2:</b> Comparaison entre les métriques Full-Reference et Reduced-Reference et Non-Reference par le PLCC. ....	55

## Chapitre IV : Expérimentation & résultats

<b>Tableau IV.1:</b> Caractéristique de l'environnement matériel de travail.....	57
<b>Tableau IV.2 :</b> Les fonctionnalités des éléments de la fenêtre accueil .....	69
<b>Tableau IV.3 :</b> Les fonctionnalités des éléments de la fenêtre fichier et les graphes .....	69
<b>Tableau VI.4:</b> Résultats de PSNR (WIFI, IP H.264, MPEG2) .....	70
<b>Tableau VI.5:</b> Résultats de SSIM (WIFI, IP H.264, MPEG2). ....	70
<b>Tableau VI.6:</b> Résultats de SSIM_diffé (WIFI, IP H.264, MPEG2). ....	71
<b>Tableau VI.7:</b> Résultats de PSNR_TCD (WIFI, IP H.264, MPEG2). ....	71
<b>Tableau VI.8:</b> SROOC de (PSNR, SSIM, SSIM_différ et PSNR_TCD). ....	72

Chapitre IV : Expérimentation & résultats

<b>Algorithme</b>	lecture de fichier YUV .....	62
<b>Algorithme</b>	PSNR_img.....	63
<b>Algorithme</b>	SSIM_img.....	64
<b>Algorithme</b>	SSIM_diffé .....	65
<b>Algorithme</b>	différence_frame .....	65
<b>Algorithme</b>	PSNR_DCT .....	66
<b>Algorithme</b>	DCT_vid .....	66
<b>Algorithme</b>	SROCC_vid.....	67



# *Introduction générale*

La vidéo numérique s'intègre de plus en plus dans notre quotidien à travers l'explosion des applications vidéo comme la télévision numérique, le cinéma numérique, la téléconférence, la vidéosurveillance, les services de partage des séquences vidéo (Youtube par exemple), la vidéo à la demande et bien d'autres.

La vidéo numérique passe obligatoirement à travers plusieurs étapes de traitement avant d'atteindre l'utilisateur final ; qui est généralement l'observateur humain. En revanche, ces traitements que subit la vidéo dégradent sa qualité visuelle. Les méthodes automatiques d'évaluation de la qualité vidéo (VQA : VideoQualityAssessment) jouent un rôle critique dans les systèmes de contrôle de la qualité des données visuelles dans plusieurs applications, à savoir le maintien des demandes en qualité de service (QoS : Quality of Service); l'estimation des performances des dispositifs d'acquisition et d'affichage vidéo ; l'évaluation des systèmes de traitement de la vidéo tel que les systèmes de compression, de transmission, ou de correction d'erreurs

La seule méthode fiable permettant d'évaluer la qualité visuelle d'une vidéo consiste à demander aux observateurs leurs avis, cette technique et connue sous le nom d'évaluation subjective de la qualité vidéo. Toutefois, l'évaluation subjective est excessivement coûteuse, lente et ne peut pas être intégrée dans un processus automatique ; d'où la nécessité d'avoir des méthodes permettant d'évaluer automatiquement la qualité visuelle de la vidéo. Néanmoins les méthodes subjectives restent toujours essentielles puisqu'elles procurent les données nécessaires pour estimer les performances d'un système automatique d'évaluation de la qualité, vu que le but principal est de se rapprocher le plus possible de la perception humaine.

Les méthodes permettant d'évaluer automatiquement la qualité de la vidéo sont connues sous le nom de méthodes objectives ou métriques objectives de la qualité vidéo. La plupart des métriques objectives proposées appartiennent à la famille des métriques à référence complète (FR : Full-Reference metrics) car elles exigent que la vidéo de référence soit en totalité disponible. En revanche, le signal vidéo de référence ne peut être acquis dans certaines applications ; ce qui a conduit les chercheurs à développer des métriques aveugles ou sans référence (NR : No-Reference metrics) où seulement la vidéo dégradée est utilisée dans le processus d'évaluation de sa qualité. Dans un troisième cas à mi-chemin entre les deux cas précédents, la vidéo originelle n'est pas disponible, mais certains de ses caractéristiques pouvant aider à évaluer la qualité de la vidéo dégradée, sont extraites et transmises sous forme

d'informations supplémentaires, et c'est le dernier type de métriques objectives de la qualité, qui est les métriques à référence réduite (RR : Reduced-Reference metrics).

Dans ce travail, nous allons proposer deux métriques permettant d'évaluer la qualité visuelle des vidéos. La première métrique opère dans le domaine spatial en utilisant le SSIM qui est calculé pour le résultat de différence de chaque deux images successives de la vidéo. La seconde métrique exploite le domaine fréquentiel en calculant la DCT de chaque deux images successives et ensuite le PSNR est calculé entre le résultat de la DCT de différence de la vidéo originale et la vidéo dégradée. Les deux méthodes sont implémentées et testées sur une partie de la base de données vidéo LIVE.

Le mémoire est organisé comme suit :

Le premier chapitre présente une synthèse des principaux concepts à connaître la technologie vidéo numérique

Le deuxième chapitre est un état de l'art de compression des vidéos, il définit le schéma de déroulement de compression vidéo, aussi que les normes de compression utilisées dans la base de donnée LIVE qui sont MPEG2, H.264, Erreur de transmission par wifi, et Erreur de transmission par IP, on présenterons leurs fonctionnement et leurs caractéristiques. A la fin de ce chapitre nous Passerons à parler sur les distorsions qui se produisent pendant la compression.

Le troisième chapitre est consacré aux métriques de qualité des vidéos, dans les deux sens subjective et objective (FR, RR et NR) en donnant quelques méthodes appartenant à chaque axe, et en terminant ce chapitre par les métriques de comparaisons entre eux.

L'environnement de travaille, les résultats expérimentaux obtenus et les commentaires de ces résultats sont présentés dans le quatrième et le dernier chapitre.



# *Chapitre I : vidéo numérique*

## I.1 Introduction

La vidéo est un domaine qui a connu une évolution importante ces dernières années. Dans ce chapitre, notre étude porte sur la vidéo numérique, qui est une vidéo analogique numérisée. L'objectif de la numérisation est d'apporter une précision, une qualité et une transmission aisée sur les réseaux de communication.

Pour bien saisir la notion d'une vidéo numérique, nous allons tout d'abord présenter les principes de base d'une vidéo en ligne, les différents types de vidéo, leur système, leur structure, leurs formats et leurs données.

## I.2 Vision humaine

La vision humaine est un processus complexe implique l'interaction quasi-simultanée des deux yeux et du cerveau au travers d'un réseau de neurones, de récepteurs et d'autres cellules spécialisées. La première étape de ce processus est la stimulation de récepteurs de lumière situés dans les yeux, la conversion du stimulus lumineux ou des images en signaux et la transmission de ces signaux électriques contenant l'information de la vision depuis chaque œil vers le nerf optique. Cette information est traitée en plusieurs étapes pour atteindre finalement le cortex visuel du cerveau [1].

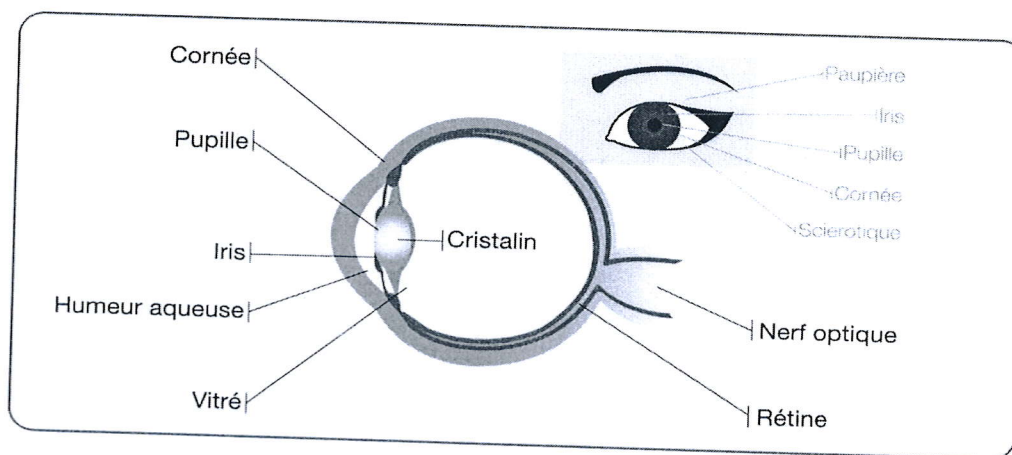


Figure I.1 : Vision humaine

La vue commence lorsque la lumière réfléchie d'un élément extérieur entre dans l'œil, la cornée oriente ces rayons lumineux vers le centre de l'œil et les fait pénétrer par la pupille. La lumière traverse le cristallin, formant ainsi une image inversée sur la surface de la rétine, membrane multicouche qui contient des millions de cellules sensibles à la lumière, appelées cônes et bâtonnets. Ces cellules photosensibles sont concentrées au centre de la rétine (fovéa).

Ce sont elles qui détectent l'image et la traduisent en série de signaux électriques qui seront transmis au cerveau. Plus précisément, ce sont les cônes (cellules qui permettent la vision des détails, des couleurs et des contrastes) qui sont rassemblées au centre de la rétine alors que les bâtonnets situent en périphérie, leur sensibilité à la lumière est environ 1000 fois plus importante que celle des cônes, donc ils permettent la vision nocturne (aussi appelée scotopique). Comme ils sont situés en périphérie de la rétine, les bâtonnets ne permettent qu'une vision des formes et des mouvements.

### I.3 Principe de base

#### I.3.1 Image numérique

**Définition :** Le terme d'image numérique désigne, dans son sens le plus général, toute image qui a été acquise, traitée et sauvegardée sous une forme codée représentable par des nombres (valeurs numérique).

**La résolution de l'image :** est le nombre de pixels par unité de longueur. Elle s'exprime en ppp (pixel par pouce) ou dpi (dot per inch). Le pouce (inch en anglais) vaut 2,54 cm.

Exemple : une feuille A4 numérisée en 300 ppp correspond à une trame de

$$\frac{300}{2,54} \times 21,0 = 2480 \text{ pixels} \quad \text{sur} \quad \frac{300}{2,54} \times 29,7 = 3508 \text{ pixels}$$

Le fichier est composé de  $2480 \times 3508 = 8699840$  pixels soit à peu près 8,7Mpixels.

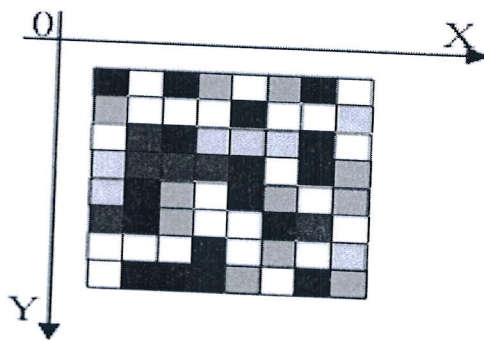


Figure I .2 : Élément d'une image (le pixel).

#### I.3.2 Vidéo

Le mot vidéo vient du latine **video** qui signifie " je vois ".il est utilisé pour désigner une succession d'images a un certaine cadence, leur principe est que L'œil humain a comme caractéristique d'être capable de distinguer environ 20 images par seconde. Ainsi, en affichant plus de 20 images par seconde, il est possible de tromper l'œil et de lui faire croire à une image animée. On caractérise la fluidité d'une



vidéo par le nombre d'images par secondes (en anglais frame rate, FPS), d'autre part la vidéo au sens multimédia du terme est généralement accompagnée de son, c'est-à-dire de données audio [4]. Il existe deux grandes familles de systèmes vidéo: les systèmes vidéo analogiques et les systèmes vidéo numériques [5].

### I.3.2.1 La vidéo analogique

La caméra balaye l'image bidimensionnelle qu'elle a devant elle par un faisceau d'électrons qui se déplace très rapidement de gauche à droite et plus lentement de haut en bas et produit une tension en fonction du temps. Elle enregistre ainsi l'intensité lumineuse, et à la fin du balayage, on a alors une trame. Le faisceau revient à l'origine pour recommencer. Le récepteur va recevoir cette intensité en fonction du temps, et pour reconstruire l'image, va répéter le processus de balayage. Il existe deux systèmes selon les paramètres du balayage:

- **Le système PAL/SECAM** (Phase Alternating Line / SEquentiel Couleur Avec Mémoire) : utilisé en Europe, ce système utilise 625 lignes (seulement 576 sont affichées), un rapport vertical/horizontal de 4/3 et un débit de 25 images par seconde.
- **Le système NTSC** (National Television Standards Committee) : Utilisé au Japon et en Amérique, ce système utilise seulement 525 lignes (483 affichées) et un débit de 30 images par seconde (figure I.3).

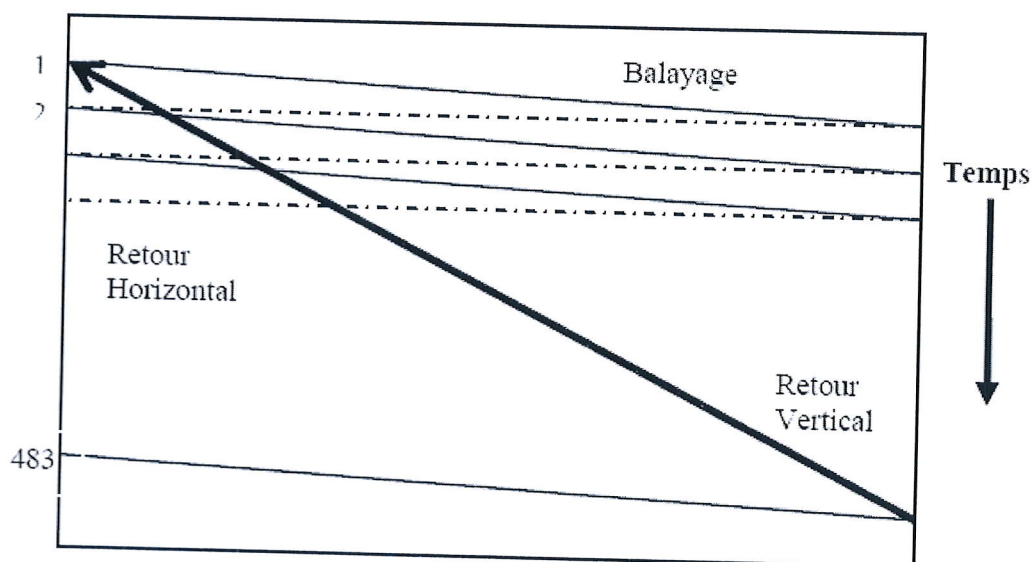


Figure I.3 : Principe de balayage utilisé pour la vidéo.

### I.3.2.2 Vidéo numérique

La vidéo numérique est une suite d'images formées d'une matrice de pixels qu'il est caractérisé par le nombre d'images par seconde (FPS). Pour obtenir des images en couleur, il faut utiliser au moins 8 bits par pixel, ce qui correspond à 256 couleurs. En fait, avec 8 bits par pixel, on obtient de la vidéo numérique noir et blanc de haute qualité. Pour la vidéo numérique couleur, on utilise 8 bits pour chaque couleur RVB, soit donc 24 bits par pixel, ce qui correspond à environ 16,8 millions de couleurs. Le principe de balayage utilisé est similaire à celui de la vidéo analogique.

### I.3.2.3 vidéos entrelacées

Les vidéos entrelacées permettent d'atteindre un taux de rafraîchissement élevé, sans augmenter la bande passante. Plutôt que de rafraîchir toute l'image, on rafraîchit les lignes paires puis on rafraîchit les lignes impaires à la prochaine image. Cette technique permet de capturer des mouvements plus rapides. Par contre, cette technique entraîne un artefact appelé peigne d'entrelacement. Cet artefact peut être corrigé par les logiciels de traitement et de montage vidéo (désentrelacement).

### I.3.2.4 Les vidéos progressives

Les vidéos progressives rafraîchissent l'image entière. Ceci permet d'éviter les artefacts d'entrelacement, mais nécessite deux fois plus d'information qu'une vidéo entrelacée. Il existe plusieurs standards de vidéo haute-définition (HD) entrelacée et progressive désignées respectivement par la lettre i (interlaced) ou p: 720p, 720i, 1080p ou 1080i. Le chiffre correspond au nombre de lignes verticales. Le nombre de lignes horizontales peut être calculé en sachant que le ratio d'aspect d'une vidéo HD est de 16/9.

- $720 / 9 * 16 = 1280$  d'où un format 1280 x 720.
- $1080 / 9 * 16 = 1920$  et on a alors 1920 x 1080.

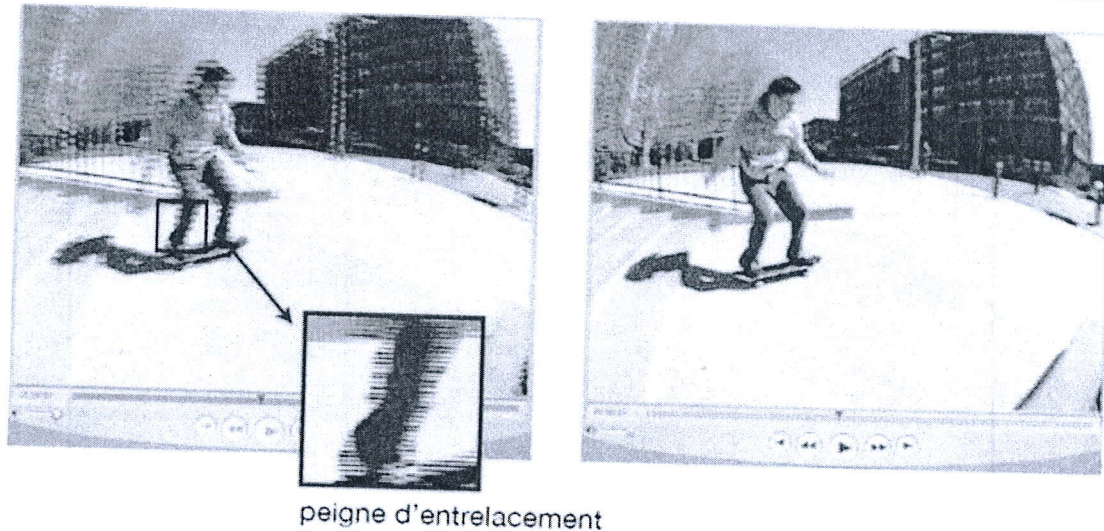


Figure I.4 : Vidéo entrelacée et Vidéo progressive.

### I.3.3 Signale

**Les signaux analogiques :** qui varient de façon continue dans le temps (intensité sonore, intensité lumineuse, pression, tension), c'est-à-dire qu'ils peuvent prendre une infinité de valeurs différentes.

**Les signaux numériques :** Les signaux numériques, sont exclusivement transmis sous forme de points sélectionnés par intervalles sur la courbe. L'ordinateur peut utiliser un signal numérique de type **binaire**, qui décrit ces points sous la forme d'une suite de valeurs minimales ou maximales correspondant respectivement au zéro et au un.

Cette suite de zéros et de uns peut ensuite être interprétée à la réception comme un ensemble de nombres représentatifs de l'information émise à l'origine. [6]

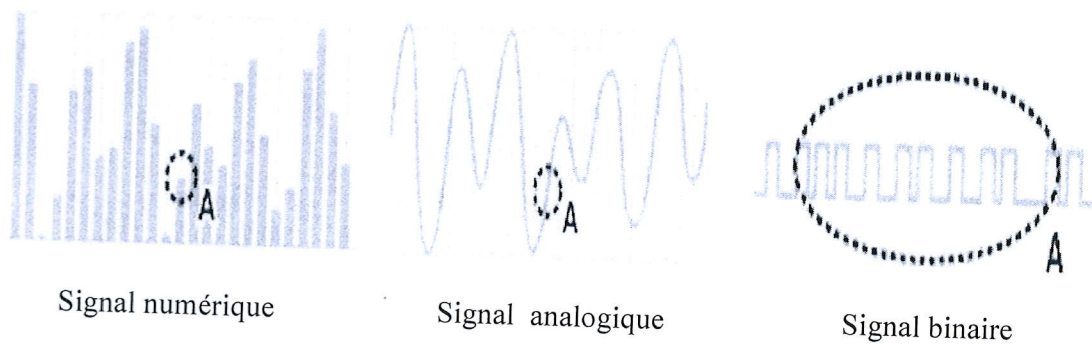


Figure I.5 : Signaux vidéo.



### I.3.4 Macroblock

Le terme macrobloc (en Anglais macroblock) désigne deux notions :

- une unité de traitement dans les formats d'image et de compression vidéo basée sur les transformées de blocs linéaires, Un macrobloc se compose généralement de  $16 \times 16$  échantillons, et est encore subdivisé en blocs de transformation, et peut être encore subdivisé en blocs de prédiction.
- un phénomène, effet ou artefact pouvant survenir lors de la compression vidéo numérique ou du décodage de ce type de signaux. Dans ce cas on évoque l'apparition de macroblocs à l'image.

### I.3.5 Block

Le block est la plus petite entité utilisée dans le codage spatial.

### I.3.6 Débit

Le débit est la quantité de données informatiques transmises par secondes pour définir un flux (audio/vidéo), elle est exprimée en **kbs** (kilobits par seconde).

## I.4 Couleur

### I.4.1 Définition

Les couleurs trouvent leur origine dans la séparation de la lumière blanche naturelle en composantes absorbées et composantes réfléchies. Toute source lumineuse visible est composée d'un mélange d'ondes électromagnétiques cohérentes, dont la longueur d'onde est comprise entre 400nm (violet) et 700nm (rouge) [7] :

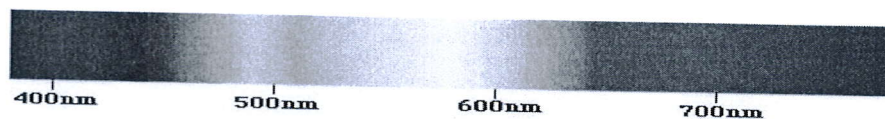


Figure I.6 : Spectre de la lumière blanche, couleurs d'arc en ciel.

### I.4.2 Espace couleur

La couleur est généralement représentée par trois composantes qui définissent un espace de couleurs. Il existe plusieurs espaces colorimétriques qui ont chacun certaines caractéristiques intéressantes.

### I.4.2 .1 Espace RVB ou RGB

L'espace RGB est l'espace vectoriel engendré par les 3 composantes primaires (Rouge, Vert, Bleu), adapté pour la capture vidéo et l'affichage. Il est le plus utilisé, puisqu'il intervient dans la plupart des appareils de prise d'images couleurs, ainsi que dans les moniteurs couleur. La plupart des formats de codage d'images (GIF, BMP, PPM...) utilisent cet espace de couleurs.

La représentation des couleurs dans cet espace donne un cube de Maxwell comme illustré dans la figure :

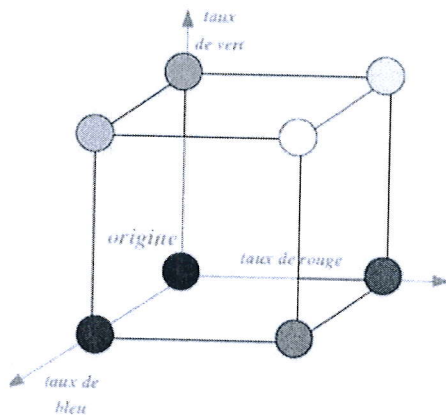


Figure 1.7 : Maxwell



Figure 1.8 : Exemple de RGB.

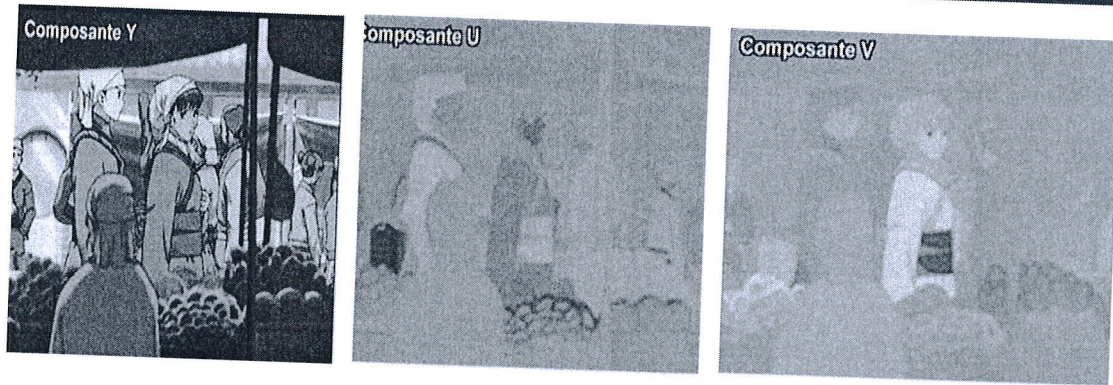
### I.4.2 .2 Espace YCbCr ou YUV

Ce système est principalement utilisé pour le stockage et la transmission des signaux vidéos, et utilisé pour les images JPEG. Ce modèle colorimétrique permet en effet de réduire la taille d'une image ainsi on transforme les composantes RGB de l'image en une composante de luminance notée **Y** et en deux composantes de chrominance notées **Cb** et **Cr**. Pour cela on utilise des formules de passage (pour un pixel).

$$y = 0.2989 * R + 0.5866 * G + 0.1145 * B \quad \dots 1$$

$$U = -0.1688 * R - 0.3312 * G + 0.5000 * B \quad \dots 2$$

$$V = 0.5000 * R - 0.4184 * G - 0.0816 * B \quad \dots 3$$



**Figure I.9** : Décomposition de l'image en composantes: luminance(Y) et chrominance (U et V).

La reconversion de l'espace de représentation YUV vers l'espace RGB est nécessaire pour l'affichage des images sur un écran. Cette reconversion s'obtient par le système d'équations suivant :

$$R = 1 * Y + 0 * U + 1.402 * V \dots 4$$

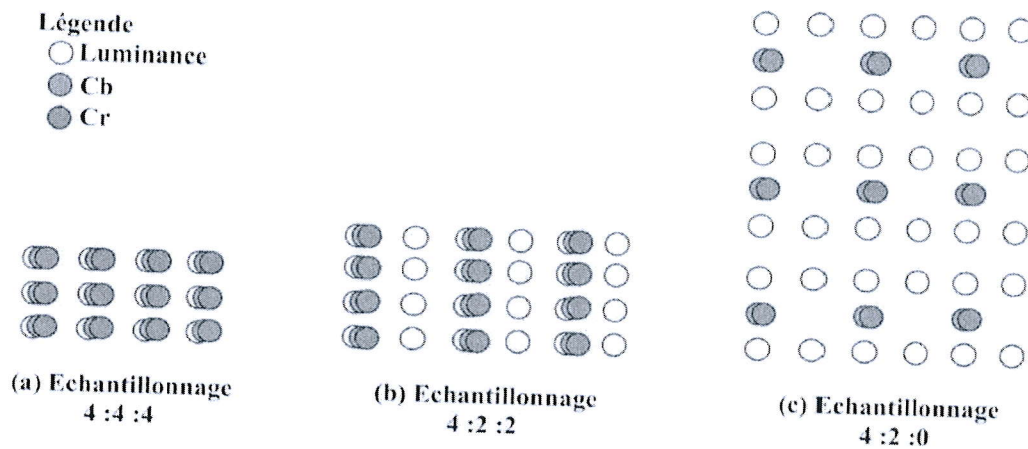
$$G = 1 * Y - 0.344 * U - 0.714 * V \dots 5$$

$$B = 1 * Y + 1.772 + 0 * V \dots 6$$

Les images vidéo comprennent trois matrices rectangulaires de données de pixels, lesquelles représentent le signal de luminance (Y) et deux signaux de chrominance (chroma Cb et Cr). Ces matrices correspondent à une représentation décomposée des trois couleurs primaires associées à chaque élément d'image. Pour la représentation des pixels vidéo numérique, il existe plusieurs formats.



## Format d'échantillonnage



**Figure I.10 :** Représentation des modèles d'échantillonnage.

- 4:4:4 : les trois composants (Y, Cb and Cr) ont exactement la même résolution et donc un échantillon de chaque composant existe à chaque position des pixels. Pour quatre échantillons luminance, il y a quatre échantillons Cb et quatre échantillons Cr.
- 4:2:2 : les composants chrominance ont la même résolution verticale que la luminance, mais la moitié de la résolution horizontale.
- 4:2:0 : les deux composantes de chrominance Cb et Cr ont la moitié de la résolution verticale et la moitié de la résolution horizontale de Y.
- 4 : 0 : 0 : consiste à ne conserver que la composante luminance Y.

### I.4.2 .3 L'espace HSV

HSV Aussi connu sous le nom de système de cône hexagonal. Le principe de cet espace est de caractériser les couleurs de façon plus intuitive et de se rapprocher de la perception humaine, Les trois composantes de cet espace sont la Teinte (Hue), la Saturation (Saturation) et l'Intensité (Value).

- **Teinte :** est un angle de  $0^\circ$  à  $360^\circ$ , représente le nom qu'on utilisera pour désigner la couleur.
- **Saturation :** c'est le taux de pureté de la couleur, qui doit varier entre la pureté maximale (couleur éclatante) et l'achromatisme (niveau de gris). la valeur est calculée à partir de 0 à 1. Lorsque la valeur est 0, la couleur est grise et lorsque la valeur est 1, la couleur est une couleur primaire.

- **Intensité** : c'est la mesure de lumineuse de la couleur, qui doit varier entre le noir absolu et le blanc.

$$\text{où } H = \arctan\left(\frac{\beta}{\alpha}\right) \dots 7$$

$$S = \sqrt{\alpha^2 + \beta^2} \dots 8$$

$$I = \frac{(R+G+B)}{3} \dots 9$$

$$\text{Avec } \alpha = R - \frac{1}{2}(G + B)$$

$$\beta = \frac{\sqrt{3}}{2}(G - B)$$

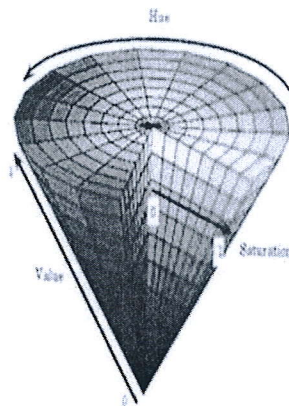


Figure I.11 : Représentation de l'espace couleur HSI.

## I.5 Système de la vidéo numérique

Un système vidéo numérique est illustré sur la figure I.12. En entrée du système, une scène visuelle est capturée, typiquement avec un appareil-photo. Elle est ensuite convertie sous forme numérique (numérisation).

Ce signal numérique peut alors être manipulé dans le domaine de différentes manières (traitement, stockage et transmission).

En sortie de système, le signal de vidéo numérique est montré à une visionneuse par la reproduction de l'image vidéo (ou la séquence vidéo) suivant un affichage deux dimensions [8].

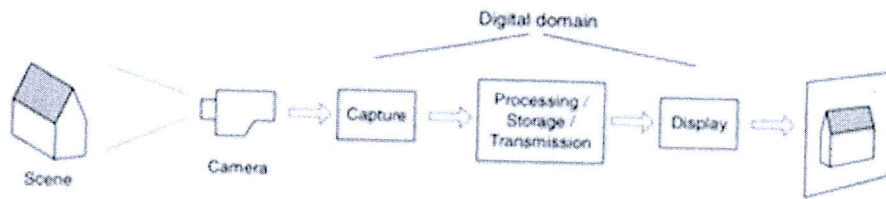


Figure 1.12 : Système de vidéo numérique : capture, traitement et affichage

## I.6 Structure d'un fichier vidéo

Un fichier vidéo est composé de plusieurs parties :

- **Le flux vidéo**, cet élément ne contient que les informations nécessaires pour afficher la vidéo. Le type de flux est déterminé par le codec utilisé (COdage/DECodage), il est aussi défini par le nombre d'images par secondes (frame per second-FPS). En fonction de l'utilisation voulue, cette donnée est différente.
- **Le flux audio**, cet élément contient la ou les pistes audio incluses dans le fichier vidéo à lire, il est possible en fonction des spécifications du fichier de contenir plusieurs pistes, comme sur un DVD par exemple.
- **Les fichiers de sous-titres**, il s'agit en général de fichiers texte contenant les données spatio-temporelles permettant de synchroniser les textes avec le film.
- **Le conteneur**, ou (container) il s'agit d'une (capsule) qui va contenir en son sein le flux vidéo et le flux audio ainsi que les fichiers de sous-titres ; ainsi que d'autres données comme une description des flux que contient le conteneur, des métadonnées (auteur, date, est.), des chapitrages. Ce conteneur a pour principale utilité de faciliter le transport et la lecture d'un fichier. La lecture va ensuite se charger de séparer les flux pour les lire.

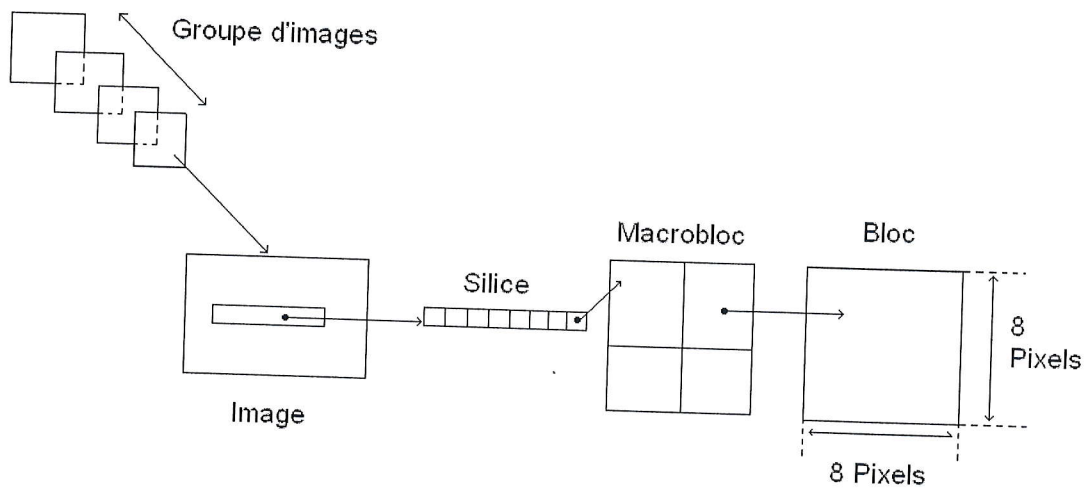
## I.7 Les données vidéo

Dans le flux vidéo, les données sont hiérarchisées selon une manière bien précise. Tout d'abord, la séquence vidéo commence par un code de début de séquence, contient un ou plusieurs groupes d'images, et se termine par un code de fin de séquence [5].

Une image est constituée de trois matrices où chaque élément de la matrice représente un pixel. A noter que les matrices U et V sont de dimensions plus petites que la matrice Y suivant le format utilisé. L'essentiel de l'information de l'image est stocké dans la matrice Y.



En effet, l'œil est plus sensible à la luminance que la chrominance. Plusieurs formats d'images existent et on les note selon les caractéristiques de leurs matrices YUV.



**Figure I.13 :** Hiérarchie des données dans le flux vidéo.

Chaque image est découpée en tranche (*slice*). Une tranche est constituée d'un ou plusieurs macroblocs (MBs) adjacents, ordonnés de gauche à droite. Ce sont des éléments importants pour la gestion des erreurs. Si une tranche contient une erreur, le décodeur passe alors immédiatement à la tranche suivante. Plus il y a de tranches dans une image, meilleur est le traitement des erreurs mais plus l'image prend de place également. Enfin, un MB est une matrice constituée de 4 blocs, chacun de taille 8 x 8 pixels. Un MB contient donc 16 x 16 pixels dans l'espace de luminance et 8 x 8 pixels dans l'espace de chrominance si on utilise le format d'image 4 : 2 : 0 .

## I.8 Les formats de la vidéo numérique

Le standard de compression vidéo peut compresser une grande variété de formats colorimétriques sous diverses résolutions. En pratique, CIF (Common Intermediate Format) est le format de base pour tout un ensemble de format vidéo. Ces formats correspondent aux tailles des images composant la vidéo [9] et sont résumés dans le tableau I.1. Le choix du format dépend de l'application et de la capacité de stockage ou de la capacité de transmission.

Le tableau I.1 donne aussi le nombre de bits requis pour représenter une image (frame) non compressée pour chaque format en supposant que l'échantillonnage est le 4 : 2 : 0 et que 8 bits sont utilisés pour la luminance et la chrominance .

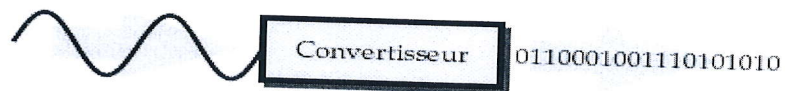
Format	Résolution luminance (hori x vert)	Résolution chrominance (hori x vert)	Bit par image	Application
S-QCIF	128 x 96	64x48	147.456	Vidéo mobile
QCIF	176 x144	88x72	304.128	Vidéo Conférence
CIF	352x288	176x144	1.216.512	Monitoring vidéo
4CIF	704x576	352x288	4.866.048	TV, DVD

**Table I.1** : formats vidéo numérique.

### I.9 La numérisation

Numériser un signal analogique consiste à transformer les grandeurs continues dans le temps en des grandeurs discontinues qui varient par palier en prenant des valeurs à intervalle de temps régulier : période d'échantillonnage  $T_e$ .

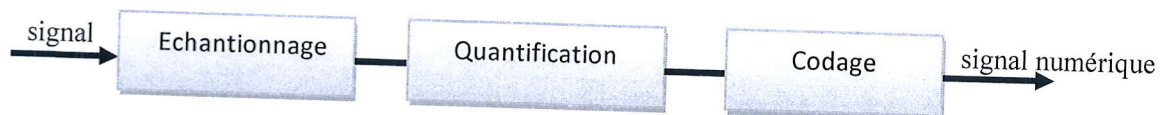
La numérisation est faite à l'aide d'un convertisseur analogique-numérique (CAN)



**Figure I.14** : convertisseur analogique-numérique (CAN).

La numérisation d'un signal nécessite trois étapes :

- Echantillonnage
- Quantification des échantillons obtenus
- Le codage des échantillons quantifiés.



**Figure I.15** : schéma de numérisation.

- **Échantillonnage** : On appelle période d'échantillonnage  $T_e$  (en secondes s), le temps entre deux mesure successives.

La fréquence d'échantillonnage  $f_e$ , correspond au nombre de mesures effectuées par

seconde. On a :  $f_e = \frac{1}{T_e} \dots 10$

- **Quantification** : Un signal numérique ne peut prendre que certaines valeurs : c'est la quantification. Elle s'exprime en bits. Cette quantification est assurée par un convertisseur (CAN). Chaque valeur est arrondie à la valeur permise la plus proche par défaut.
- **Codage** : On appelle codage la transformation des différentes valeurs quantifiées en langage binaire.

### I.10 Conclusion

Dans ce premier chapitre nous avons présentée les principes de base de la vidéo numérique et espace de représentation de couleur. Leur système et leur structure, ainsi que les données qui lui compose. À la fin de ce chapitre en parle sur la numérisation, l'effet de passage d'un signale analogique ver un signale numérique (échantillonnage et quantification).



*Chapitre II : Traitement effectué  
sur la vidéo*

## II.1 Introduction

Une vidéo contient une quantité importante de données, et malgré l'augmentation de la puissance des processeurs et des capacités des périphériques de stockage, nous avons besoin de la représenter dans un format plus précis. C'est la raison d'apparition de la compression vidéo. Dans ce chapitre, Nous définissons des concepts principaux de compression vidéo, ainsi que les normes de compression MPEG2, H.264, Erreur de transmission par IP et Erreur de transmission par wifi.

Les différents traitements effectués sur une vidéo conduisent des pertes des informations et des distorsions, ce qui minimise la qualité de la vidéo. Dans la dernière partie de ce chapitre nous parlerons sur les distorsions liées à la compression vidéo.

## II.2 Compression

**La compression vidéo** est une méthode de compression de données, qui consiste à réduire la quantité de données, en minimisant l'impact sur la qualité visuelle de la vidéo. L'intérêt de la compression vidéo est de réduire les coûts de stockage et de transmission des fichiers vidéo, les deux méthodes de compression **sans perte** et **avec perte** sont utilisées pour la compression des séquences vidéo.

**La compression sans perte** Elle garantit que le fichier compressé puis décompressé est identique à l'original. Ce type de compression est obligatoire pour les fichiers binaires et les fichiers texte.

**La compression avec perte** Elle admet la perte des données considérées redondantes ou inutiles en échange d'un meilleur taux de compression. Un fichier compressé puis décompressé selon cette méthode sera différent de l'original. Les images, les sons et la vidéo utilisent la compression avec perte. Ceux sont par exemple les formats JPEG pour les images ou MPEG pour le son et/ou la vidéo.

La compression des données vidéo numériques n'est possible que si les redondances qui existent dans une vidéo sont éliminées et il en existe essentiellement trois types [10]:

- La redondance spatiale : corrélation entre les pixels voisins.
- La redondance temporelle : corrélation entre les images successives de la vidéo.
- La redondance psycho-visuelle : propriétés du système visuel humain.

## II.2.1 Types de redondance

### II.2.1.1 Redondance spatiale

C'est la corrélation statistique entre les pixels d'une image appelée redondance intra frame. Chaque image, prise indépendamment des autres, présente des zones uniformes plus ou moins grandes dans les quelles les valeurs des intensités des pixels sont très voisines.

On peut diminuer cette redondance en codant chaque image séparément. Ce type de codage est appelé codage INTRA. L'utilité de ce codage est de pouvoir accéder de façon aléatoire à chaque image individuellement.

### II.2.1. 2 Redondance temporelle

Cette redondance est une corrélation statistique entre les pixels d'images dans une séquence vidéo. L'intervalle de temps entre la prise d'images est très court et donc la similarité sera forte. Deux images qui se suivent dans une séquence vidéo sont quasiment identiques sauf dans le cas d'un changement de plan. Le but est alors de ne stocker que la différence, c.à.d. ce qui a été modifié entre deux images successives. Ce type de codage est appelé codage INTER.

### II.2.1. 3 Redondance psycho-visuelle

En plus d'éliminer les redondances temporelles et spatiales, la redondance psycho-visuelle est généralement réduite elle aussi. La mesure la plus significative est une résolution réduite du détail des couleurs par rapport au détail de la luminance qui permet un meilleur rapprochement avec les caractéristiques de la perception humaine. C'est dans ce but que la première étape d'une compression vidéo est le changement d'espace couleur (RVB vers YUV ou YCbCr). Les codeurs vidéo utilisent différents taux de sous-échantillonnage de la couleur selon les besoins de la compression. Par exemple : Les algorithmes de compression vidéo pour le Web et les DVD utilisent le schéma 4:2:0 .

## II.2.2 Définition du codeur

Un codeur ou CODEC (COmpression/DECompression) vidéo est un outil pouvant être logiciel (software), matériel ou une combinaison des deux permettant la compression et/ou la décompression des vidéos numériques. Les codeurs vidéo utilisent, en général, des formats de compression standards par exemple: H.264, MPEG-2 et MPEG-4 : codecs dérivés des standards MPEGs. DivX , Xvid , DV , Apple Pro Res, MJPEG et JPEG 2000.



### II.2.3 Transformation en cosinus Discrète

La Transformation en Cosinus Discrète (DCT) permet de transformer un bloc d'une composante en un ensemble de fréquences décrivant le même ensemble (c'est un changement de représentation isomorphe). Qui sera ordonné dans une matrice de coefficients afin que la valeur moyenne (coefficient DC) soit le premier en haut à gauche. Les fréquences verticales sont représentées par les coefficients sur les lignes de la matrice, et les fréquences horizontales sont représentées par les coefficients sur les colonnes de la matrice.

La définition formelle de la DCT à deux dimensions est la suivante :

$F(u,v)$  est la transformée,  $c$ 'est la fonction qui donne la valeur pour le couple de fréquence  $(u,v)$ .

$$F(u, v) = \frac{2}{N} * C(u) * C(v) * \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x, y) \cos\left(\frac{(2x+1)u\pi}{2N}\right) \cos\left(\frac{(2y+1)v\pi}{2N}\right) \dots 11$$

Avec :  $u, v, x, y = 0, 1, 2, \dots, N - 1$

Ou  $x$  et  $y$  sont les coordonnées spatiales

Ou  $u$  et  $v$  sont les coordonnées dans la transformée

Et avec :

$$C(u), C(v) = \begin{cases} \frac{1}{\sqrt{2}} & \text{pour } u, v = 0 \\ 1 & \text{sino} \end{cases}$$

Et la transformation inverse, l'IDCT est défini ainsi :

$$f(x, y) = \frac{2}{N} \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} C(u)C(v)F(u, v) \cos\left(\frac{(2x+1)u\pi}{2N}\right) \cos\left(\frac{(2y+1)v\pi}{2N}\right) \dots 12$$

En pratique dans les encodages MPEG, la DCT est utilisée sur des blocs. Le bloc est transformé dans le domaine fréquentiel par la DCT. Au moment du décodage on applique ce que l'on appelle l'IDCT (Inverse Direct Cosine Transform) qui permet de repasser au domaine spatial [11].

### II.2.4 Schéma de base de compression

Quel que soit le type de codeur (voix, image fixe, vidéo), un certain nombre d'outils communs permettent de réaliser la compression du signal. Ces outils, brièvement décrits ci-dessous [12]:

- **La prédiction** : Ce mécanisme permet d'établir un ensemble de valeurs prédites sur les échantillons du signal d'entrée. Généralement, cette estimation est basée sur les informations des échantillons déjà encodés du signal. La différence entre les échantillons du signal original et leur prédiction permet d'obtenir les résidus de prédiction.

- **La transformation** : Ce procédé consiste à projeter les résidus de prédiction dans une base permettant de réduire la corrélation statistique entre les échantillons des résidus de prédiction. Cette opération permet de rassembler l'énergie du signal sur un faible nombre d'échantillons. Dans cette catégorie, la transformée en cosinus discrète TCD (Discrete Cosine Transform ou DCT) est la plus utilisée pour sa simplicité algorithmique et ses performances.
- **La quantification** : Ce procédé permet d'approximer les échantillons du signal transformé afin de réduire la quantité d'information à transmettre. Parmi tous les outils de compression, la quantification est la seule opération irréversible, induisant des pertes d'information.
- **Le codage entropique** : Ce mécanisme permet de représenter efficacement les symboles issus du quantificateur en prenant en compte leur fréquence d'apparition. Ainsi, un mot de code de longueur variable CLV (Variable Length Coding ou VLC) est associé à chaque symbole en fonction de la statistique de la source. Les mots de code les plus courts sont attribués aux symboles fréquents, et inversement, des mots de code plus longs sont réservés aux symboles les moins probables. Les codeurs utilisent souvent ce procédé pour encoder les paramètres de compression, les plus connus étant le codage de Huffman et le codage arithmétique.

Trois modes de codage sont possibles et conduisent à distinguer trois types d'image :

- Les images **INTRA** (key frames), dites de **type I** : Ces images sont codées intégralement, sans aucune référence aux images voisines de la séquence vidéo. C'est la redondance spatiale qui est exploitée et éliminée à l'aide de la TCD. Cette opération est réalisée après changement du plan de couleur de l'image du RVB vers YCbCr ou YUV. Chacun de ces plans est décomposé en blocs de 8x8 pixels et transformés en matrices de coefficients fréquentiels classés selon l'amplitude des signaux. Ces coefficients sont ensuite quantifiés et codés en utilisant un codage entropique comme montre la figure II.1. L'inconvénient des images I est qu'elles consomment beaucoup plus de bits, mais, d'un autre côté, elles ne génèrent pas beaucoup d'artefacts.

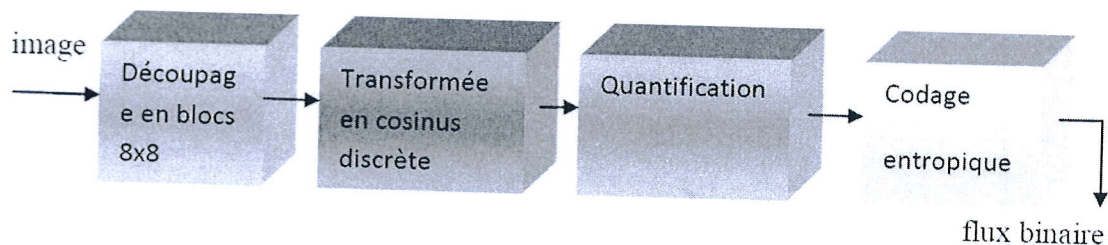


Figure II.1 : Schéma du codage INTRA.

- Les images **INTER** prédites ou à compensation de mouvement, dites de **type P** : Les images « Prédicatives » exploitent à la fois la redondance spatiale et la redondance temporelle des images



d'une séquence vidéo. Elles sont codées à partir de l'image « I » ou « P » précédente à l'aide de vecteurs de mouvement obtenus par estimation de mouvement. Les images sont découpées en blocs de 16 x 16 pixels. Les vecteurs de mouvement sont ensuite calculés en fonction du déplacement de chacun de ces blocs de pixels d'une image à la suivante puis codés avec un Codage à Longueur Variable (CLV).

- Les images **INTER** prédites bidirectionnelles, dites de **type B** : qui fait référence à une image antérieure ainsi qu'à une image future comme montrées dans la figure II.2. Elles sont codées avec une estimation du mouvement et elles ne sont pas utilisées par tous les codeurs. Les images de type I et P servent de référence aux images P et B, les images de type B ne servent jamais de référence. Le codage d'une image P et B est illustré par la figure II.3.

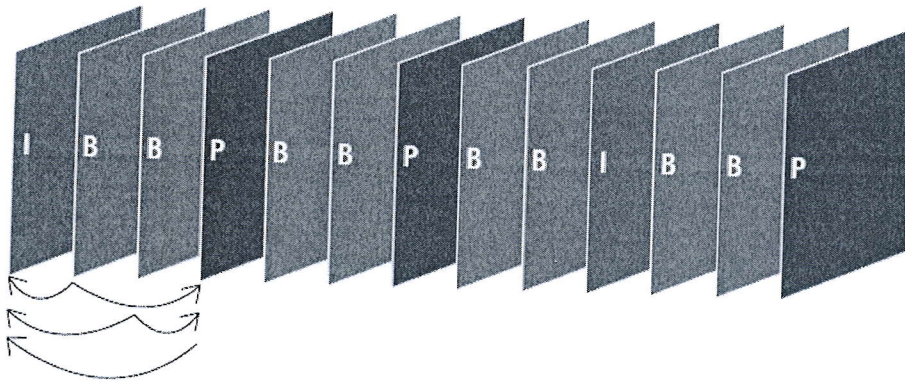


Figure II.2 : Séquence type avec des images I, B et P.

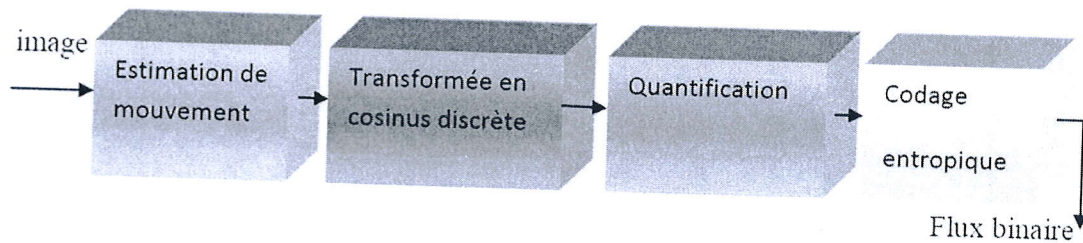


Figure II.3 : Schéma du codage INTER prédiction.

### II.3 Norme de compression



### II.3.1 MPEG2

**MPEG** (Moving Picture Experts Group) est le groupe de travail de l'ISO (International Organization for Standardization) et de la CEI (Commission Electrotechnique Internationale) pour les technologies de l'information. Ce groupe d'experts est chargé du développement de normes internationales pour la compression, la décompression, le traitement et le codage de la vidéo, de l'audio et de leur combinaison. L'activité de ce groupe de travail a commencé en 1988.

#### II.3.1.1 Définition

**MPEG 2:** cette norme est enregistrée à l'ISO sous le code ISO/IEC 13818, et a été finalisée en 1996. Elle a été développée pour la compression de la vidéo de qualité télévision à un débit de 4 à 6 Mbits/s. Un peu plus tard, MPEG-2 intégra la compression de la télévision à haute définition (TVHD). Aujourd'hui, MPEG-2 est aussi le format utilisé pour stocker les films sur DVD. Le standard permet donc désormais la compression progressive ou entrelacée de vidéo selon des débits allant de 1.5Mb/s à 100Mb/s [13].

#### II.3.1.2 Schéma de l'encodeur MPEG2

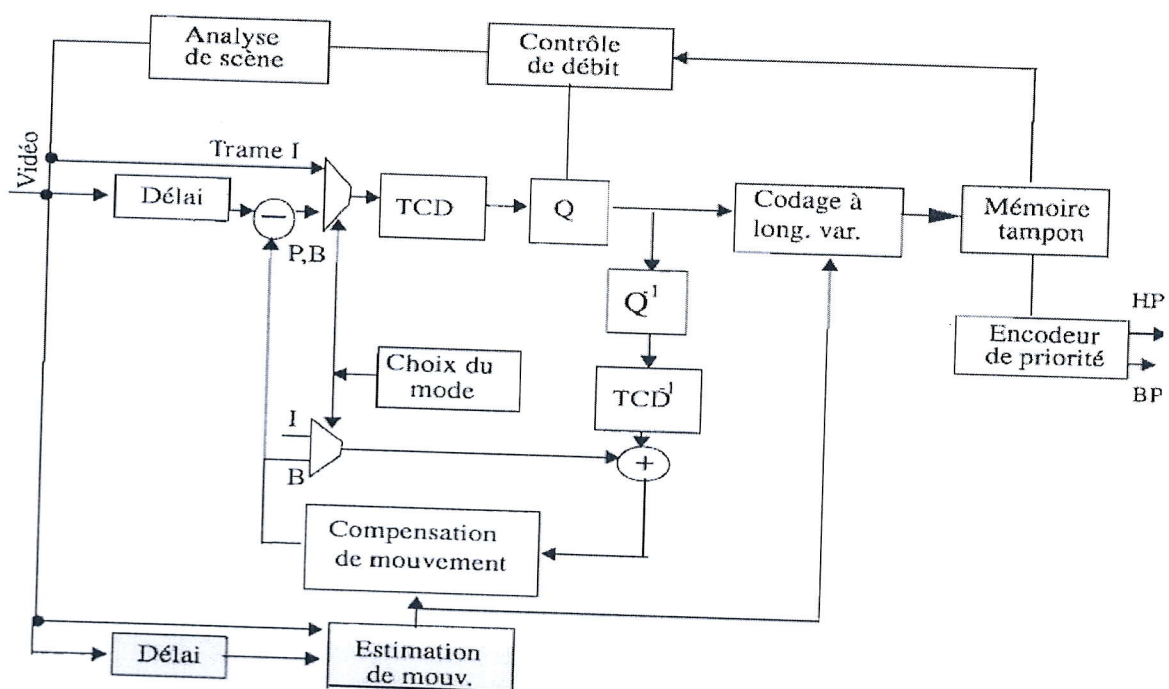


Figure II.4 : Encodeur MPEG-2.

Au niveau du codeur la DCT est appliquée à chaque bloc de luminance et de chrominance de  $8 \times 8$  et, après la sortie de la TCD, chaque des 64 coefficients DCT est quantifié de façon uniforme (Q).

Après la transformation et la quantification, le bloc de fréquences quantifiées peut contenir beaucoup de coefficients nuls (zéros). Pour obtenir une compression de donnée supplémentaire, un codage arithmétique est utilisé. Le nombre de coefficients nuls consécutifs est augmenté par un balayage des blocs de 8 X 8 pixels des basses fréquences vers les hautes fréquences.

À la suite du codage des zéros, on effectue un codage à longueur variable, c'est le codage de Huffman qui est le plus couramment utilisé. Les données encodées sont par la suite transmises ou stockées [14].

### II.3.1.2 Schéma de décodeur MPEG-2

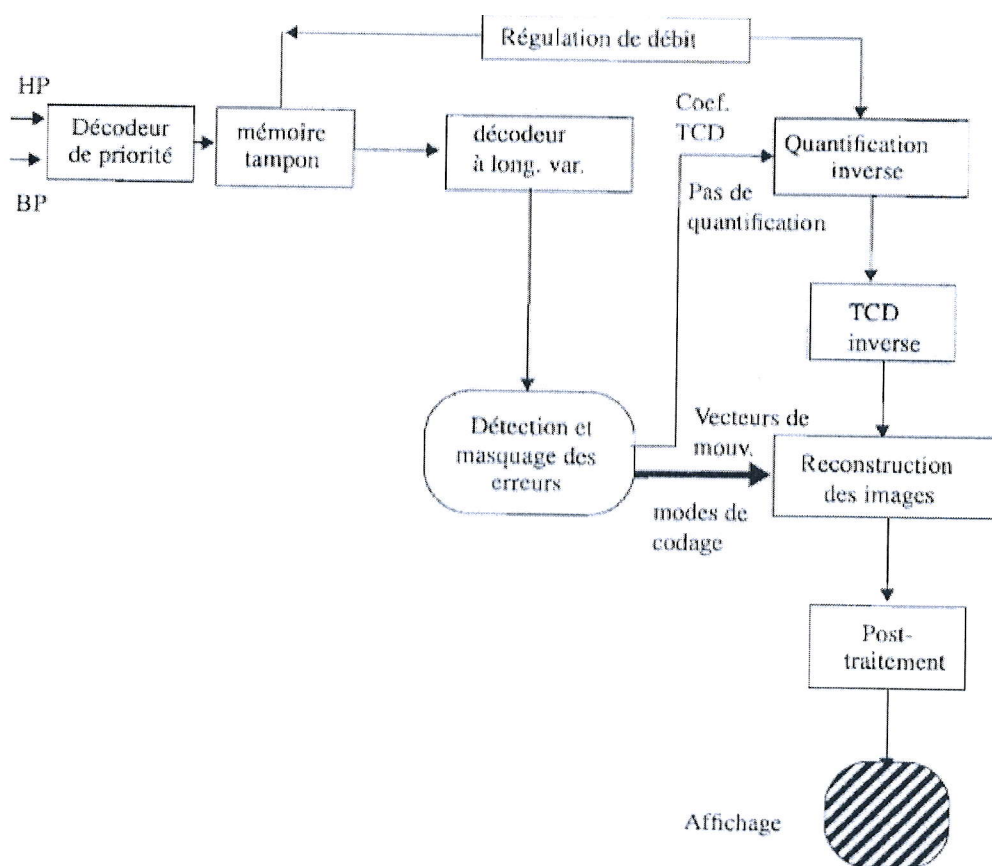


Figure II.5 : Décodeur MPEG-2.

Le décodeur [14] reçoit une séquence MPEG-2 comme entrée et effectue les opérations inverses pour produire des fichiers contenant des composantes Y, U et V. Le nombre de fichiers produits dépend de la fréquence d'affichage (25 ou 30 images/s) et de la longueur de la séquence d'images encodées (par exemple une séquence vidéo de 5 secondes contient  $30 \times 3 \times 5 = 450$  fichiers YUV). La figure II.5 illustre le schéma bloc du décodeur MPEG-2. Lorsqu'un seul canal est disponible et qu'on ne fait pas usage du codage pyramidal ou du partitionnement des données, seul le canal de basse priorité (BP) est utilisé.

## II.3.1.4 Caractéristiques de la norme MPEG2

Prédiction	<b>Prédiction Intra</b>
	<ul style="list-style-type: none"> <li>• Le codage intra-image génère les trames I.</li> <li>• Le codage intra-image consiste le codage de chaque image indépendamment des autres.</li> </ul>
	<b>Prédiction inter</b>
	<ul style="list-style-type: none"> <li>• Le codage inter-image génère les trames P et B.</li> <li>• Le codage inter-image à coder les différences existant entre des trames successives .La trame Pest prédite à partir de la trame I. La trame B est une interpolation de la trame I et la trame P.</li> <li>• L'interpolation se fait en calculant la moyenne entre deux blocs. Ceci a l'avantage de réduire le bruit dans l'image de moitié. Toutefois, l'interpolation a certains désavantages: notamment, le délai, la complexité et l'espace mémoire nécessaire pour garder les deux trames requises pour coder la trame B.</li> <li>• Le codage inter-image utilise une estimation de mouvement entre deux images consécutives. La différence entre les images s'effectue alors entre un bloc pixel et le bloc correspondant dans l'image suivante déplacée du vecteur de mouvement.</li> </ul>
transformation	<ul style="list-style-type: none"> <li>• Le MPEG2 Utilise la transformée en cosinus discrète réelle (<b>TCD</b>) (8* 8) pixel.</li> <li>• Si la transformation inverse était appliquée après la <b>TCD</b>, l'image originale serait parfaitement reconstituée.</li> </ul>
Codage entropique	<ul style="list-style-type: none"> <li>• Le MPEG2 Utilise le codage <b>Huffman</b> et le codage <b>arithmétique</b> qui consiste à coder une suite de coefficients nuls par un seul coefficient nul suivi du code indiquant le nombre de répétitions.</li> </ul>
Transmission	<ul style="list-style-type: none"> <li>• Les données encodées sont par la suite transmises ou stockée la <b>Figure II.6</b> illustre un exemple de la structure des différents types de trames codées et l'ordre de transmission par rapport à l'ordre de réception.</li> <li>• Si les trames de type B ne sont pas utilisées (ce qui résulte en un délai de codage plus court et une complexité réduite), on a une séquence semblable à celle-ci: 1 PPP 1 PPP 1 PPP 1.</li> </ul>



Table II.1 : Caractéristiques de la norme MPEG2.

Entrée de l'encodeur:	I B B P B B P B B I B B P
Sortie de l'encodeur: Séquence codée et transmise:	P B B P B B I B B P B B
Sortie du décodeur:	I B B P B B P B B I B B P

Figure II.6 : Exemple d'ordre de transmission et de réception des trames I, P et B.

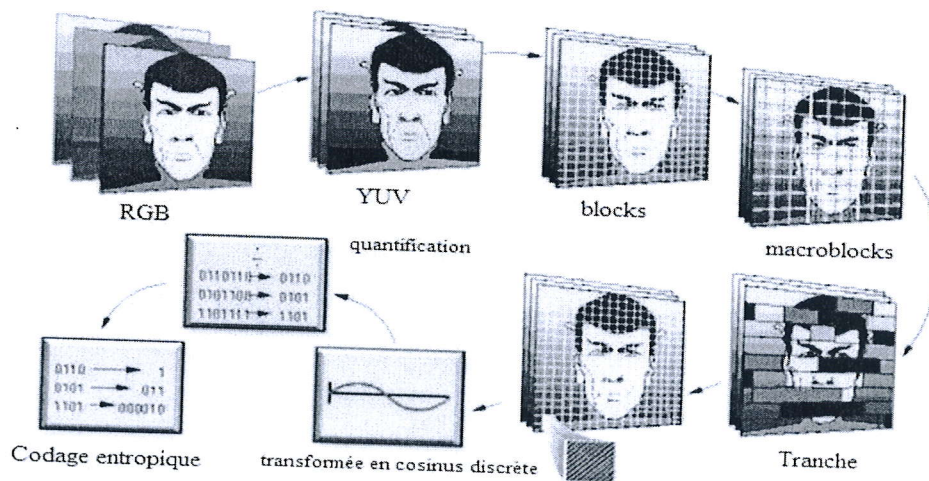


Figure II.7 : Exemple de codeur MPEG2.

### II.3. 2 H.264

#### II.3.2.1 Définition

Développé par le groupe JVT en 2003, le H.264 appelé aussi MPEG-4 Part 10 ou AVC (Advanced Vidéo Coding) est un standard très performant pour la compression vidéo. En comparaison avec le MPEG-2, le H.264 délivre des vidéos de haute définition avec seulement la moitié du taux binaire et même le quart [15].

#### II.3.2.2 Schéma de l'encodeur H.264 /AVC

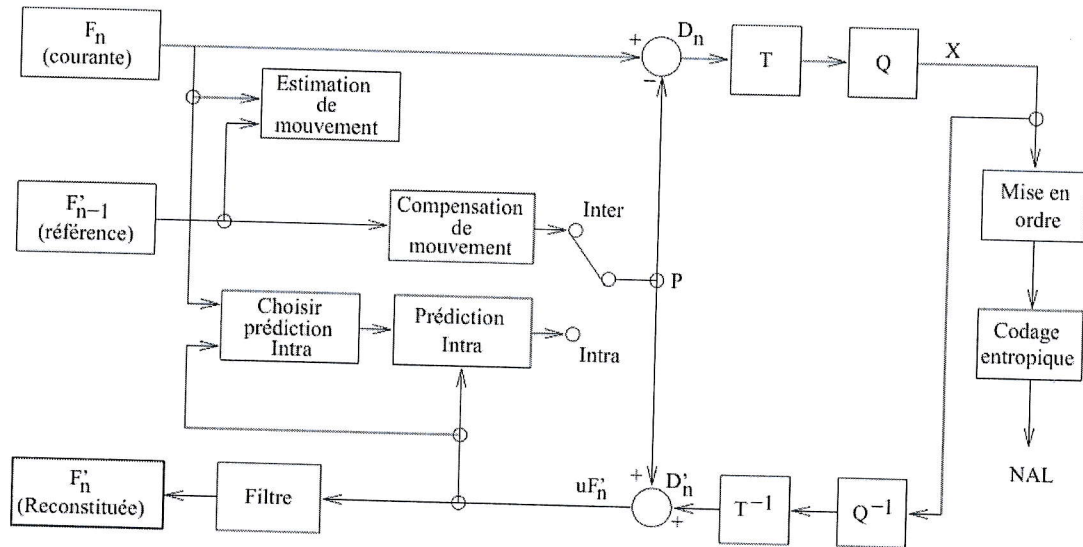


Figure II. 8 : Schéma bloc de d'encodeur H.264.

L'image courante  $F_n$  à coder est présentée à l'entrée de l'encodeur [16]. Elle est traitée par unités de macrobloc (MB), où chaque MB peut être codé en mode inter ou intra selon la nature des redondances. Si les redondances appartiennent à la même image comme c'est le cas de la figure II.9, la prédiction spatiale (Intra) est la mieux adaptée. Dans le cas d'une redondance temporelle, c'est-à-dire une ressemblance entre des images successives comme sur la figure II.10, la prédiction Inter est utilisée.

Le bloc prédit P est ensuite soustrait du bloc à coder et la différence résulte en un bloc résiduel  $D_n$ . Une transformation T est appliquée aux échantillons du bloc résiduel, et les coefficients de la transformation subissent une quantification Q. Les coefficients X seront réordonnés, puis codés par un codage entropique. Le flux binaire résultant passe alors dans la couche NAL, où il sera organisé en paquets, adaptés pour le stockage ou la transmission. Pour alimenter le processus de prédiction.

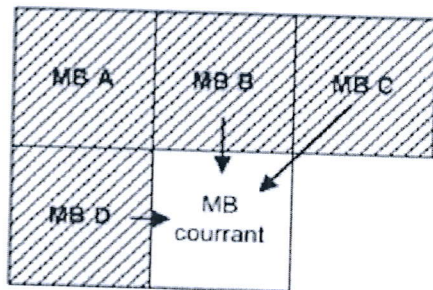


Figure II.9 : Prédiction Intra.

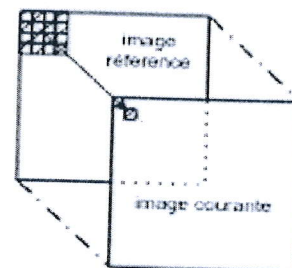


Figure II.10 : Prédiction Inter.

II.3.2.3 Schéma de décodeur H.264/AVC

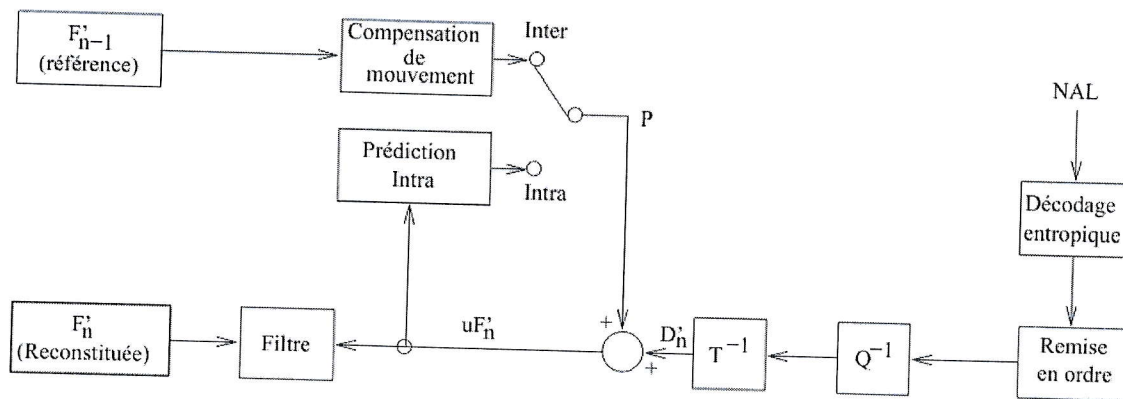


Figure II.11 : Structure du décodeur H.264/AVC.

Le décodeur reçoit en entrée le flux binaire codé venant de la couche NAL, puis applique la quantification inverse  $Q^{-1}$ , suivie par transformation inverse  $T^{-1}$ , pour reconstituer le bloc d'information résiduelle. En utilisant l'en-tête de flux binaire, qui contient les paramètres de codage (le bloc origine, le bloc résidu et le mode de prédiction), le décodeur forme aussi le bloc prédit P, qui a été utilisé pendant l'encodage. Le bloc prédit est alors ajouté au bloc résiduel pour former le bloc courant à décoder. Le bloc décodé est ensuite filtré et envoyé pour l'affichage [16].

### II.3.2.4 Caractéristiques de la norme H.264/AVC

Prédiction	Prédiction Intra
	<p>La prédiction Intra utilise deux types de prédiction pour un Macrobloc de luminance selon son homogénéité (16*16 et 4*4) et un troisième type pour la chrominance (8*8).</p> <ul style="list-style-type: none"> <li>• <b>Prédiction Intra 16 * 16</b> : Il est appliqué à un Macrobloc entier fait recourt à 4 modes obéissant chacun à une direction et à des équations caractéristiques (voir figure II.12).</li> <li>• <b>Prédiction Intra 4 * 4</b> : Si la prédiction Intra 4* 4 a été sélectionnée, un Macrobloc a 16 choix de direction de prédiction (voir figure II.13).</li> <li>• <b>Prédiction Intra 8 * 8</b> : Les 4 modes de prédiction sont similaires aux modes de prédiction de la luminance 16 * 16. La seule différence provient de la taille des blocs manipulés qui est alors de 8* 8 pixels.</li> <li>• La valeur des données résiduelles est généralement faible (Prédiction Intra et</li> </ul>



	<p>Inter). Cette propriété des résidus est exploitée afin d'atteindre des taux de compression élevés.</p>
	<p><b>Prédiction inter</b></p>
	<p>Le modèle de prédiction inter est fondé sur la recherche et la compensation de mouvement dans le but d'accomplir l'appariement de blocs entre deux trames.</p> <ul style="list-style-type: none"> <li>• <b>Partitionnement des macroblocs</b> : Un macrobloc contient <math>16 \times 16</math> valeurs de luminances qui peuvent être décomposées de quatre manières distinctes : une partition <math>16 \times 16</math>, deux partitions <math>16 \times 8</math>, deux partitions <math>8 \times 16</math> et quatre partitions <math>8 \times 8</math>. De plus, chacun des sous-macrobllocs <math>8 \times 8</math> peut également être décomposé en : deux partitions <math>8 \times 4</math>, deux partitions <math>4 \times 8</math>, et quatre partitions <math>4 \times 4</math>. Le tout est résumé à la figure II.15</li> <li>• <b>Compensation de mouvement</b> : Cette technique divise une image en série de macro-blocs. Il est possible de composer ou de « prédire » une nouvelle image (une image P, par exemple) bloc par bloc en recherchant un bloc identique dans une image de référence, on utilise la méthode de block matching. Si un bloc référent est trouvé, l'encodeur code tout simplement l'emplacement dans l'image de référence où se situe le bloc identique. Le codage du vecteur de mouvement utilise moins de bits que le codage du contenu réel d'un bloc (voir figure II.16).</li> </ul>
<p><b>Prédiction multi-images</b></p>	<ul style="list-style-type: none"> <li>• Le H.264 offre la possibilité de ne pas utiliser une seule référence mais plusieurs. Ainsi la prédiction de mouvement peut utiliser jusqu'à 16 différentes images de références selon le profil utilisé. Cette possibilité est illustrée sur la figure II.17.</li> </ul>
<p><b>transformation</b></p>	<ul style="list-style-type: none"> <li>• Le standard <b>H.264</b> utilise la transformée en cosinus discrète entière (DCT) ayant les mêmes propriétés que la DCT classique.</li> <li>• L'avantage majeur de cette version de la DCT est la génération de coefficients entiers et non décimaux.</li> <li>• Possibilité d'éviter les problèmes des implémentations liés à la partie fractionnaire et de réduire le débit.</li> <li>• Deux tailles de sous fenêtrés <math>4 \times 4</math> et <math>8 \times 8</math> sont utilisées dans le standard H.264 pour le calcul de cette transformation par DCT, ce qui permet de limiter la complexité de calcul.</li> </ul>

<p><b>codage entropique</b></p>	<p>Le standard H.264 utilise deux modes de codage adaptatif au contexte:</p> <ul style="list-style-type: none"> <li>• Le codage <b>VLC</b> : (CAVLC Context Adaptive Variable Length Coding) sert à coder les coefficients quantifiés issus de la transformation mathématique. Pour ce codage des tables de codage sont utilisées pour chaque élément syntaxique et l'utilisation de ces tables varie auprès les valeurs déjà décodées.</li> <li>• Le codage <b>CABAC</b> : (Context Adaptive Binary Arithmetic Coding) obtient la bonne performance de compression grâce à, premièrement, la sélection modèle de probabilité pour chaque élément syntaxique en fonction du contexte de l'élément, deuxièmement, l'adaptation probabilité estime en basant sur des statistiques locales et troisièmement, l'utilisation de codage arithmétique se place au codage à longueur variable.</li> <li>• une méthode utilise pour coder la plupart des éléments de synchronisation comme les entêtes : (UVLC-Universal Variable Length Coding) une table universelle de mots de code.</li> </ul>
<p><b>Filtre Anti-Bloc</b></p>	<ul style="list-style-type: none"> <li>• le H.264 est appliquée le filtrage anti effet de bloc ou <b>anti-bloc</b> à tous les macroblochs avant la reconstruction de l'image.</li> <li>• ce processus de filtrage consiste à lisser les bords horizontaux et verticaux de tous les macroblochs en utilisant des techniques de filtrage adaptatif.</li> <li>• la qualité d'image reconstruite objective et subjective est améliorée (voir figure II.18).</li> </ul>

Table II.2 : Caractéristiques de la norme H.264.

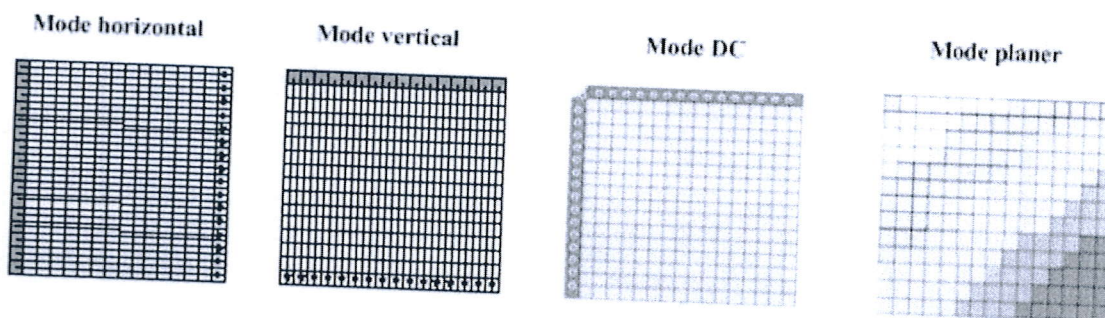


Figure II.12 : Mode d'Intra prédiction 16\*16.



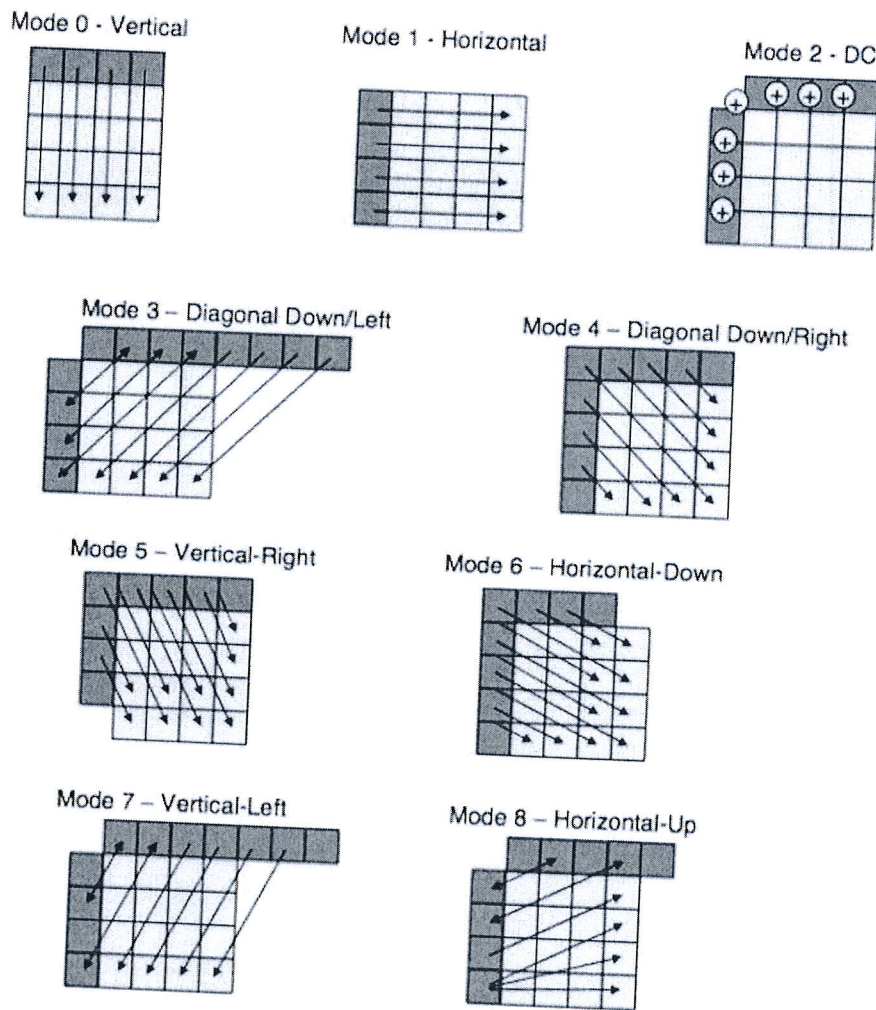


Figure II.13 : Mode d'Intra prédiction 4\*4.



Image source d'origine



Image intra-prédite





Image résiduelle

Image finale

Figure II.14 : Exemple d'intra-prédiction du H.264.

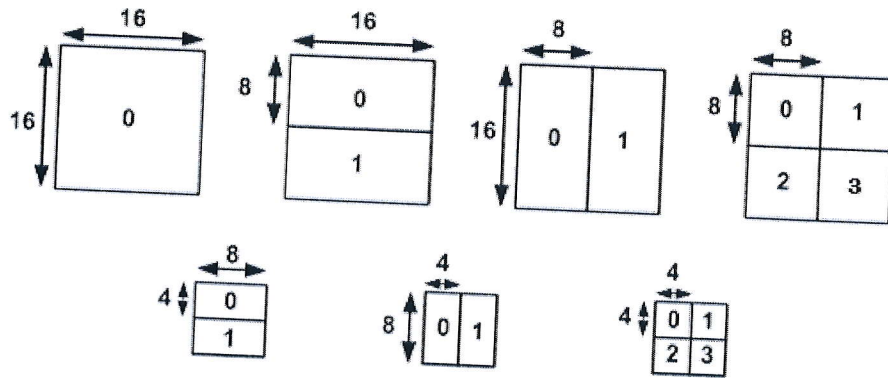


Figure II.15: Partitions de macroblocs et de sous macroblocs.

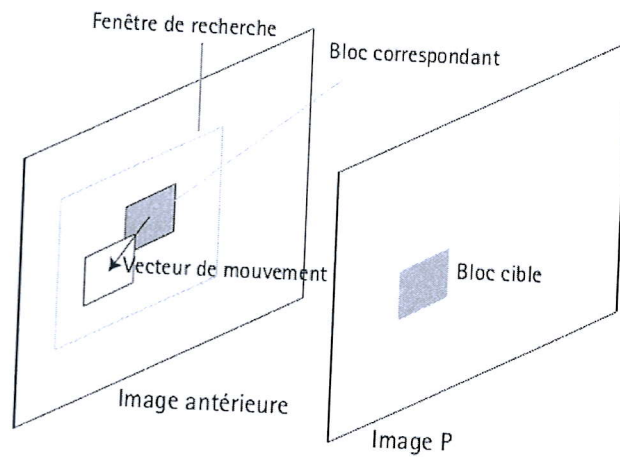


Figure II.16 : Exemple de la compensation de mouvement basée sur les blocs.

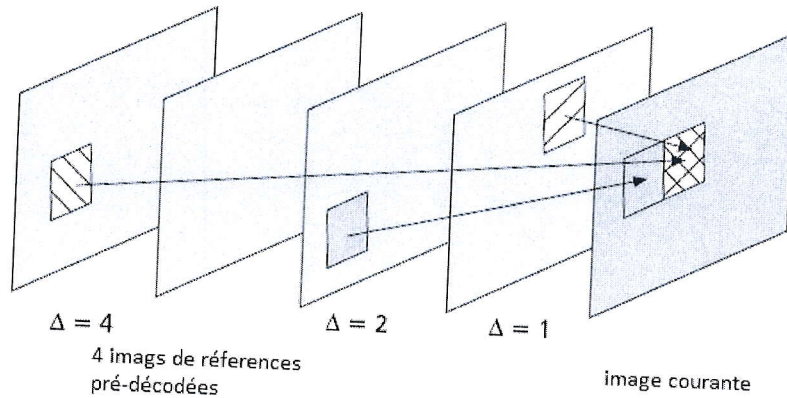


Figure II.17 : Exemples de prédiction multi-références.



Figure II .18 : image compressée.



Figure II.19 : image filtrée.

### II.3. 3 Codeur Erreur de transmission IP

#### II.3.3.1 Protocole IP

IP signifie "Internet Protocol", Protocole Internet [17]. Il représente le protocole réseau le plus répandu. Il permet de découper l'information à transmettre en paquets, de les adresser, de les transporter indépendamment les uns des autres et de recomposer le message initial à l'arrivée. Il est souvent associé à un protocole de contrôle de la transmission des données appelé TCP, on parle ainsi du protocole TCP/IP. Cependant, TCP/IP est un ensemble de protocole dont les plus connus sont :

- TCP - Transmission Control Protocol - Couche 4(couche application) - Transport en mode connecté.
- UDP - User Datagram Protocol - Couche 4 - Transport en mode non connecté.

### II.3.3.2 Description générale du modèle TCP/IP

Afin de pouvoir appliquer le modèle TCP/IP (Transmission Control Protocol/Internet Protocol) [17] à n'importe quelles machines, le système de protocoles TCP/IP a été décomposé en plusieurs couches. Les données (paquets d'informations) qui circulent sur le réseau sont traitées successivement par chaque couche, qui vient rajouter un élément d'information (appelé en-tête) puis sont transmises à la couche suivante.

### II.3.3.3 Système de transmission de la vidéo par paquets

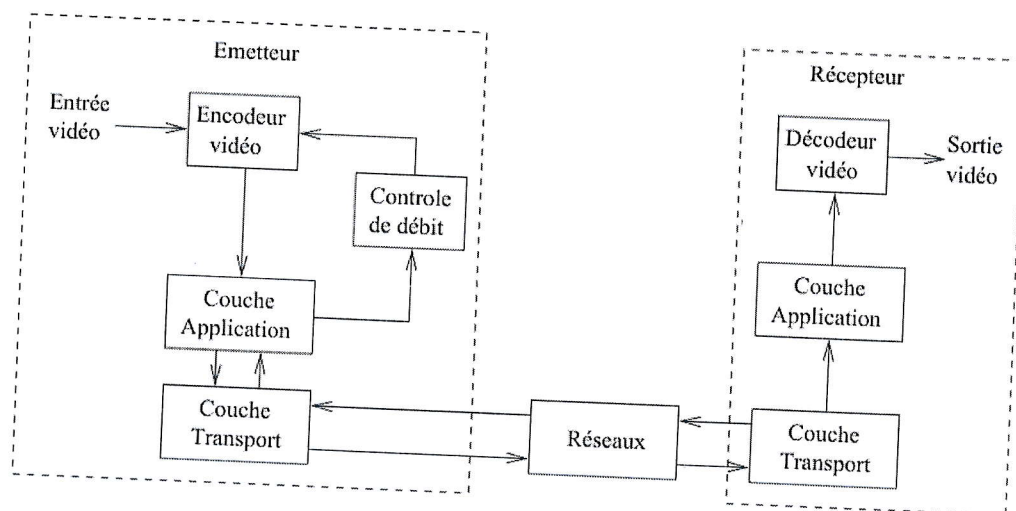


Figure II.20 : Système de transmission de la vidéo par paquets.

Le transmetteur et le récepteur sont reliés à un réseau de transport, établissant la liaison entre eux [15].

Au niveau du transmetteur, le codeur vidéo a la charge de compresser le signal vidéo, de formater les données codées en paquets et de les expédier à la couche transport, pour transmission. Le codeur vidéo doit aussi adapter le débit source à la capacité du canal de transmission. Notons que le canal se réfère, Le contrôle de débit est assuré par un module de contrôle de débit qui informe le codeur de source sur l'état du canal.

Au niveau du récepteur, les paquets reçus sont traités au niveau de chaque couche, où les en-têtes propres au protocole de communication sont ôtés et les informations utiles envoyées vers le décodeur vidéo qui se charge de décompresser la vidéo codée et de l'afficher dans l'ordre et la synchronisation requis.



### II.3.3.4 Caractéristiques de la transmission

La transmission de données sur un réseau se fait par paquets PES (packetized elementary streams - flux élémentaires découpés), de taille constante, par exemple. Un paquet de transport est composé de 188 octets, et contient 4 octets d'en-têtes et les données du flux. Comme un paquet au sens du réseau ne peut contenir les données que d'un seul PES à la fois.

#### Les composants d'un en-tête du paquet:

- Un code de début de paquet, pour faciliter la synchronisation et l'identification d'un nouveau paquet.
- Un code indiquant le début d'un nouveau PES, ou la transmission de données hors flux.
- Un numéro d'identification du flux (utilisé dans le cas de plusieurs flux multiplexés).
- Un indicateur de priorité de transmission.
- Un compteur de continuité, incrémenté à chaque fois qu'un nouveau paquet de transport est généré pour un flux donné, utilisé pour détecter la perte de paquets [18].

**Commutation de paquets :** La commutation de paquets [19] est une technologie basée sur l'informatique dans laquelle les connexions entre abonnés utilisent les mêmes lignes physiques. Lorsqu'un paquet arrive dans un commutateur, il est stocké en mémoire (en file d'attente) tant que la voie de sortie dont il a besoin est occupée ; lorsque celle-ci devient libre le paquet est envoyé vers le noeud suivant et de proche en proche, ce paquet atteint sa destination par un mécanisme de partage de lignes plus efficace.

La commutation de paquets peut conduire à la perte de paquets suite au débordement de file d'attente dans les noeuds de commutation.

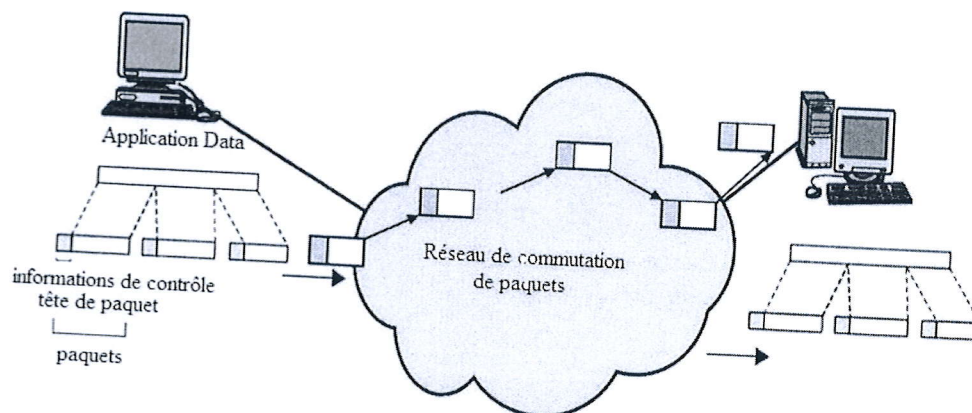
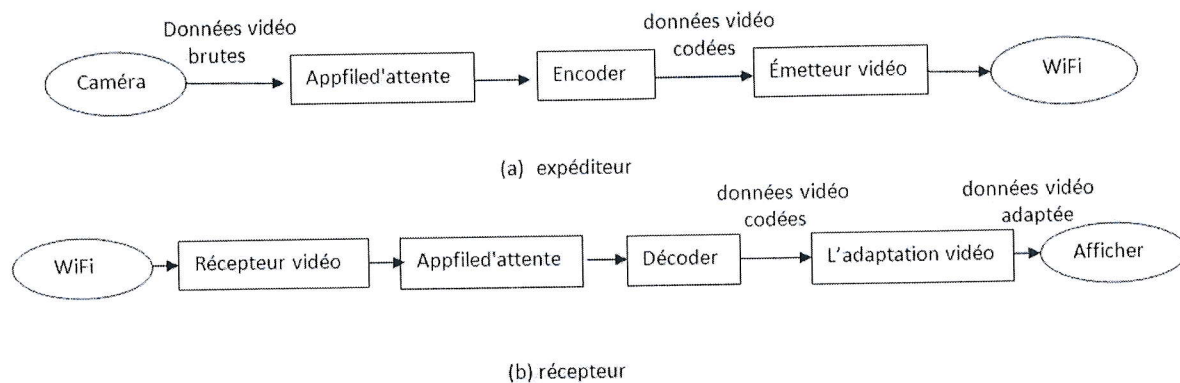


Figure II.21 : Principe de la commutation de paquets.

### II.3.4.3 Diagramme émetteur / récepteur



**Figure II.22** : Diagramme émetteur / récepteur.

Au niveau d'**émetteur** les données de la vidéo brute est extraites de la caméra / fichier et stockées dans la file d'attente (à des fins de mise en mémoire tampon). Ces données brutes sont ensuite transmises au codeur qui code les données en utilisant le codec sélectionné et une technique de codage. La trame vidéo codée est alors transmis via l'air par le module Wi-Fi au récepteur.

Au niveau de **récepteur** quand une trame vidéo codée arrive, elle est tamponnée et transmis au décodeur. Avant d'être affichées sur l'écran, les images vidéo peut nécessiter l'adaptation aux spécifications matérielles (par exemple, la résolution de l'écran). Le récepteur sera continuer à recevoir des trames jusqu'à la fin de la vidéo ou la fermeture de la connexion du client [21].

### II.3.4.4 Erreurs de transmission par un réseau sans-fil

Les principaux problèmes d'une communication sur un réseau sans-fil [22] se résument comme suit:

- la mobilité engendre une rapidité de déconnexion et de reconnexion en fonction de la distance entre le mobile et le point d'accès ainsi que les obstacles qui les séparent. En conséquence, un grand nombre de ruptures de signal sont constatées sur ces réseaux.
- le changement de chemin entre l'émetteur et le récepteur en cas de changement d'AP (Access Point) peut causer un taux élevé de paquets acheminés en désordre.
- le taux d'erreur dans les réseaux sans-fil est élevé, ce qui cause généralement des pertes aléatoires de paquets. Ce taux élevé s'explique par des interférences liées à



l'environnement ou provoquées par d'autres communications travaillant sur une même bande de fréquence ou sur des bandes de fréquence très étroites.

#### II.4 Distorsions liées à la compression vidéo numérique

Quelles que soient les performances de la technique de compression, il existera toujours des distorsions dans la vidéo reconstruites dont la qualité peut être évaluée objectivement et subjectivement. Différents types de distorsions peuvent être identifiés dans une vidéo numérique décompressée [23]:

**Le flou (blurring)** : se manifeste comme une perte des détails spatiaux. Il est dû à la suppression des coefficients hautes fréquences, via une quantification grossière.

**L'effet de bloc (blocking effect)** : ce type de distorsion est caractérisé par l'apparition des blocs visuels dans la séquence décompressée. Cet effet est dû à la différence de quantification indépendante aux différents blocs (en général 8x8 pixels) dans les schémas de compression utilisant une DCT par bloc, qui introduit des discontinuités à la frontière des blocs adjacents.

**Un effet (escalier)**: apparaitre sur les contours obliques. Cet effet est dû au fait que les fonctions de base de la DCT sont plus adaptées à la représentation des contours horizontaux et verticaux. Les contours, dont l'orientation n'est ni horizontale ni verticale, nécessitent davantage de coefficients hautes fréquences pour obtenir une représentation précise.

**Un effet (Bleeding)** : est une bavure des couleurs entre deux zones de chrominance fortement différentes. Il est le résultat de la suppression des coefficients hautes fréquences dans les composantes chromatiques.

**L'effet d'ondulation ou de sonnerie (ringing)** : il a perçu dans les hauts bords du contraste Cet effet est un résultat direct de la quantification provoquant des irrégularités dans la reconstruction de la vidéo. L'effet de ringing se produit à la fois sur la luminance et la chrominance.

**Le papillotement (flickering)** : apparait principalement dans les zones texturées. La texture des blocs de ces zones est compressée avec des pas de quantifications variant au cours du temps, ce qui a pour effet de créer des papillotements dans ces zones.

**L'effet (mosquito)** : est une distorsion temporelle observée principalement sur les zones faiblement texturées comme des fluctuations de la luminance (ou de la chrominance) autour des contours à fort contraste ou des objets en mouvement. Il est la conséquence de la variation du choix des paramètres de codage d'une même zone d'une scène dans les images successives d'une vidéo.



**Perte du paquet (paket loss) :** Se produit pendant la transmission où quelques frames de la vidéo sont perdus, cela réduit la qualité perçue de la vidéo.

Voila quelques exemples de ces artéfacts :

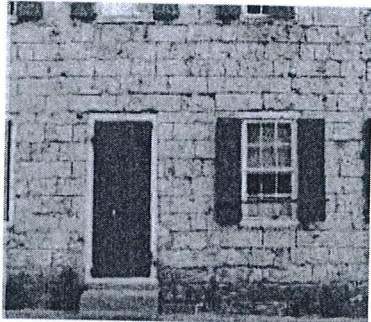


Figure II.23 : Effet de fleu

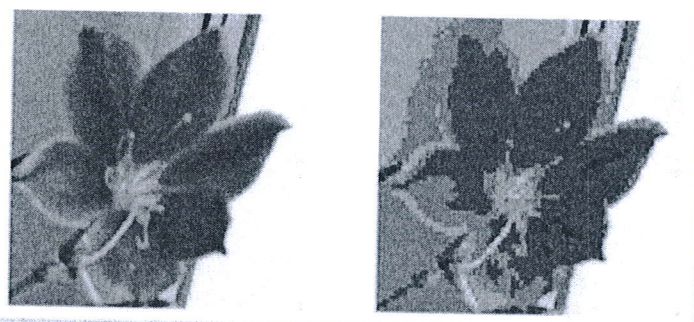


Figure II.24 : Effet de block.

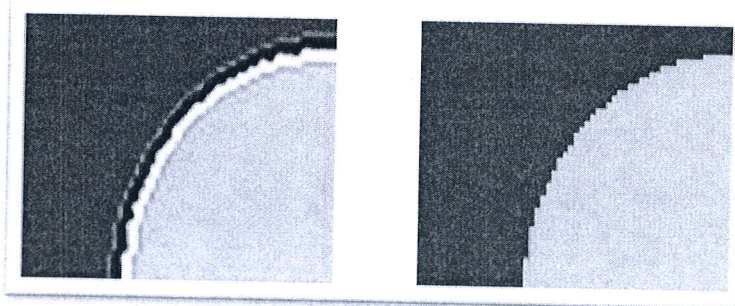


Figure II.25 : Effet de ringing.

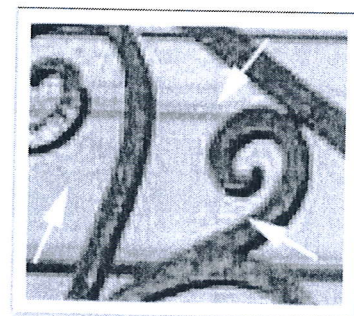


Figure II.26 : Effet de Mosquito.

## II.5 Conclusion

Ce deuxième chapitre se décompose de trois parties. La première partie présente la technique de compression des vidéos, les types de redondances, définition de codeur, la méthode de transformation TCD et le schéma de base d'une compression. Tandis que la deuxième partie consiste à présenter quatre formats de normes de compression MPEG2, H.264, Erreur de transmission IP et Erreur de transmission par wifi, On donne pour chacune son mode de fonctionnement et ses caractéristiques. Dans la dernière partie on cite les distorsions liées à la compression de la vidéo numérique.

*Chapitre III : Métriques  
d'évaluation de qualité des vidéos*

### III.1 Introduction

Ces dernières années, l'évolution des applications multimédia est de plus en plus importante. Dans plusieurs applications d'imagerie telles que la restauration, le stockage, l'imagerie médicale, l'impression, la recherche, la transmission et la compression, l'image ou la vidéo peut subir plusieurs dégradations ou distorsions. Cependant, l'évaluation de la qualité s'avère nécessaire pour contrôler le niveau visuel de la dégradation ou pour contrôler les améliorations apportées sur la vidéo. On distingue deux grandes approches : l'évaluation subjective et l'évaluation objective qu'on va étudier dans ce chapitre, et les méthodes de comparaisons entre les deux. L'évaluation objective contient trois grandes classes, avec référence complète (RC), avec référence réduite (RR) et sans référence (SR). On va représenter quelques métriques pour chacune de ces classes.

### III.2 Métriques subjectives d'évaluation de qualité de la vidéo

Les métriques subjectives nécessitent une normalisation des conditions de tests. Cette normalisation facilite l'appréciation des résultats et minimise l'influence de paramètres perturbateurs. L'I.T.U (International Telecommunication Union) propose plusieurs recommandations dont par exemple la recommandation BT.500-10 [ITU-R Rec.BT.500-10 00]. Cette recommandation contient un certain nombre de règles pour normaliser l'environnement de test. Ces règles initialement prévues pour des tests de qualité d'images de télévision peuvent être utilisées plus généralement pour l'évaluation subjective de qualité de vidéos.

Les métriques subjectives d'évaluation de qualité consistent à présenter à des observateurs des vidéos potentiellement dégradées et dont ils doivent juger la qualité. Ces vidéos peuvent être accompagnées de leur version originale. Ces métriques subjectives peuvent être divisées en trois grandes familles :

- les méthodes comparatives.
- les méthodes à simple stimulus.
- les méthodes à double stimuli.

#### III.2.1 Méthodes comparatives

Ces méthodes consistent à présenter deux vidéos aux observateurs qui doivent caractériser la relation entre les deux présentations [24]. Pour ce faire, les observateurs peuvent avoir à disposition deux types d'échelle d'évaluation : une échelle d'évaluation par catégorie ou une échelle d'évaluation continue. Les deux types d'échelle indiquent la présence de différences perceptibles, et parfois le degré de différences perceptibles, comme celles indiquées sur la figure



III.1. Les échelles par catégorie limitent le choix des observateurs à un ensemble de catégories définies sémantiquement, alors que les échelles continues offrent davantage de souplesse en permettant aux observateurs de choisir tout point d'une droite tracée entre les différents qualificatifs.

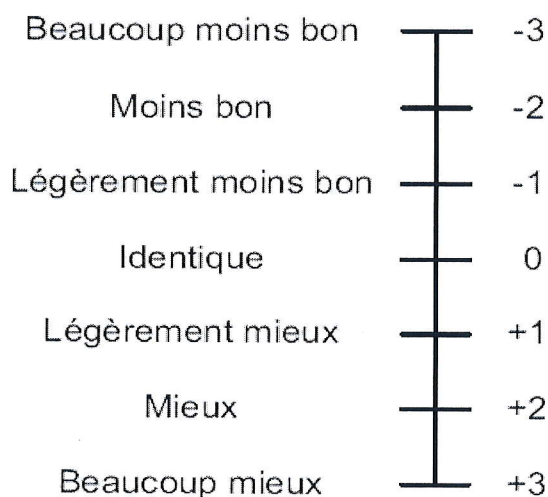


Figure III.1 : Echelle comparative de l'I.T.U.

### III.2.2 Méthodes à simple stimulus

Ces méthodes consistent à présenter à l'observateur une seule vidéo à partir de laquelle il doit juger et noter la qualité globale [25]. Elles sont nommées SSCQS (Single Stimulus Continuous Quality Scale) si elles utilisent une échelle continue de notation de la qualité, ou SSIS (Single Stimulus Impairment Scale) si l'échelle de qualité n'est constituée que de quelques catégories. Le sigle ACR (Absolute Category Rating) est aussi utilisé dans la littérature pour désigner les méthodes à simple stimulus de type SSIS.

En pratique, les échelles continues de qualité sont en réalité discrétisées mais conservent un nombre de catégories supérieures à celui dont l'observateur est conscient. Souvent, une échelle à cent valeurs est utilisée.

Les différents types d'échelle sont présentés sur la figure III.2 (a, b). Pour les échelles à cinq catégories, l'I.T.U. recommande l'utilisation d'une échelle à cinq notes (de qualité ou de dégradations). Cependant, si la dynamique des dégradations est importante, des échelles à six ou sept notes, voire même à onze notes, peuvent permettre une évaluation plus juste des dégradations importantes.



**Figure III.2** – Exemple d'échelle de notation.

(a) continue permettant d'évaluer la qualité d'une vidéo.

(b) par catégories de l'I.T.U. permettant d'évaluer la qualité (à gauche), ou les dégradations (à droite) d'une vidéo.

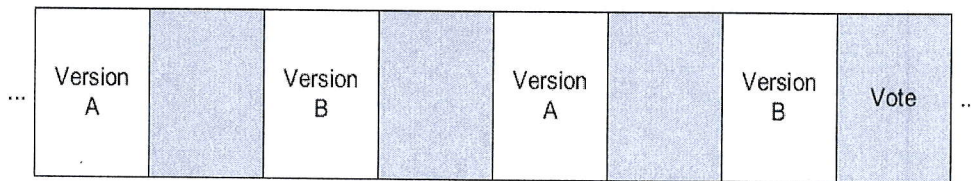
Dans le cadre de l'évaluation subjective de la qualité de vidéos, d'autres méthodes ont été imaginées. Contrairement aux méthodes SSCQS ou SSIS ne permettant d'obtenir qu'une note globale de la vidéo à évaluer, la méthode appelée SSCQE (Single Stimulus Continuous Quality Evaluation) permet d'obtenir une évaluation continue (deux notes par seconde) de la qualité de la vidéo. Dans cette méthode, les observateurs notent la qualité de la vidéo présentée au moyen d'un dispositif coulissant qu'ils déplacent dans un sens, ou dans l'autre, sur une échelle de notation continue, en fonction de leur perception momentanée de la qualité de la vidéo. La durée des vidéos présentées avec cette méthode est généralement plus importante qu'avec les méthodes SSCQS ou SSIS (de l'ordre de dix secondes en SSCQS ou SSIS, contre plusieurs minutes en SSCQE).

### III.2.3 Méthodes à double stimuli

A la différence des méthodes à simple stimulus, ces méthodes consistent à présenter à l'observateur deux vidéos : une version de référence et une version à évaluer qui est potentiellement dégradée [24]. Comme pour les méthodes à simple stimulus SSCQS et SSIS, les méthodes à double stimuli sont nommées DSCQS (Double Stimuli Continuous Quality Scale) si elles reposent sur une échelle continue de notation de qualité, et DSIS (Double Stimuli Impairment Scale) si cette échelle de qualité ne contient que quelques catégories. Le sigle DCR

(Degradations Category Rating) est aussi utilisé dans la littérature pour désigner les méthodes à double stimuli de type DSIS.

Dans la méthode DSCQS les vidéos sont présentées par paire. Chaque paire contient une référence (non dégradée) et une version à juger. Il n’y a pas d’a priori sur l’ordre des présentations et chaque paire peut être présentée plusieurs fois avant que l’observateur note la qualité. L’observateur doit noter la qualité des deux présentations. La figure III.3 illustre le déroulement de l’évaluation d’une paire avec la méthode DSCQS.



**Figure III.3 :** Illustration de la méthode DSCQS.

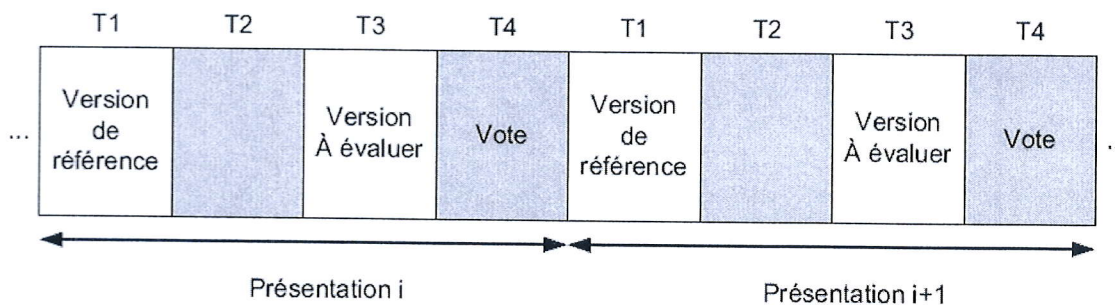
Les versions A et B sont présentées deux fois, la référence peut être indifféremment la version A ou la version B.

La méthode DSIS consiste, elle, à présenter aux observateurs les images ou les vidéos dans un ordre particulier : d’abord la référence (clairement identifiée), puis seulement après la version dont la qualité est à évaluer.

A la suite de quoi l’observateur doit noter la qualité de la deuxième version par rapport à la première version (la référence). Chaque séance obéit à plusieurs règles (par exemple sa durée est limitée à une quarantaine de présentations). Une présentation comprend quatre phases :

- T1 : affichage de la référence,
- T2 : temps mort de séparation,
- T3 : affichage de la version à évaluer,
- T4 : temps de vote.

La figure III.4 illustre le déroulement d’une présentation avec la méthode DSIS.



**Figure III.4 :** Illustration de la méthode DSIS sur deux présentations consécutives.



L'I.T.U. recommande d'utiliser une échelle à cinq notes (de qualité ou de dégradations). Certains auteurs utilisent des échelles différentes comme le laboratoire LIVE 1 qui effectue des tests avec des notes entre un et cent.

Dans le cadre de l'évaluation subjective de la qualité de vidéos, il existe aussi une méthode d'évaluation continue de la qualité à double stimuli. Cette méthode est appelée DSCQE (Double Stimulus Continuous Quality Evaluation) [25]. Le fonctionnement est le même que pour la méthode SSCQE, à la différence que la vidéo originale est présentée en même temps que la vidéo à évaluer. Dans cette méthode, les observateurs notent la qualité de la vidéo présentée au moyen d'un curseur qu'ils déplacent dans un sens ou dans l'autre selon les variations de qualité qu'ils observent. De même que pour les méthodes à simple stimulus, la durée des vidéos présentées avec cette méthode est généralement plus importante qu'avec les méthodes DSCQS ou DSIS.

#### III.2.4 Méthodes à stimuli multiple

Ces méthodes consistent à présenter à l'observateur plusieurs vidéos à évaluer en laissant à l'observateur une grande liberté sur l'ordre de présentation des différentes versions. Dans la méthode SAMVIQ (Subjective Assessment Methodology for Video Quality), l'observateur doit évaluer une référence dite cachée car non identifiée explicitement par l'observateur et plusieurs versions dégradées. Cette méthode est très différente des autres méthodes présentées en particulier dans la façon dont les vidéos sont présentées à l'observateur. Il est donné beaucoup plus de liberté à l'observateur qui peut revoir plusieurs fois chaque version et qui peut corriger des notations. L'observateur peut comparer les versions dégradées entre elles, ainsi que par rapport à la référence explicite. Dans cette méthode une échelle de notation continue est utilisée (cf. figure III.2(a)). L'intérêt de cette méthode est de rendre plus cohérentes les diverses notations et donc de réduire les erreurs de notation. Cependant, en contrepartie de la liberté donnée à l'observateur, la durée des tests n'est plus réellement maîtrisée.

La durée de présentation de chaque vidéo est comparable avec les autres méthodes d'évaluation non continue (de l'ordre de dix secondes).

### III.3 Avantages et inconvénients des métriques subjectives

Les avantages et les inconvénients [26] des indicateurs subjectifs sont:

- Les résultats obtenus sont valides pour des vidéos compressées ou non
- Une moyenne des mesures de plusieurs observateurs est un indicateur valide quant à la qualité globale des vidéos ou images.

Mais :

- Les méthodes et critères de tests sont nombreux
- La configuration matérielle est lourde et difficilement reproductible
- De nombreux observateurs doivent être sélectionnés et surveillés
- La complexité générale est énorme et pose surtout des problèmes de temps.

Ces inconvénients sont la cause d'apparition de l'approche objective pour les éviter, Elle est automatisée et moins coûteuse que l'approche subjective et elle ne nécessite pas d'interaction avec l'utilisateur. Ces mesures sont divisées en trois catégories : avec référence complète(RC), avec référence réduite (RR) et sans référence (SR).

### III.4 Métriques objective d'évaluation de qualité de la vidéo

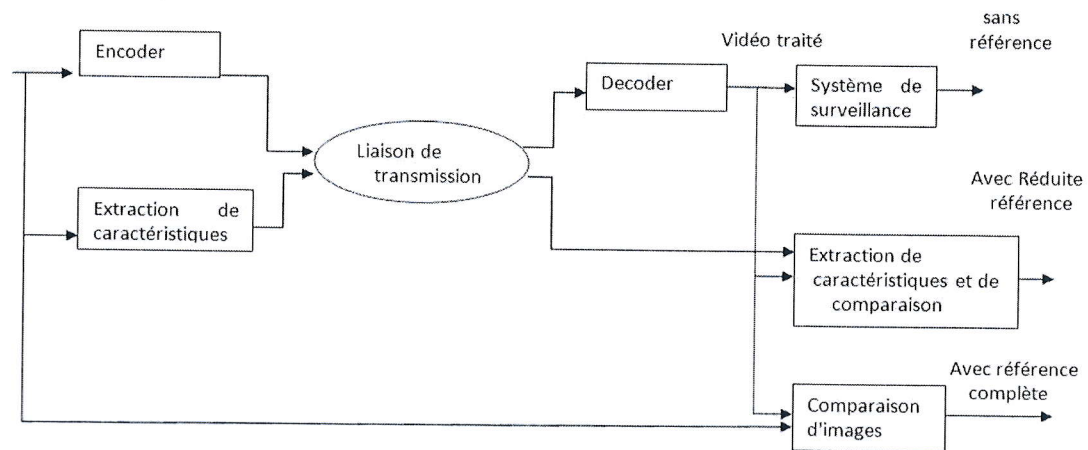


Figure III.5 : Métriques objective d'évaluation de qualité de la vidéo.

#### III.4.1 Les métriques avec Références Complètes (RC)

Les métriques avec Références Complètes [27] réalisent une comparaison directe entre la vidéo jugée et la vidéo de référence. Cette approche nécessite la disponibilité d'une image ou séquence de référence, ce qui représente une restriction importante à l'usage d'une telle mesure.

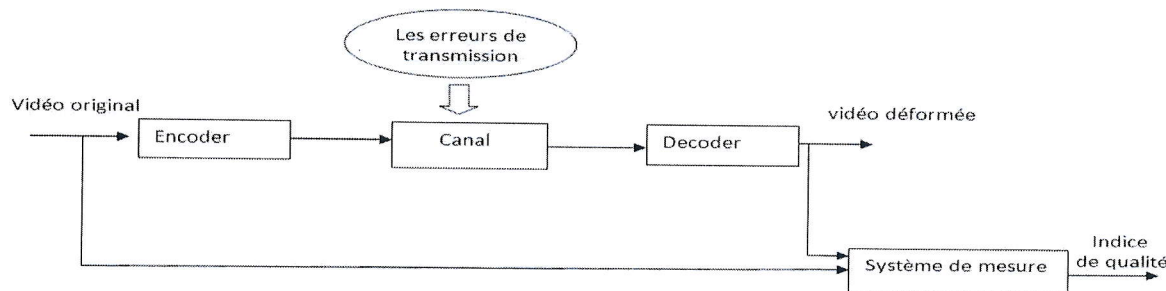


Figure III.6 : diagramme de fonctionnement des métriques RC.

#### III.4.1.1 PSNR (Peak Signal to Noise Ratio)

Le PSNR est une métrique utilisée pour évaluer la fidélité d'une image ou vidéo (par rapport à celle de référence). C'est un critère qui traite l'image purement comme un signal numérique. Il se base sur l'erreur quadratique moyenne MSE (Mean Squared Error) liée à la différence entre l'image dégradée et celle de référence. La valeur du PSNR est exprimée en décibel (dB) et donnée par l'équation suivante [27]:

$$\text{PSNR} = 10 \log_{10} \frac{L^2}{\text{MSE}} \quad \dots 13$$

MSE (Mean square error) est l'erreur quadratique moyenne (EQM) définie par :

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2$$

- L : la dynamique des valeurs des pixels (L = 255 pour un codage des pixels sur 8 bits).
- N : le nombre de pixels dans l'image ou la vidéo.
- $x_i$  et  $y_i$  : les  $i$ èmes pixels de l'image ou vidéo originale et dégradée, respectivement.

Une valeur de PSNR supérieure à 34 dB indique une assez bonne qualité d'image. MSE et PSNR sont largement utilisés car leur implémentation est facile et leur complexité est faible. Mais plusieurs tests ont démontré qu'ils ne sont pas corrélés suffisamment avec la perception humaine, où l'image visuellement plus dégradée a un PSNR plus élevé et une MSE plus petite que l'image dont la qualité visuelle est meilleure [28].

#### VI.1.2 SSIM (Structural Similarity Index)

SSIM (Structural Similarity Index) est une métrique qui base sur la similarité structurelle entre deux images comme le cas de l'oeil humain qui est sensible aux changements dans la structure [29]. Cette métrique est basée sur trois composantes : la luminance, le contraste et la structure de l'image.



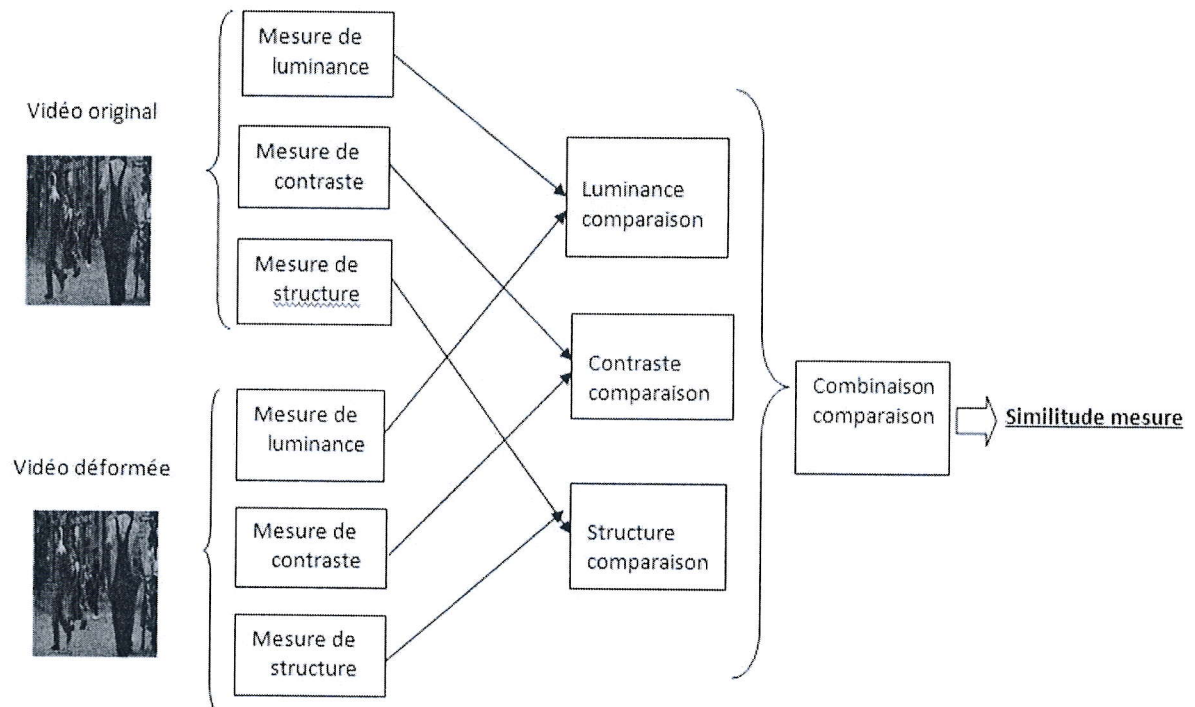


Figure III.7: diagramme de fonctionnement de SSIM.

Soit  $X = \{x_i | i=1,2,\dots, N\}$  et  $Y = \{y_i | i=1,2,\dots, N\}$  l'image ou la vidéo originale et compressée respectivement. L'équation de la métrique SSIM est donnée par :

$$SSIM(x, y) = [l(x, y)]^\alpha [c(x, y)]^\beta [s(x, y)]^\gamma \dots 14$$

$$l(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \quad c(x, y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \quad s(x, y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}$$

- $\alpha, \beta,$  et  $\gamma$  sont des paramètres qui définissent l'importance des trois composantes
- $l(x, y)$  qui est la comparaison de la luminance.
- $c(x, y)$  est la comparaison du contraste comparaison.
- $s(x, y)$  est la comparaison structurelle.
- $C_1, C_2, C_3$  sont des constantes.
- $\mu_x, \mu_y$  sont les moyennes.
- $\sigma_x, \sigma_y, \sigma_{xy}$  sont les variances.

La valeur de SSIM est comprise entre 0 et 1. Plus la valeur de SSIM proche de 1, plus les vidéos se ressemblent.

### III.4.2 Les métriques avec Références Réduites (RR)

Les métriques objectives de qualité avec référence réduite [27] diffèrent de critères avec référence complète par le fait que l'image ou la vidéo originale n'est pas disponible. La référence est alors remplacée par une représentation formée de descripteurs pertinents (attributs), dont on doit estimer la qualité est basée uniquement sur ces descripteurs.

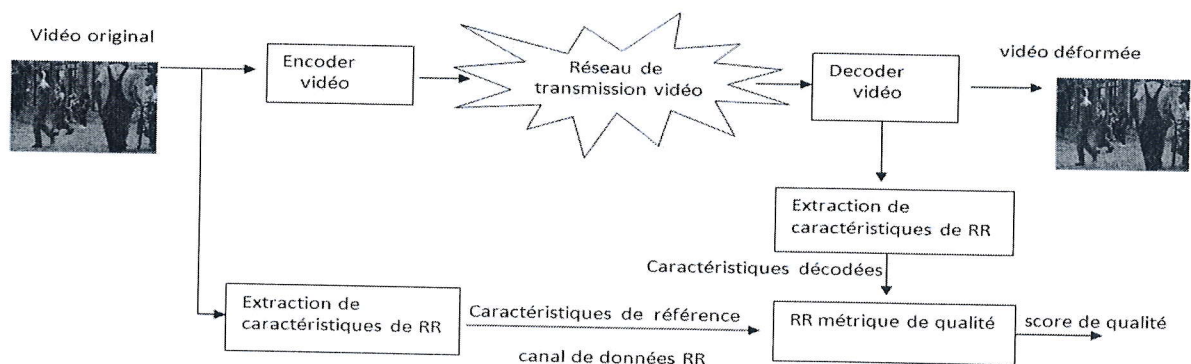


Figure III.8 : diagramme de fonctionnement des métriques RR.

Dans la littérature, rares sont les critères d'évaluation de la qualité d'une vidéo avec référence réduite. Ceci est dû au choix difficile de descripteurs qui soient une fidèle représentation de la vidéo et à la complexité de leur extraction.

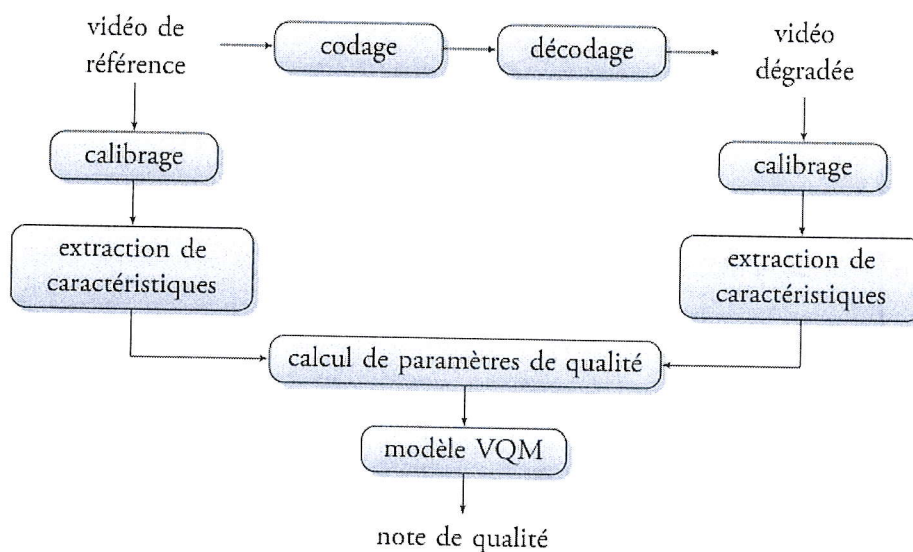
#### III.4.2.1 Métrique de qualité vidéo VQM

VQM (Video quality metric) [30] est une métrique objective de qualité à référence réduite développée par l'ITS (Institute for Telecommunication Sciences), la branche d'ingénierie du NTIA (National Telecommunications and Information Administration) au ministère américain du commerce. Cette métrique est proposée par Wolf et Pinson [Wolf et Pinson, 2002].

C'est la métrique ayant obtenu les meilleurs résultats lors de la seconde campagne d'évaluation de critères objectifs de qualité vidéo réalisée par VQEG (Video Quality Experts Group) en 2003. Les auteurs fournissent un logiciel permettant de mesurer directement la qualité de séquences avec le critère HAL.

VQM extrait des caractéristiques d'une séquence vidéo dégradée et de sa référence pour évaluer l'amplitude des dégradations. VQM est inspiré de phénomènes haut niveau de la

perception visuelle. En particulier, Les différentes phases d'opération de VQM sont données [31] figure III.9.



**Figure III.9 :** Structure globale du critère de qualité vidéo VQM [Wolf et Pinson, 2002].

**Calibrage :** Cette étape calibrer la vidéo échantillonnée en préparation pour l'extraction de caractéristiques. Elle estime et corrige le décalage spatial et temporel ainsi que le contraste et la luminosité de décalage de la séquence vidéo traitée par rapport à la séquence vidéo d'origine.

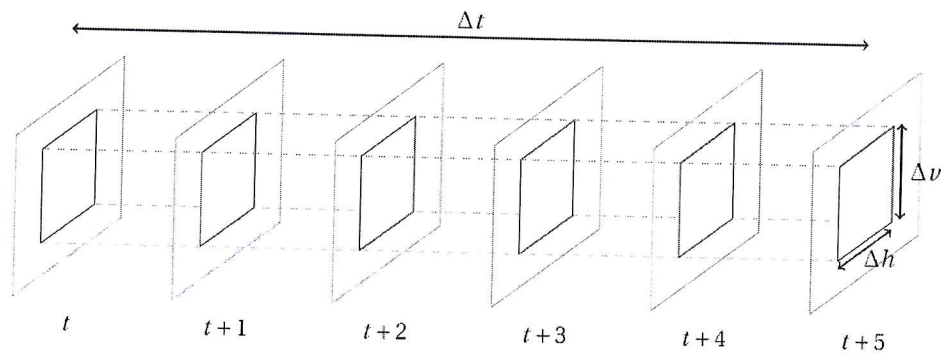
**Extraction de caractéristiques :** Cette étape extrait un ensemble de caractéristiques de qualité qui caractérise les changements de perception dans le domaine spatiale et temporelle, et les propriétés de chrominance des régions spatio-temporelles des flux vidéo en utilisant une fonction mathématique. Une telle région est un ensemble de pixels défini par ses deux dimensions spatiales ( $\Delta h$ ,  $\Delta v$ ) et sa dimension temporelle  $\Delta t$ . La figure III.10 présente la un exemple des régions de taille  $(\Delta h, \Delta v, \Delta t) = (8, 8, 6)$ .

VQM utilise six caractéristiques. Deux caractérisent l'activité spatiale, une troisième les distorsions dans les composantes chromatiques, une quatrième le contraste local, une cinquième la quantité d'information temporelle et enfin la sixième est le produit de la caractéristique du contraste local et de celle de l'information temporelle.

**Calcul de paramètres de qualité :** Cette étape calcule un ensemble de paramètres de qualité qui décrivent des changements de la perception de la qualité de la vidéo en comparant les caractéristiques extraites de la vidéo traitée avec celles extraites de la vidéo originale.

VQM est calculée en utilisant une combinaison linéaire des paramètres calculés à partir des étapes précédentes.

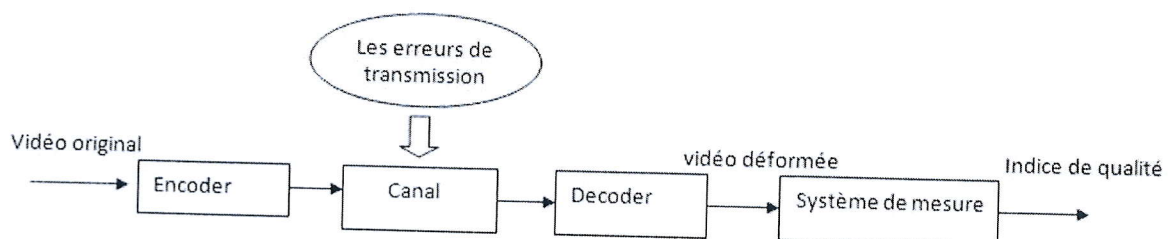




**Figure III.10 :** Tube d'extension spatio-temporelle ( $\Delta h$ ,  $\Delta v$ ,  $\Delta t$ ) utilisé dans VQM [Wolf et Pinson, 2002].

### III.4.3 Les métriques Sans Référence (SR)

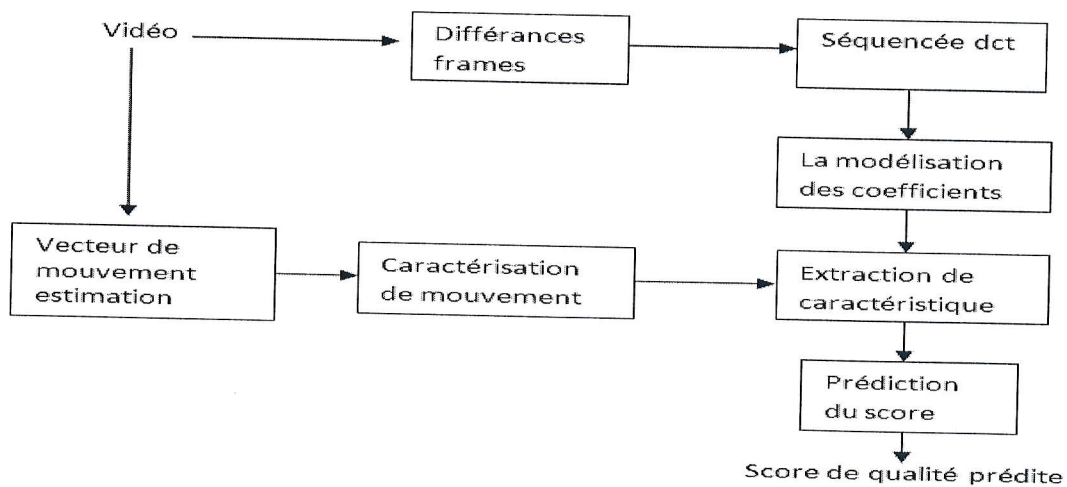
Les métriques Sans Référence (SR) prennent en considération uniquement la vidéo dont on doit estimer la qualité. Elles n'ont pas besoin donc, d'informations de référence [27].



**Figure.III.11 :** diagramme de fonctionnement des métriques NR.

La difficulté ici consiste à distinguer les défauts du contenu ordinaire de l'image ou la vidéo, une distinction que les humains sont capables d'effectuer à partir du contexte et de l'expérience.

### VI.3.1 Métrique de qualité vidéo BLIIND



**Figure III.12 :** Structure de la métrique BLIIND.

Dans cette métrique, les dimensions spatiales et temporelles des signaux de la vidéo est analysés et évalués conjointement. Le comportement d'une vidéo est analysé le long de la dimension temporelle dans deux façons distinctes [32]:

- 1) La différence du frame: les statistiques de différences des frames sont analysées sous le modèle NVS.
- 2) En analysant les types de mouvement qui se produit dans la vidéo et par des caractéristiques de pondération dérivés dans le cadre du modèle NVS (natural video statistics) de l'étape précédente en conséquence.

L'extraction des caractéristiques est effectuée dans le domaine fréquentiel local (DCT), grâce à l'observation que les statistiques des coefficients DCT changent avec le degré et le type de distorsion de l'image.

#### III.4.3.1.1 Modèle de la cohérence du mouvement

Les auteurs caractérisent le contenu temporel d'une vidéo à l'aide d'une structure 2D (dimensions) le modèle de tenseur appliqué aux vecteurs de mouvement calculés d'une vidéo. Un algorithme simple d'estimation de mouvement est appliqué en  $n \times n$  blocs pour déterminer l'emplacement des blocs dans une trame dans la trame consécutive dans le temps [32].

Le tenseur de cohérence du mouvement résume les directions prédominantes de mouvement sur les quartiers. La cohérence 2D de tenseur de mouvement à un pixel donné est donné par:

$$S = \begin{bmatrix} f(M_x) & f(M_x, M_y) \\ f(M_x, M_y) & f(M_y) \end{bmatrix} \dots 15$$

$$f(V) = \sum_{l,k} w[i,j] V(i-l, j-k)^2$$

Où  $M_x(i, j)$  et  $M_y(i, j)$  sont les vecteurs de mouvement horizontal et vertical aux pixels  $(i, j)$  respectivement.

$W$  est une fenêtre de dimension  $m \times m$  sur laquelle le calcul de tenseur effectué est localisé. Les valeurs propres de la cohérence du mouvement tenseur transmettent des informations sur l'alignement spatial des vecteurs de mouvement à l'intérieur de la fenêtre de calcul. Le rapport différence entre deux valeurs propres est un indicateur de degré d'anisotropie du mouvement local (dans la fenêtre), ou à quel point le mouvement est sollicité vers une direction particulière. Ceci est effectivement quantifié par la mesure de cohérence:

$$C = \left( \frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2} \right)^2 \dots 16$$

Où  $\lambda_1$  et  $\lambda_2$  sont les valeurs propres de la cohérence du mouvement tenseur.

Le mouvement global peut être présenté en raison du mouvement de la caméra ou autre grand mouvement dominant. La vitesse du mouvement global peut affecter la perception du contenu de la scène. Par conséquent, notre modèle représente l'ampleur du mouvement global. Elle est calculée simplement le mode des grandeurs de vecteurs de mouvement entre toutes les deux trames consécutives [33].

La valeur absolue de la différence entre le mode et la moyenne des vecteurs de mouvement amplitude par trame est calculée et divisée par l'ampleur du mouvement moyen par image. En d'autres mots, le mode  $M$  et la moyenne  $E$  des grandeurs de vecteurs de mouvement (correspondant à deux trames consécutives), respectivement sont :

$$M = \text{mode}_{\{i=1 \dots m\}} \left( \sqrt{(M_{X(i)})^2 + (M_{Y(i)})^2} \right) \dots 17$$

$$E = \frac{1}{m} \sum_{i=1}^m \sqrt{(M_{X(i)})^2 + (M_{Y(i)})^2} \dots 18$$

Où  $m$  est le nombre de vecteurs de mouvement par trame.



La mesure globale de caractérisation de mouvement est donné par :

$$G = \frac{|E - M|_{ave}}{1 + M_{ave}} \quad \dots 19$$

Les quantités  $M$  et  $|E - M|_{ave}$  sont alors la moyenne des images d'une séquence vidéo résultant en  $M_{ave}$  et  $|E - M|_{ave}$  respectivement.

Cette quantité représente la fraction de mouvement attribué au mouvement non global  $|E - M|_{ave}$  sur le mouvement global  $M_{ave}$ . En soustrayant  $M$  (mouvement global) du mouvement moyen  $E$ , nous obtenons un résidu, et de déterminer quelle fraction de la moyenne le mouvement est contribué par ce que résiduel. Le mouvement global  $G$  est utilisé comme une caractéristique pour la prédiction de la qualité.

#### III.4.3.1.2 Modèle NVS (natural video statistics)

Un bon modèle NVS (natural video statistics) doit capturer le comportement statistique régulier et prévisible des vidéos naturelles. Un tels modèles peuvent être utilisés pour mesurer le degré des distorsions dans les signaux vidéo depuis des distorsions peut prévisible modifier ces statistiques [33].

Le modèles NVS peuvent être considérés comme des duals bas niveaux des modèles perceptifs depuis le HVS (système visuelle humaine) est supposée ont évolué en ce qui concerne les statistiques sur les environs environnement visuel au cours des millénaires.

### III.5 Indicateurs de performance de critères objectifs de qualité

Afin d'évaluer la pertinence de métriques de qualité, il est nécessaire de confronter les notes prédites avec le jugement des observateurs. Pour cela il est nécessaire de constituer une base de test d'images ou de vidéos [24].

La qualité des images ou des vidéos de cette base de test est ensuite évaluée subjectivement au moyen de tests subjectifs d'évaluation de qualité et objectivement par les métriques de qualité dont on veut évaluer les performances. Pour évaluer les performances d'une métrique de qualité, on dispose donc de deux séries de données :

- les MOS, issus des tests subjectifs et correspondants au jugement d'un observateur moyen,
- les notes objectives, issues de la métrique objective à évaluer, notées MOSp.

### III.5.1 Coefficient de corrélation linéaire : indicateur de précision

Le coefficient de corrélation linéaire (Pearson linear correlation coefficient), noté CC ou PLCC exprime la dépendance linéaire entre les mesures objectives  $MOS_p$  et les notes subjectives MOS. Il est donné par la relation suivante [24]:

$$CC = \frac{\sum_{j=1}^{N_{deg}} \sum_{k=1}^{N_{im}} (MOS_{jk} - \overline{MOS})(MOS_{pjk} - \overline{MOS_p})}{\sqrt{\sum_{j=1}^{N_{deg}} \sum_{k=1}^{N_{im}} (MOS_{jk} - \overline{MOS})^2} \cdot \sqrt{\sum_{j=1}^{N_{deg}} \sum_{k=1}^{N_{im}} (MOS_{pjk} - \overline{MOS_p})^2}} \quad \dots 20$$

avec :

–  $MOS_{jk}$  et  $MOS_{pjk}$  étant respectivement les MOS et  $MOS_p$  pour l'image ou la vidéo dégradée issue de l'image ou vidéo originale k ayant subi la dégradation j.

–  $N_{im}$  et  $N_{deg}$  étant respectivement les nombres d'images ou de vidéos originales et le nombre de versions dégradées.

–  $\overline{MOS}$  et  $\overline{MOS_p}$  étant respectivement le MOS moyen et le  $MOS_p$  moyen.

La valeur du coefficient de corrélation linéaire est comprise entre  $-1$  et  $1$ . Plus sa valeur est proche de  $-1$  ou de  $1$ , plus la dépendance linéaire entre les deux séries de nombres est forte.

### III.5.2 Coefficient de corrélation de rang : indicateur de monotonie

Le coefficient de corrélation de rang, noté SROCC (Spearman rank order correlation coefficient), est une mesure de la monotonie, c'est-à-dire qu'il caractérise le degré avec lequel les mesures objectives  $MOS_p$  et les notes subjectives MOS évoluent dans le même sens [24]. Ces deux séries de nombres évoluent dans le même sens si la fonction  $f(MOS) = MOS_p$  (ou  $f(MOS_p) = MOS$ ) est monotone. Pour calculer le SROCC, les MOS et  $MOS_p$  de toutes les vidéos sont ordonnés (par ordre croissant ou décroissant) afin de déterminer le rang de chacun, puis le SROCC est donné par la relation suivante :

$$SROCC = 1 - \frac{6 \sum_{j=1}^{N_{deg}} \sum_{k=1}^{N_{im}} d_{jk}^2}{N_{tot}(N_{tot}^2 - 1)} \quad \dots 21$$

avec :

- $d_{jk}^2$  : différence de classement, ou de rang, entre le MOS et  $MOS_p$  de la vidéo jk (vidéo originale k ayant subi la dégradation j).

- $N_{im}$ ,  $N_{deg}$  et  $N_{tot}$  représentent respectivement le nombre de vidéos originales, le nombre de dégradations, le nombre total de vidéos évaluées ( $N_{tot} = N_{tot} \cdot N_{deg}$ ).

Un coefficient de corrélation de rang proche de 1 est recherché car alors cela signifie que la métrique de qualité classe les vidéos, de la moins bonne à la meilleure qualité, selon le même ordre que les observateurs (SROCC = -1 indiquerait un classement dans l'ordre inverse).

### III.5.2 Outlier ratio : indicateur de cohérence

Cet indicateur permet de mesurer l'aptitude de la métrique de qualité à prédire une note de qualité qui ne soit pas trop éloignée du MOS [24]. Cet indicateur, appelé Outlier ratio et noté OR, exprime le nombre d'images ou de vidéos mal notées par rapport au nombre d'images ou de vidéos testées. Il est déterminé par le rapport suivant :

$$OR = \frac{\text{Nombre de notes aberrantes}}{\text{Nombre total de notes}} \quad \dots 22$$

ou une note est déclarée « aberrante » si elle est en dehors de l'intervalle (intervalle de confiance à 95% environ):  $MOS_{jk}$  et  $MOS_{pjk}$

$$[MOS_{jk} - 2SE_{jk}, MOS_{jk} + 2SE_{jk}]$$

ou  $SE_{jk}$  désigne l'erreur type de la moyenne  $MOS_{jk}$  des notes subjectives pour l'image ou la vidéo originale k ayant subi la dégradation j. L'erreur type  $SE_{jk}$  (standard error) est calculée à partir de l'écart type  $\sigma_{jk}$  selon la relation :

$$SE_{jk} = \frac{\sigma_{jk}}{\sqrt{N_{obs}}} \quad \dots 23$$

ou  $N_{obs}$  correspond au nombre d'observateurs ayant noté l'image ou la vidéo jk. La proportion la plus faible possible de notes « aberrantes » est souhaitée.

### III.5.4 Erreur de prédiction de la qualité

Dans la littérature l'erreur quadratique moyenne, ou plutôt sa racine carrée RMSE (Root mean square error), est utilisée comme indicateur complémentaire [24]. Le RMSE permet de mesurer la distance L2 entre les mesures objectives  $MOS_p$  et les notes subjectives MOS. Elle est donnée par la relation suivante :

avec :



$$RMSE = \sqrt{\frac{\sum_{j=1}^{N_{deg}} \sum_{k=1}^{N_{im}} (MOS_{jk} - MOS_{pjk})^2}{N_{im} \cdot N_{deg}}} \quad \dots 24$$

–  $MOS_{jk}$  et  $MOS_{pjk}$  représentent respectivement les MOS et  $MOS_p$  pour l'image ou la vidéo originale k ayant subi la dégradation j.

–  $N_{im}$  et  $N_{deg}$  représentent respectivement les nombres d'images ou de vidéos originales, le nombre de dégradations.

Plus RMSE est faible, plus la dispersion des  $MOS_p$  autour de leur MOS respectif est faible. Cet indicateur est souvent utilisé pour comparer plusieurs métriques de qualité entre elles. La métrique de qualité ayant la plus faible RMSE est recherchée.

Les métriques ont été testées séparément sur la base de données LIVE qui contient les codecs spécifiques (MPEG2, H264, Erreur de transmission par wifi, et Erreur de transmission par IP). le SROCC et le PLCC entre les scores subjectifs et objectifs (FR, RR et NR) les plus performantes pour la référence sont donnés respectivement dans les Tables I et II [33].

Métriques	F-R			R-R			N-R	
	PSNR	SSIM	MOVIE	STMAD	VQM	RRED	NIQE	BLIIND
<b>MPEG-2</b>	0.643	0.786	0.905	0.929	0.905	0.905	0.523	0.869
<b>H.264</b>	0.714	0.881	0.881	0.952	0.786	0.905	0.541	0.839
<b>Wireless</b>	0.691	0.691	0.786	0.810	0.762	0.762	0.280	0.815
<b>IP</b>	0.600	0.543	0.771	0.771	0.771	0.771	0.276	0.779

**Table III.1** comparaison entre les métriques Full-Reference et Reduced-Reference et Non-Reference par le SROCC.

Métriques	F-R			R-R			N-R	
	PSNR	SSIM	MOVIE	STMAD	VQM	RRED	NIQE	BLIIND
<b>MPEG-2</b>	0.696	0.805	0.955	0.942	0.943	0.904	0.490	0.924
<b>H.264</b>	0.698	0.851	0.919	0.947	0.850	0.892	0.579	0.893
<b>Wireless</b>	0.798	0.634	0.920	0.904	0.943	0.806	0.387	0.951
<b>IP</b>	0.733	0.726	0.895	0.901	0.896	0.816	0.443	0.946

**Table III.2** comparaison entre les métriques Full-Reference et Reduced-Reference et Non-Reference par le PLCC.

VQM et vidéo RRED [34] sont les plus performants dans les approches réduit référence. D'autre part, MOVIE [35] et ST-MAD [36] sont très compétitifs (en termes de prédiction de performance).

BLIIND surpasse clairement le NIQE, le PSNR et le SSIM, bien que BLIINDS n'atteigne pas le niveau de performance de MOVIE et ST- MAD, mais sa performance est presque aussi bon et beaucoup moins coût de calcul, car les vidéos de référence ne peuvent pas être incluses dans les ensembles de test, ce qui conduise à des corrélations trompeusement plus élevées.

### III.6 Conclusion

Dans ce chapitre, nous avons présenté les métriques de l'évaluation subjective et objective de la qualité des vidéos, dans la première partie de ce chapitre nous détaillons Les différents protocoles de tests subjectifs aussi citer leur avantage et inconvénients. Dans la deuxième partie nous avons présenté les trois métriques de l'évaluation objective (RC, RR, SR) et leurs fonctionnements, dans la troisième partie nous avons présenté quelques méthodes de comparaisons entre les deux approches (objective, subjective).

*Chapitre IV : Expérimentation &  
résultats*



## IV.1 Introduction

L'évaluation de la qualité des vidéos représente un noyau agissant dans la conception des systèmes de communication vidéo, pendant le traitement de la vidéo. Notre travail est de faire une évaluation automatique d'une partie de la base de données LIVE en implémentant quelques métriques de qualité visuelle des vidéos.

Ce chapitre est ciblé vers une présentation de l'environnement matériel et logiciel utilisés pour le développement de notre application, nous expliquerons éventuellement notre choix relatif aux langages de programmation et des outils utilisés. Nous donnerons ensuite une description des résultats obtenus accompagné de quelques prises d'écran.

## IV.2 Problématique

1. Comment évaluer la qualité d'un fichier vidéo ?
2. Quelle est la mesure prise comme critère d'évaluation de la qualité d'une vidéo ?
3. Comment comparer entre les métriques d'évaluation d'un vidéo ?

## IV.3 Environnement du travail

Nous avons présenté dans cette partie les différents outils utilisés dans notre application, les outils matériels, le choix des outils logiciels et la base de donnée LIVE.

### IV.3.1 Environnement matériel

Nous rappelons les caractéristiques de l'ordinateur sur lesquelles nous avons développé l'application parce qu'elles peuvent donner une idée sur les conditions du travail. L'application a été développée sur un ordinateur portable Lenovo qui se caractérise Par :

Processeur	Intel(R) pentium(R) CPU N3530@2.16GHz 2.16 GHz
Mémoire installée (RAM)	2.00Go (1.89Go utilisable)
Type de système	Système d'exploitation 64bits

**Tableau IV.1 :** Caractéristique de l'environnement matériel de travail.

### IV.3.2 Environnement logiciel

Nous utilisons MATLAB comme langage de programmation version : R2013a Sun Feb 17 18:06:36 2013.



**Matlab** (abréviation de " **Matrix Laboratory** "), est un environnement informatique conçu pour le calcul matriciel, Le but initial de Cleve Moler, concepteur de ce logiciel en 1981 était de fournir un accès aisé aux logiciels de haute qualité issus des projets LINPACK (LIN pour « linéaire » = linéaire) et EISPACK (EIS pour « eigenvalues » = valeurs propres) [37]. Le Matlab est développé par la société Mathworks (<http://www.mathworks.com>). L'élément de base est une matrice dont la dimension n'a pas à être fixée. Matlab est un outil puissant qui permet la résolution de nombreux problèmes dans différents domaines (traitement d'image, traitement du signal, optimisation, contrôle ...).

- Programmation infiniment plus rapide pour le calcul et pour l'affichage.
- Une librairie très riche.
- Un langage interprété, c'est-à-dire que les instructions sont exécutées immédiatement après avoir été tapées.
- Possibilité d'exécuter du code en dehors du programme.
- Possibilité d'inclure un programme en C/C++.
- Sa facilité d'emploi avec des nombres complexes et ses possibilités d'affichages graphiques en font un outil intéressant pour bien différents types d'applications.
- Code facile à comprendre et très lisible.
- Une aide interactive très bien faite est disponible pour toutes les commandes MATLAB.

### IV.3.3 description de la Base de données LIVE

La base de données LIVE contient dix vidéos originales de haute qualité avec une grande variété de contenus comme des vidéos de référence. Pour pouvoir tester au maximum les métriques objectives de la qualité vidéo, Un ensemble de 150 vidéos dégradées ont été créées à partir de ces vidéos de référence (15 vidéos dégradée par référence) en utilisant quatre types de distorsion différents:

- Wireless (Erreur de transmission par wifi)
- IP (Erreur de transmission par IP)
- H.264

- MPEG2

Chaque vidéo de la base de données a été évaluée par 38 humains, dans une étude de stimulation unique avec élimination de référence cachée, où les chercheurs ont marqué la qualité vidéo sur une échelle continue de la qualité. Mean Opinion Scores (MOS) obtenu à partir des évaluations subjectives, ainsi que la référence et les vidéos dégradées sont disponibles dans la base de données [38].

Tous les fichiers vidéo ont le format YUV 4: 2: 0 sans séquence audio, avec résolution spatiale de 768x432 pixels, aussi que toutes les séquences vidéo sont de type progressif

#### IV.3.3 .1 Nommage de fichier

Chaque nom de fichier a le format suivant "xx # \_ \$ fps.yuv".

"**Xx**" : représente la séquence vidéo de source / référence, qui a été utilisé pour créer la vidéo de test. Il y a dix vidéos de référence dans la base de données LIVE et les formats utilisés pour chacun sont:

"**bs**" Blue sky, "**mc**" Mobile and Calendar, "**pa**" Pedestrian Area, "**pr**" Park run, "**rb**" Riverbed, "**rh**" Rushhour, "**sf**" Sunflower, "**sh**" Shields, "**st**" Station, "**tr**" Tractor

"**#**" Est un nombre qui indique la catégorie de la distorsion qui a créer la vidéo de test. Il existe quatre catégories de distorsion dans la base de données LIVE :

- "1" la vidéo de référence d'origine.
- "2", "3", "4", "5" distorsions sans fil (Wireless) quatre vidéos de test par référence.
- "6", "7", "8" distorsions IP, trois vidéos de test par référence.
- "9", "10", "11", "12" compression H.264, quatre vidéos de test par référence.
- "13", "14", "15", "16" compression MPEG-2, quatre vidéos de test par référence.

"**S**" : Représente le nombre qui indique la fréquence d'images de la séquence vidéo images par seconde (fps). Toutes les vidéos de la base de LIVE ont une fréquence de 25 fps ou 50 fps.

Exemple : bs1\_25fps.yuv: est la vidéo de référence d'origine. bs6\_25fps.yuv, bs7\_25fps.yuv et bs8\_25fps.yuv sont trois tests vidéos obtenues à partir de la référence en utilisant des distorsions IP.

- "pa", "rb", "rh", "sf", "sh", "st" et "tr" ont 250 frames avec fréquence de 25fps = 10 seconds de la durée de la vidéo.
- "bs" a 217 frames avec fréquence de 25fps= 8.68 seconds de la durée de la vidéo.
- "mc", "pr" et "sh" ont 500 frames avec fréquence de 50fps = 10 seconds de la durée de la vidéo.



### IV.3.3 .2 Format de données subjectives

Les données subjectives représentent la mesure prise comme critère d'évaluation de la qualité d'une vidéo, ces données sont fournies dans deux fichiers :

- "Live\_video\_quality\_seqs.txt" qui contient les noms des vidéos.
- "Live\_video\_quality\_data.txt" qui contient les notes subjectives correspondant.

Chaque fichier dispose de 150 lignes correspondant 150 vidéos dégradées de la base de données.

"Live\_video\_quality\_data.txt" contient les MOS suivie par l'écart type des scores MOS pour la vidéo nommée en ligne i de "live\_video\_quality\_seqs.txt". Par exemple :

La ligne 50 de fichier live\_video\_quality\_seqs.txt est "Tr6\_25fps.yuv" et la ligne 50 du live\_video\_quality\_data.txt est "73,4730 11,2189". Cela signifie que les MOS pour la vidéo "tr6\_25fps.yuv" est 73,4730 et l'écart-type des scores MOS est 11,2189.

## IV.4 Scénario de l'application

On va d'abord lire le fichier YUV, puis faire l'évaluation de cette vidéo en utilisant leur référence. Par implémentation de ces méthodes (PSNR, SSIM, SSIM\_diffé et PSNR\_DCT), après cette étapes on compare entre ces méthodes par le SROOC.

Début :

1. Lire la séquence vidéo (.yuv).
2. Afficher les informations de cette vidéo.
3. Afficher la frame numéro i de la vidéo chargée.
4. Evaluer la qualité de la vidéo dans le domaine spatial :
  - Implémentation de la métrique PSNR.
  - Implémentation de la métrique SSIM.
  - Implémentation de la métrique SSIM\_diffé.
5. Evaluer la qualité de la vidéo dans le domaine fréquentiel:
  - Calculer la DCT de chaque vidéo (originale et dégradée).
  - Calculer le PSNR de résultat de la DCT.
6. Comparer entre les méthodes implémentées (PSNR, SSIM, SSIM\_diffé et PSNR\_DCT) par le SROOC.

Fin.

Nous utilisons dans notre application six vidéos origines qui ont le même nombre de frames et même fréquence, "pa" "rb" "st" "rh" "sf" et "tr" et leurs dégradations :

- "2", "3", "4", "5" (Wireless).
- "6", "7", "8" (IP).
- "9", "10", "11", "12" (H.264).
- "13", "14", "15", "16" (MPEG-2).

#### IV.5 Visualisation des vidéos YUV

Les séquences YUV d'entrée peuvent être visualisées à l'aide du lecteur « YUV Player ». Version utilisée dans notre application est Version 0.0.2 released, voila le lien pour le téléchargement : <http://www.yuvtoolkit.com>.

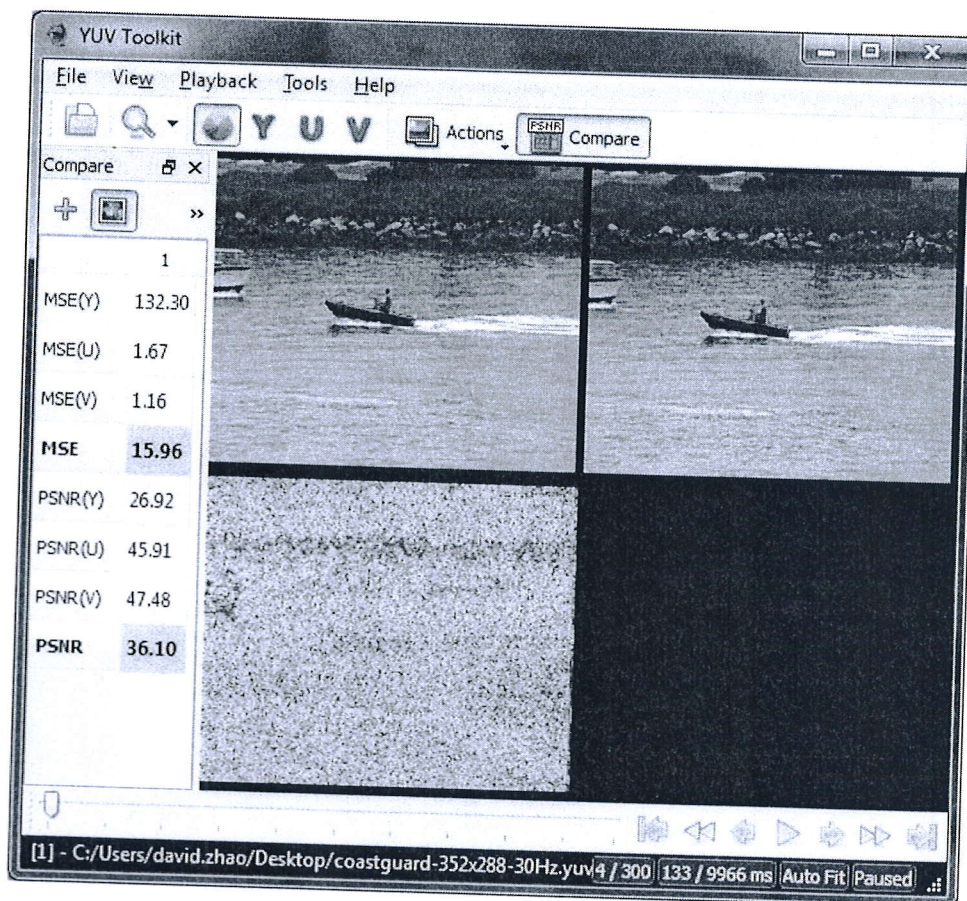


Figure VI.1: lecture YUV Toolkit.

### IV.5.1 Algorithme de lecture d'une vidéo YUV

Pour afficher un fichier YUV, le « YUV Player » doit lire chaque frame de la séquence comme une matrice (.mat en Matlab), puis convertir l'espace de couleur YUV de la séquence vidéos à l'espace de couleur RGB pour pouvoir l'affichage.

On peut résumer le processus de la lecture d'un fichier YUV dans les étapes suivantes :

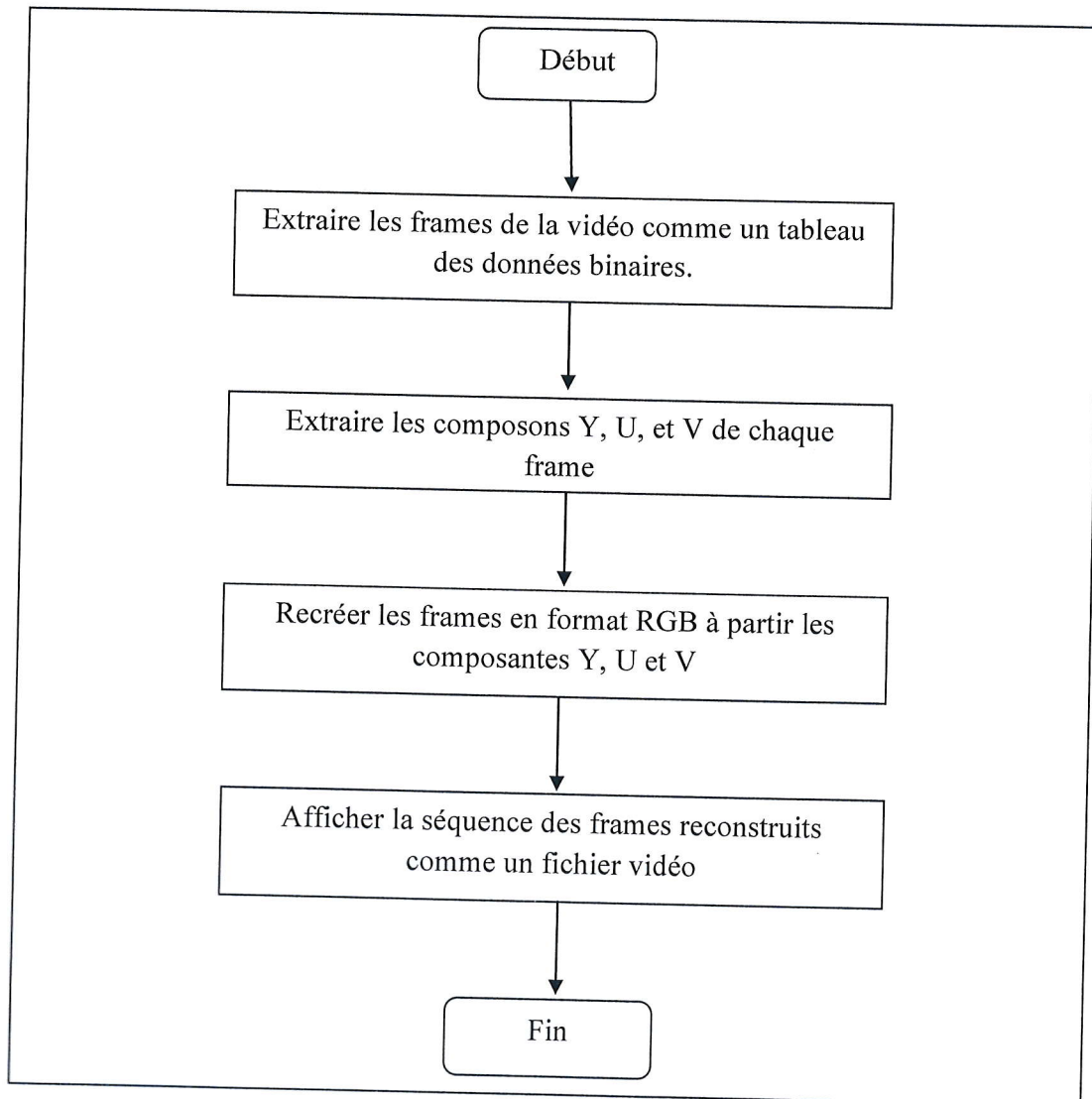


Figure IV.1: Algorithme lecture de fichier YUV.

### IV.6 Les principales fonctions implémentées

Nous donnerons une idée générale de la phase pratique de notre application qui consiste à décrire les fonctions fondamentales avec leurs rôles.



<u>Fonction</u>	<u>rôle</u>
- lecture ()	Lire le fichier YUV
- PSNR_img (img1, img2)	calculer le PSNR d'image originale et dégradée
- SSIM_img (img1, img2)	calculer le SSIM d'image originale et dégradée
- différence (vid)	calculer la différence des composantes YUetV
- dct_vid (vid)	calculer la dct de composante Y
- SROOC_vid (list_sub, list_obj)	calculer le SROOC

## IV.7 Evaluation de la qualité de la vidéo dans le domaine spatial

### IV.7.1 Implémentation de la métrique PSNR

Pour calculer le PSNR des vidéos en sélectionnant la vidéo originale et la vidéo dégradée en utilisant l'algorithme qui calcule le PSNR des images suivant :

---

#### Algorithme de PSNR\_img

---

##### Variable:

Haut, Lrg : hauteur et largeur de l'image

X[i,j], Y[i,j] : valeur de pixel de l'image originale et l'image dégradée respectivement

Mse : erreur quadratique moyenne

psnr : valeur de psnr

err : variable initialiser avec 0

entrée : img\_org, img\_deg

sortie : valeur de psnr

##### Début :

```

    Pour i=1 jusqu'à H faire :
        Pour j=1 jusqu'à L faire :
            err = err + ((x(i,j)) - (y(i,j)))2;
        fin
    fin
    mse = (err / (Haut*Lar));
    psnr = 10*log10(1*/mse);
fin

```

---

### IV.7.2 Implémentation de la métrique SSIM

La métrique SSIM est une métrique qui entraîne de simuler le système visuelle humaine, elle est basée sur trois composantes : la luminance, le contraste et la structure de l'image. Elle est implémentée selon l'algorithme suivant :

---

#### Algorithme de SSIM\_img

---

##### Variable:

Haut, Lrg : hauteur et largeur de l'image

moy1, moy2: les moyennes

var1, var2, var12: les variances

c1,c2,c3,α,β,γ =1

**entrée** : img\_org, img\_deg

**sortie** : valeur de sim

##### Début :

```

Pour i =1 jusqu'à haut faire
    Pour j =1 jusqu'à lar faire
        Calculer (moy1, moy2 ,var1, var2, var12)
    fin
fin

lumin = ((2*moy1*moy2)+c1)/(((moy12)+(moy22))+c1);
contra = ((2*var1*var2)+c2)/(((var12)+(var22))+c2);
struct = (var12+c3)/((var1*var2)+c3);
sim = (lumin)α*(contra)β*(struct)γ;
fin

```

---

### IV.7.3 Implémentation de la métrique SSIM\_diffé

Nous proposons une métrique qui consiste à calculer la différence entre chaque deux images successives qui implémentée l'algorithme **différence\_frame** puis calculer le SSIM de cette différence résultantes à l'objectif de prenez l'aspect temporele. Elle est implémentée comme montre l'algorithme suivant :

---

**Algorithme de SSIM\_diffé**

---

num: nombre de frames

**Début :**

```
    Sélectionner la vidéo originale
    Sélectionner la vidéo dégradée
    Pour chaque vidéo faire
        Pour i=1 jusqu'à num
            Extraire les composantes Y U et V
            Calculer la différence de chaque composante Y U et V
            Convertir les résultats obtenus à l'espace RGB
        fin
    fin
    Calculer le SSIM de ces deux vidéos reconstruites à partir cette différence.
fin
```

---

---

**Algorithme de différence\_frame**

---

num: nombre de frames

**Début :**

```
    Pour i=1 jusqu'à num-1
        Extraire les composantes Y U et V de la vidéo
        Calculer les différences (Y(i)- Y(i+1)), (U(i)- U(i+1)), (V(i)- V(i+1))
        Convertir les résultats obtenus à l'espace RGB
    fin
    Return la table des frames obtenus
fin
```

---



## IV.8 Evaluation de la qualité de la vidéo dans le domaine fréquentiel

### IV.8.1 Implémentation de la métrique PSNR\_DCT

Dans le domaine fréquentiel nous proposons une métrique qui basée sur la Transformation Cosinus Discrète (DCT) elle est appliquée sur la composante Y, qui implémente l'algorithme DCT\_vid puis calculer le PSNR de résultat obtenue en implémentée l'algorithme suivant :

---

#### Algorithme de PSNR\_DCT

---

num: nombre de frames

**Début :**

```

    Sélectionner la vidéo originale
    Sélectionner la vidéo dégradée
    Pour chaque vidéo faire
        Pour i=1 jusqu'à num
            Extraire les composantes Y
            Calculer la dct
        fin
    fin
    Calculer le PSNR de ces deux vidéos reconstruites
Fin
```

---



---

#### Algorithme de DCT\_vid

---

num: nombre de frames

**Début :**

```

    Pour i=1 jusqu'à num-1
        Extraire les composantes Y de la vidéo
        Calculer la différence (Y(i)- Y(i+1))
        Calculer la dct de la différence
    fin
    Return la table de dct obtenus
fin
```

---

## IV.9 Comparaison entre les métriques d'évaluation de la qualité des vidéos

Pour comparer les résultats obtenus dans les deux évaluations avec l'évaluation subjective, nous avons utilisées la méthode de comparaison SROCC, en implémentant l'algorithme suivant :

---

### Algorithme de SROCC\_vid

---

#### Variable:

Nim : nombre de vidéo original

Ndeg : nombre de dégradation

Ctr : variable initialiser avec zéro

**Entrée :** List\_obj : liste des valeurs objective

List\_subj : liste des valeurs subjective

**Sortie :** valeur de srocc

#### Début :

Trier la liste des valeurs objective

Trier la liste des valeurs subjective

**Pour** (chaque valeur objective et subjective dans la liste trier)

    //Obtenir le range de chaque valeur

    a= le range de valeur objective

    b= le range de valeur subjective

**fin**

    Ctr = Ctr + (a - b)<sup>2</sup>

    Ntot = Nim \* Ndeg;

    sroc = 1 - (((6 \* Ctr) / (Ntot \* ((Ntot<sup>2</sup>) - 1)));

**fin**

---

## IV.10 Présentation de l'application

Dans cette partie nous donnerons un aperçu général de la phase pratique de notre application.

La figure IV.2 représente la page d'accueil de notre application qui contient un menu permettant l'accès à d'autres fenêtres. Les fonctionnalités des menus qui sont trouvées dans la figure IV.2 et la figure IV.3 sont expliquées dans le tableau IV.2 et le tableau IV.3.





Figure IV.2 : L'interface principale de l'application.

Lorsqu'on choisit la fonction Entrer à l'application du menu fichier la fenêtre suivante sera affichée.

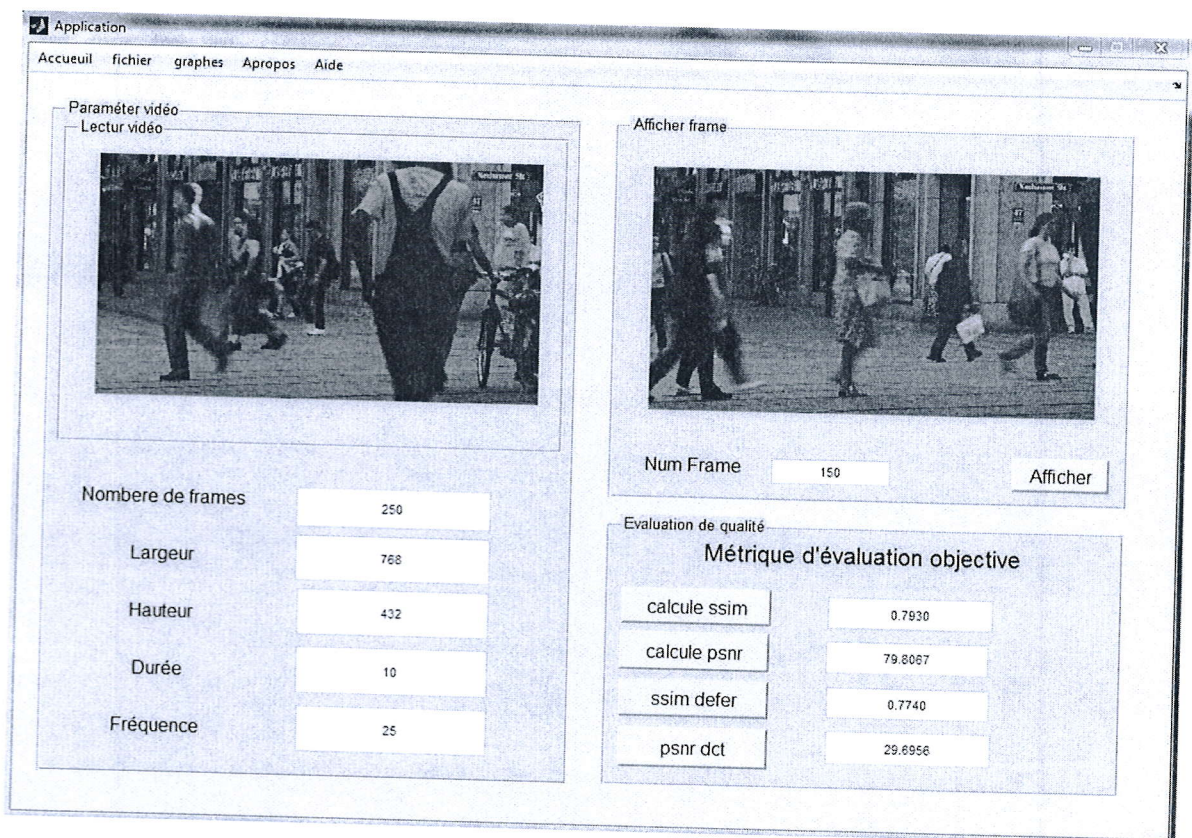


Figure IV.3 : La fenêtre Entrer à l'application.



Menu	Fonctionnalités
Fichier	<b>Entrer à l'application</b> : Ouvrir la fenêtre de traitement qui permet d'entré à l'application. <b>Quitter</b> : Pour quitter l'application.
Apropos	Pour plus d'information sur l'application.
Aide	Information pour nous contacter.

Tableau IV.2 : Les fonctionnalités des éléments de la fenêtre accueil.

Menu	Fonctionnalités
Fichier	<b>Nouveau</b> : pour rouvrir l'application. <b>Lecture_vidéo</b> : pour choisir le vidéo que l'on souhaite lire puis le traiter. <b>Fermer</b> : pour quitter l'application.
Les graphe	<b>PSNR</b> : pour dessiner nuage de point des valeurs psnr des vidéos. <b>SROCC</b> : pour dessiner les courbes de srocc pour chaque méthode( PSNR , SSIM, SSIM_def, PSNR_dct).

Tableau IV.3 : Les fonctionnalités des éléments de la fenêtre fichier et les graphes.

## IV.11 Résultats expérimentaux

### IV.11.1 Résultats de PSNR

Normes	Vidéos	pa	rb	st	rh	sf	tr
Wifi	Deg2	84.3500	75.6334	79.1102	78.1269	77.0822	74.4881
	Deg3	81.1526	78.0891	81.4349	78.9259	78.5772	75.6464
	Deg4	79.8067	78.2147	84.7644	82.8123	83.5015	75.7522
	Deg5	76.6495	79.5390	85.1108	80.3453	81.5423	78.8773
IP	Deg6	79.1807	76.9183	85.3233	83.4644	79.4350	74.1659
	Deg7	84.2639	78.0206	86.1388	79.6614	80.8572	80.0524
	Deg8	83.2648	79.3926	86.7977	86.6321	83.6507	79.3522
H.264	Deg9	84.2994	79.6950	76.4976	61.9529	85.4367	80.2126

	Deg10	82.3993	76.4976	85.1659	83.6403	83.5707	78.9952
	Deg11	76.9812	76.9812	84.9212	82.8020	82.8947	82.8947
	Deg12	80.1594	76.2104	84.4439	81.6410	82.1101	82.1101
MPEG2	Deg13	85.3676	80.0492	85.1659	88.3319	89.0114	82.1119
	Deg14	84.2791	78.4849	84.1300	84.5298	87.0597	80.6088
	Deg15	78.0217	78.0217	82.9351	82.7036	85.4453	78.5710
	Deg16	81.9891	78.0217	82.0276	81.3081	81.3081	78.5710

**Tableau IV.4:** Résultats de PSNR (WIFI, IP H.264, MPEG2).

**IV.11.2 Résultats de SSIM**

Normes	Vidéos	pa	rb	st	rh	sf	Tr
Wifi	Deg2	0.7942	0.7720	0.7915	0.7943	0.8293	0.7718
	Deg3	0.7939	0.7726	0.7910	0.7943	0.8301	0.7732
	Deg4	0.7930	0.7729	0.7917	0.7952	0.8311	0.7723
	Deg5	0.7947	0.7734	0.7918	0.7954	0.8305	0.7729
IP	Deg6	0.7942	0.7723	0.7915	0.7955	0.8300	0.7734
	Deg7	0.7943	0.7729	0.7915	0.7949	0.8304	0.7732
	Deg8	0.7942	0.7724	0.7917	0.7958	0.8310	0.7731
H.264	Deg9	0.7941	0.7731	0.7914	0.7956	0.8311	0.7732
	Deg10	0.7940	0.7729	0.7914	0.7955	0.8309	0.7730
	Deg11	0.7939	0.7726	0.7914	0.7954	0.8308	0.7728
	Deg12	0.7938	0.7723	0.7914	0.7953	0.8307	0.7712
MPEG2	Deg13	0.7945	0.7735	0.7918	0.7961	0.8322	0.7743
	Deg14	0.7944	0.7733	0.7918	0.7959	0.8320	0.7742
	Deg15	0.7944	0.7733	0.7918	0.7959	0.8320	0.7740
	Deg16	0.7943	0.7733	0.7917	0.7958	0.8319	0.7740

**Tableau IV.5:** Résultats de SSIM (WIFI, IP H.264, MPEG2).



## IV.11.3 Résultats de SSIM\_différ

Normes	Vidéos	pa	rb	st	rh	sf	Tr
Wifi	Deg2	0.7741	0.7619	0.7771	0.7772	0.7746	0.7771
	Deg3	0.7741	0.7627	0.7770	0.7773	0.7746	0.7770
	Deg4	0.7740	0.7621	0.7769	0.7774	0.7744	0.7687
	Deg5	0.7739	0.7622	0.7769	0.7773	0.7744	0.7692
IP	Deg6	0.7740	0.7622	0.7769	0.7773	0.7742	0.7769
	Deg7	0.7740	0.7620	0.7770	0.7772	0.7744	0.7689
	Deg8	0.7741	0.7620	0.7769	0.7773	0.7743	0.7691
H.264	Deg9	0.7741	0.7625	0.7771	0.7775	0.7743	0.7771
	Deg10	0.7742	0.7628	0.7771	0.7775	0.7743	0.7771
	Deg11	0.7741	0.7634	0.7772	0.7776	0.7743	0.7696
	Deg12	0.7743	0.7638	0.7772	0.7777	0.7744	0.7697
MPEG2	Deg13	0.7740	0.7616	0.7769	0.7771	0.7747	0.7769
	Deg14	0.7732	0.7617	0.7769	0.7771	0.7746	0.7700
	Deg15	0.7739	0.7618	0.7771	0.7771	0.7745	0.7690
	Deg16	0.7739	0.7618	0.7772	0.7772	0.7745	0.7692

Tableau IV.6: Résultats de SSIM\_diffé (WIFI, IP H.264, MPEG2).

## IV.9.4 Résultats de PSNR\_TCD

Normes	Vidéos	pa	rb	st	rh	sf	Tr
Wifi	Deg2	32.8267	25.9543	28.1791	28.4653	26.8219	24.7176
	Deg3	30.6664	27.8166	30.3924	28.4300	28.9316	25.1877
	Deg4	29.6956	28.0087	33.8086	32.3960	32.0197	25.6930
	Deg5	25.7526	29.1978	33.9945	31.2731	30.6478	27.8369
IP	Deg6	29.6082	26.7150	30.0217	33.0743	29.0407	23.9535
	Deg7	32.1247	28.0601	34.9307	32.3960	29.9457	28.3745
	Deg8	31.8045	28.0485	35.3207	34.9653	32.4369	28.0257
H.264	Deg9	32.9457	29.0541	34.5924	33.7482	33.6691	28.6534



	Deg10	31.2185	27.9973	34.2857	32.6939	32.1589	27.6795
	Deg11	29.9667	26.6481	34.0056	31.8988	31.5570	27.0057
	Deg12	29.1512	25.9502	33.5057	30.8138	30.9761	26.6531
MPEG2	Deg13	33.8886	29.3656	34.3140	37.0408	36.5236	30.2399
	Deg14	32.8872	27.9982	33.3715	33.5610	34.7018	29.0357
	Deg15	32.0810	27.5916	32.2563	31.7679	33.1242	27.4450
	Deg16	30.7620	27.5916	31.3968	30.3654	32.0339	26.3598

**Tableau IV.7:** Résultats de PSNR\_TCD (WIFI, IP H.264, MPEG2).

#### IV.11.5 Résultats de SROOC de (PSNR, SSIM, SSIM\_différ et PSNR\_TCD)

Les métriques implémentées sur une partie de la base de donnée LIVE ont été testées et comparées avec les résultats subjectives séparément par la méthode SROOC comme montre le tableau suivant :

Métriques Distorsion	PSNR	SSIM	SSIM_différ	PSNR_TCD
Wifi	0.9729	0.9788	0.9828	0.9693
IP	0.9876	0.9899	0.9902	0.9882
H.264	0.9791	0.9762	0.9792	0.9721
MPEG2	0.9727	0.9806	0.9801	0.9728

**Tableau IV.8:** SROOC de (PSNR, SSIM, SSIM\_différ et PSNR\_TCD).

La figure IV.4, figure IV.5, figure IV.6, figure IV.7 et La figure IV.8 représente le graphe de PSNR, SSIM, SSIM\_def, PSNR\_dct et SROCC respectivement.

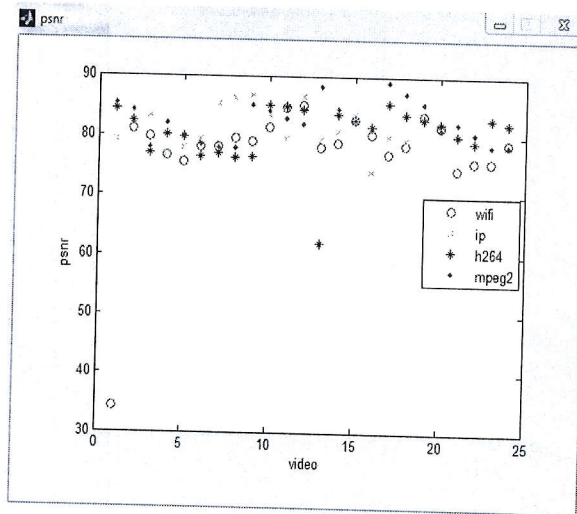


Figure IV.4 : PSNR des vidéos.

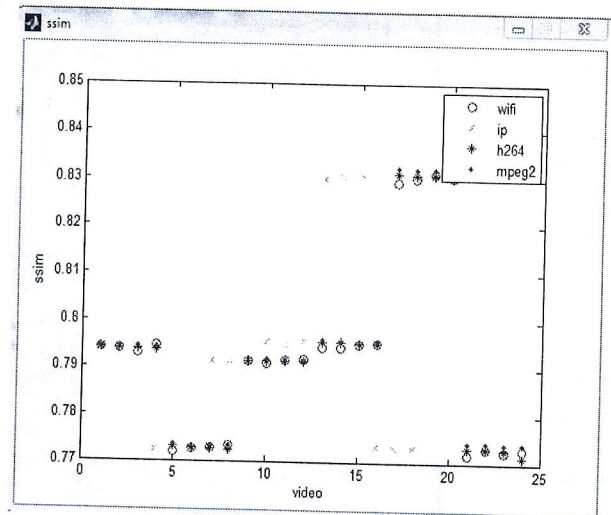


Figure IV.5 : SSIM des vidéos.

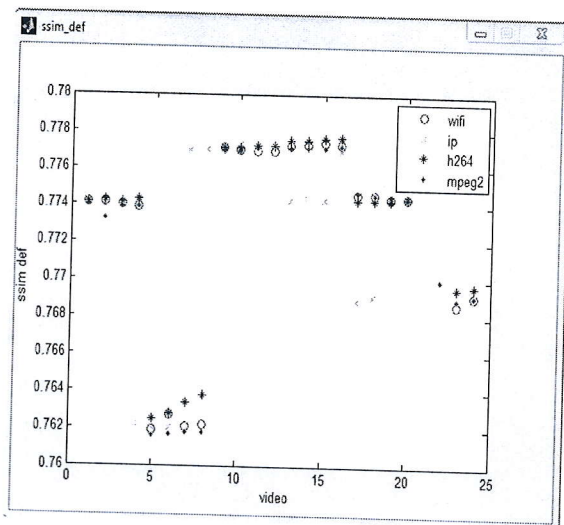


Figure IV.6 : SSIM\_def des vidéos.

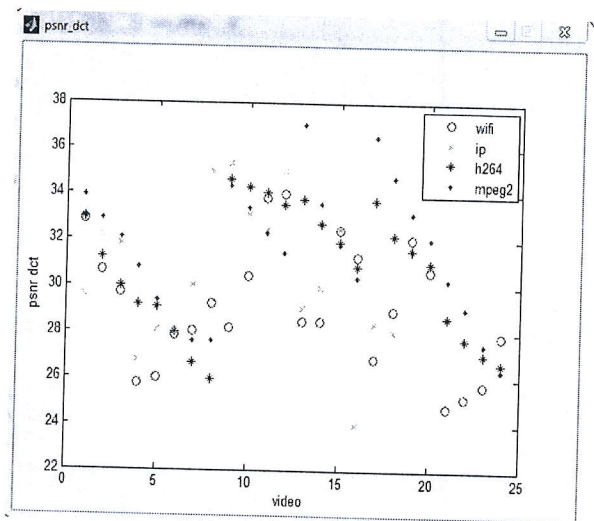


Figure IV.7 : PSNR\_dct des vidéos.

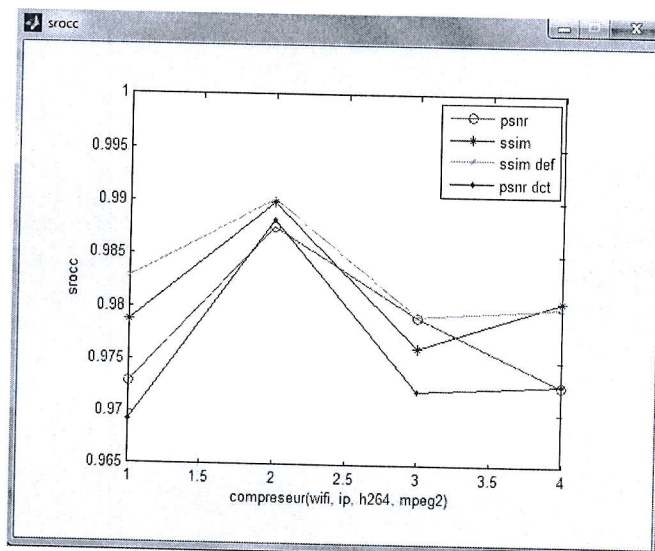


Figure VI.8 : SROOC de (PSNR, SSIM, SSIM\_différ et PSNR\_TCD).

Les résultats obtenus représentés dans la figure VI.2 montrent l'efficacité de la métrique SSIM, car elle est basée sur la similarité structurelle de l'image, tendit que l'efficacité de PSNR est moins que le SSIM.

La métrique SSIM de la différence est la plus performante grâce à l'ajout d'un autre axe au SSIM traditionnel qui est l'aspect temporelle ce qui abouti les meilleures résultats (la plus proche aux résultats subjectives), la métrique la moins efficace parmi ces métriques est le PSNR dans le domaine fréquentiel à cause d'une incompatibilité entre les fréquences passe bas et haut.

#### IV.12 Conclusion

Dans ce chapitre nous avons présenté la réalisation de notre application, qui consiste à une implémentation des métriques objectives dans les deux domaines spatial et fréquentiel, dans le premier domaine nous avons implémenté le PSNR, SSIM et le SSIM de résultat de la différence de chaque deux images successives de la vidéo, bien que le deuxième consiste à calculer le PSNR de la DCT de la différence de la vidéo originale et la vidéo dégradée. Ainsi que la comparaison entre ces métriques par le SROOC. Nous atteindre que le SSIM est plus performantes que le PSNR, d'autre part le SSIM de la différence est mieux que le SSIM.



## *Conclusion générale*

## Conclusion générale

Cette étude traite l'évaluation de la qualité des vidéos numériques en se basant sur les métriques subjective et objective, les métriques subjectives utilisent des sujets humains pour pouvoir évaluer la qualité visuelle, Ces méthodes sont couteuses et prennent beaucoup de temps. C'est la raison de la naissance des métriques objectives, à fin de passer au domaine automatique tout en cherchant à simuler le système visuelle humaine qui est un système très complexe et par conséquence très difficile à simuler.

Les métriques objectives sont classées en trois sous ensembles: avec référence complète (RC), référence réduite (RR) et sans référence (SR), ce classement est basé sur la présence ou l'absence de la vidéo originale utilisée au cours de la phase d'évaluation de la qualité, la métrique avec référence complète utilise la vidéo originale et la vidéo dégradée, dans la métrique référence réduite la vidéo originale est remplacée par un ensemble de caractéristiques, la dernière métrique utilise uniquement la vidéo dégradée. Il existe divers méthodes efficaces pour ces métriques qui nous nous décrivions dans le troisième chapitre (PSNR, SSIM, BLIND, VQM...).

Dans notre travaille en va implémenter quelques métriques objectives, ces métriques sont testées sur une partie de la base de données LIVE, Cette base de données qui contient dix vidéos originales (naturelles non compressées) et quinze vidéos dégradées pour chacune originale donc au totale cent cinquante vidéos, les chercheurs ont choisi quatre normes standards qui sont : MPEG-2 (4 vidéos pour chaque vidéo originale), H.264 (4 vidéos pour chaque vidéo originale), erreur de transmission par IP (3 vidéos pour chacune originale), et la même chose sur un réseau sans files ou wifi (4 vidéos pour une vidéo originale). Toutes ces vidéos ont le format YUV 4:2:0 et une résolution 768x432, avec une durée de 10 secondes, sauf une seule vidéos de 8.68 seconde, la fréquence de ces vidéos prendre les valeurs suivantes 25 image/seconde ou 50 image/seconde. En débutant par le SSIM et le PSNR.

La métrique d'évaluation que nous avons proposés exploite le SSIM dans le domaine spatial en utilisant le SSIM qui calculer pour la différence de chaque deux images successives de la vidéo et ces faits pour introduise l'aspect temporelle au SSIM traditionnel.

L'autre métrique celle qui opère dans le domaine fréquentiel utilise la DCT pour faire le passage de domaine spatial vers le domaine fréquentiel. Cette DCT est calculé pour les résultats de différence de chaque deux images successives de la vidéo, ensuite le PSNR est calculée entres les résultats de l'étape précédente de la vidéo originale et la vidéo dégradée.

L'évaluation de l'efficacité des deux métriques proposées doit passer obligatoirement par une comparaison avec les résultats subjectives (c'est le cas pour toutes les métriques), le SROCC est la méthode de comparaison entre les résultats objectives et subjectives.

Les résultats obtenus montrent l'efficacité de la métrique SSIM, car elle est basée sur la similarité structurelle de l'image, tandis que l'efficacité de PSNR est moins que le SSIM.

La métrique de SSIM de la différence donner les meilleurs résultats grâce à un critère supplémentaire au SSIM appliqués dans la phase précédente, qui est l'aspect temporelle ce qui améliorer les performances, d'autre part l'efficacité la plus faible dans notre environnement est celle de PSNR dans le domaine fréquentiel.



**Résumé**

L'évaluation de la qualité vidéo est d'une très grande importance, puisque elle permet d'évaluer l'efficacité des services offerts, surtout lorsque le dernier maillon de la chaîne est l'observateur humain (qui est vrai dans la plupart des cas). Dans notre travail on essaye d'investir le domaine d'évaluation objective de la qualité des séquences vidéo.

Le premier chapitre présente quelques notions de base sur la vidéo numérique telle que les espaces de représentation couleurs, ainsi que la hiérarchie de composition des séquences vidéo.

Le second chapitre effectue une analyse des traitements les plus répandus que peut subir une vidéo brute, ainsi que les dégradations reliées à chaque traitement.

Le troisième chapitre résume les différentes techniques d'évaluation de la qualité vidéo, ces techniques peuvent être subjectives (effectuées par l'observateur humain), ou objectives (effectuées d'une façon automatique).

Le quatrième chapitre de notre mémoire, est le chapitre dédié à la présentation de l'application, les résultats obtenus et l'analyse de ces résultats. Dans notre application, nous avons implémenté quatre métriques de qualité vidéo, qui sont le PSNR le SSIM le SSIM-Dif, et le PSNR-DCT. Les performances de ces techniques sont testées sur une partie de la base de vidéo LIVE, et leurs corrélations avec les résultats de l'évaluation subjective sont mesurées grâce au SROOC.

**Mots clés :** qualité vidéo, objective, subjective, PSNR, SSIM, DCT.

**Abstract**

The evaluation of the video quality is of great importance, since it allows to evaluate the effectiveness of services, especially when the last link in the chain is the human observer (which is true in most cases). In our work we try to invest the field of objective assessment of the quality of video sequences.

The first chapter presents some basics on digital video such as color representation spaces and the composition hierarchy movies.

The second chapter performs an analysis of the most answered treatments that can undergo a raw video, as well as degradations related to each treatment.

The third chapter summarizes the different technical evaluation of video quality, these techniques can be subjective (performed by the human observer) or objective (performed in an automatic way).

The fourth chapter of our brief is the chapter dedicated to the presentation of the application, the results and analysis of these results. In our application, we implemented four video quality metrics, which are PSNR SSIM SSIM-Diff and PSNR-DCT. The performance of these techniques are tested on some basic video LIVE, and their correlations with the results of the subjective evaluation are measured through SROOC.

**Keywords:** video quality, objective, subjective, PSNR, SSIM, DCT

## Liste de références

---

- [1] [http://www-energie.arch.ucl.ac.be/eclairage/guide\\_confort.htm](http://www-energie.arch.ucl.ac.be/eclairage/guide_confort.htm).
- [2] <http://www.map.toulouse.archi.fr/works/panoformation/imagenum.htm>.
- [3] Sabin André, Documentaliste – Sciences de l’information, 2003, vol. 40, n°6.383
- [4] M.Benabdellah, Outils De Compression Et De Crypto-compression : Application Aux Images Fixes Et Vidéo, Thèse de Doctorat, Université Mohammed v-Agdal, faculté des sciences, rabat, 20 Juin 2007.
- [5] Ahmed Ben Atitallah, —Etude et Implantation d’Algorithmes de Compression d’images dans un Environnement Mixte Matériel et Logiciel, thèse de doctorat en électronique, Université Bordeaux 1, Juillet 2007.
- [6] [www.adobe.com/motion/events/pdfs/dvprimer.pdf](http://www.adobe.com/motion/events/pdfs/dvprimer.pdf)
- [7] Antoine Manzanera, ENSTA/UEI -Traitement Et Reconnaissance d'Images
- [8] L.E. RICHARDSON, video codec design : Developing Image and Video Compression Systems , John Wiley & Sons, Ltd, UK, 2002.
- [9] Yun Q. Shi et Huifang Sun, Image and Video Compression for Multimedia Engineering, Fundamentals, Algorithms, and Standards II, CRC Press, USA, 2000.
- [10] HARIZE Saliha. Etude et implémentation d’un codeur vidéo sur un système embarqué de type FPGA : *Applications aux transformations orthogonales du codeur H264/AVC*. Thèse doctorat, Université Mokhtar-Annaba 2013/2014.
- [11] Paul-Kenji Cahier, (Techniques de compression vidéo des standards MPEG) cours, Université de NICE-SOPHIA ANTIPOLIS.
- [12] Cédric MARIN, Vers une solution réaliste de décodage source-canal conjoint de contenus multimédia, thèse de doctorat en physique, Université Paris-sud 11, 2009.
- [13] Standard MPEG-2: ISO/IEC 13818-2, Information Technology – Generic Coding of Moving Pictures and Associated Audio Information.
- [14] Pascal Nsame, modélisation d'un système de transmission d'images animées sur réseau sans fil,mémoire présenté à l'université du Québec à trois-rivières comme exigence partielle de la maîtrise en physique 1997.
- [15] H.264/MPEG-4 Part 10 or AVC (Advanced Video Coding) - H.264 learning from WikiPedia.
- [16] Mohamed Azni, codage conjoint source canal : Modèles et Application à la transmission vidéo, Thèse de doctorat, Université Abderrahmane MIRA – BEJAIA.2010.
- [17] Nessrine Zeglam. Codec de la vidéo appliqué à la visiophonie sur IP, mémoire de magister,ednstic, le 15Juillet 2008.



- [18] Zeraoulia khaled, transmission vidéo sur des liaisons sans fil, mémoire de magister, université Houari Boumediene, 2005.
- [19] Z, Mammeri, cours de réseaux, université Paul Sabatier (Toulouse III).
- [20] DI Gallo Frédéric, WiFi L'essentiel qu'il faut savoir.2003.
- [21] Deepak Gupta, Preeti Yadvendra "Analysis of android video streaming over Wi-Fi/Ad-hoc Network" International Journal of Advanced Research in Computer Science, vol.-4 issue: 8, May-June 2013 ISSN No. 0976-5697.
- [22] Wassim Ramdan. Amélioration de transmission de contenus vidéo et de données dans les réseaux sans fil, thèse de doctorat. French : Université de Franche-comté. 2011.
- [23] Franco Almada Valdez, Objective video quality assesement concidring frame and display time variation, Diploma thesis ,uiversidade Federal do rio grande do sul. July 2012.
- [24] Alexandre NINASSI. Thèse de Doctorat de l'Université de Nantes de la perception locale des distorsions de codage a l'appréciation globale de la qualité visuelle des images et vidéos. 2009.
- [25] T. Alpert & J.-P. Evain. Evaluation subjective de la qualité : Les méthodes SSCQE et UER-revue technique, 1997.
- [26] <https://www.tbs-sct.gc.ca >pem-mep-fra>.
- [27] Franco Almada Valez, diploma thesis Objectif video quality assesement concidring frame and display time variation, University Fedral Do Rio Grande Do Sul. Porto Alegre, 2012.
- [28] S.Winkler and P.Mohandas. "The Evolution of Video Quality Measurement: From PSNR to Hybrid Metrics", 2008
- [29] Songnan Li, Lawrence Chun-Man Mak, and King Ngi Ngan, Visual Quality Evaluation for Images and Videos. The Chinese University of Hong Kong.
- [30] Fadi Boulos. Transmission d'images et de vidéos sur réseaux à pertes de paquets : mécanismes de protection et optimisation de la qualité perçue. Traitement du signal et de l'image. Université de Nantes, 2010.
- [31] Yubing Wang, Survey of Objective Video Quality Measurements, MA 01748, USA
- [32] Michele A. Saad, Alan C. Bovik blind quality assesement of videos using a model of natural scene statistics and motion coherency, University of Texas at Austin.
- [33] Michele A. Saad, Alan C. Bovik, Fellow, IEEE, and Christophe Charrier, Member,

- IEEE, Blind Prediction of Natural Video Quality, MARCH 2014.
- [34] R. Soundararajan and A. C. Bovik, "Video quality assessment by reduced reference spatio-temporal entropic differencing," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 4, pp. 684–694, Apr. 2013.
- [35] K. Seshadrinathan and A. C. Bovik, "Motion tuned spatio-temporal quality assessment of natural videos," *IEEE Trans. Image Process.*, vol. 19, no. 2, pp. 335–350, Feb. 2010.
- [36] P. V. Vu, C. T. Vu, and D. M. Chandler, "A spatiotemporal most apparent distortion model for video quality assessment," in *Proc. IEEE ICIP*, Sep. 2011.
- [37] Professeurs Quentin Louveaux, Olivier Bruls, et Frédéric Nguyen, Introduction a MATLAB, Cours de 1re année du grade de bachelier en sciences de l'ingénieur, orientation ingénieur civil. Année académique 2008-2009.
- [38] LIVE Video Quality Database, 2009 [Online]. Available [http://live.ece.utexas.edu/research/quality/live\\_video.html](http://live.ece.utexas.edu/research/quality/live_video.html).

