



Codage et compression multimédia

Compte-rendu TP6

Codage vidéo

Louis Jean
Master 1 IMAGINE
Université de Montpellier
N° étudiant : 21914083

26 mars 2024

Table des matières

1	Introduction	2
2	Prédiction spatiale	3
2.1	Choix d'une image couleur	3
2.2	Modes de prédiction spatiale	4
2.2.1	Premier mode : la prédiction verticale	4
2.2.2	Deuxième mode : la prédiction horizontale	5
2.2.3	Troisième mode : la prédiction moyenne	6
2.3	Carte de décision	7
2.4	Différence	8
3	Transformation par DCT entière	9
3.1	Transformation sur chaque sous-macrobloc	9
3.2	Comparaison des histogrammes pré et post transformation	10
3.3	Reconstruction de l'image par DCT entière inverse	11
3.4	Comparaison des histogrammes pré et post reconstruction	12
4	Quantification par QP	13
4.1	Comparaison des histogrammes pré et post quantification	14
4.2	Reconstruction après quantification	15
5	Prédiction temporelle	16
5.1	Choix d'une seconde image	16
5.2	Méthode de prédiction temporelle	18
6	Conclusion	19

1 Introduction

Ce travail pratique explore les principes de la compression vidéo à travers la mise en œuvre de techniques de prédiction spatiale et temporelle, fondamentales pour les algorithmes de compression modernes tels que H.264. À travers l'application de transformations en cosinus discrètes et de stratégies de quantification adaptatives, ce TP vise à approfondir notre compréhension des mécanismes qui permettent de préserver la qualité visuelle de l'image tout en minimisant la taille occupée par les données.

2 Prédiction spatiale

2.1 Choix d'une image couleur

Pour ce TP, j'ai choisi une image tirée du jeu Counter-Strike : Global Offensive, plus précisément d'une finale d'un major (comparable à des championnats du monde), représentant l'une des actions les plus marquantes de la scène e-sportive du jeu. Je l'ai découpée au format 2048×1024 pixels (multiple de 16 pour le découpage en macroblocs), transformée au format ppm, nommée *nikomiss1.ppm*, puis passée dans l'espace YCbCr, car c'est la luminance qui nous intéresse dans ce TP.



Figure 1: *nikomiss1.ppm*



Figure 2: Canal Y de *nikomiss1.ppm*

2.2 Modes de prédiction spatiale

Le but de la prédiction spatiale est de réduire la redondance spatiale au sein d'une image. Elle repose sur le postulat que dans une image, des pixels voisins ont de fortes chances de se ressembler. On va donc découper l'image en macroblocs de taille 16×16 , pour faire des traitements dessus.

2.2.1 Premier mode : la prédiction verticale

Dans ce mode, chaque pixel d'un macrobloc est prédit à partir des pixels situés directement au-dessus de lui. La prédiction verticale est particulièrement efficace pour les régions de l'image où les variations de contenu se font principalement dans la direction horizontale.



Figure 3: Prédiction verticale

2.2.2 Deuxième mode : la prédiction horizontale

La prédiction horizontale, en contraste avec la prédiction verticale, utilise les pixels situés immédiatement à gauche du bloc en cours de codage pour prédire les valeurs des pixels à l'intérieur du bloc. Ce mode de prédiction est basé sur l'hypothèse que les pixels horizontalement adjacents partagent des caractéristiques similaires. La prédiction horizontale s'avère efficace dans les parties de l'image où les variations verticales dominent.



Figure 4: Prédiction horizontale

2.2.3 Troisième mode : la prédition moyenne

La prédition moyenne est un mode de prédition spatiale qui calcule la valeur moyenne des pixels autour du macrobloc à coder et utilise cette valeur moyenne comme prédition pour l'ensemble du bloc. Cette méthode est particulièrement utile dans des régions de l'image où les variations de luminance sont faibles et uniformément distribuées.



Figure 5: Prédiction moyenne

2.3 Carte de décision

À partir de ces prédictions, on peut prendre des décisions pour savoir laquelle utiliser dans un macrobloc donné. Pour cela, on calcule les distances euclidiennes entre un macrobloc prédit par une méthode, et un macrobloc dans l'image originale. Pour chaque macrobloc, on choisit la méthode la plus proche de l'image originale. On peut alors dresser une carte de décision.



Figure 6: Carte de décision

Ici, j'ai mis à 0 les macroblocs pour lesquels la prédiction verticale a été choisie, à 255 les macroblocs pour lesquels la prédiction horizontale a été privilégiée, et à 128 les macroblocs pour lesquels la prédiction moyenne a été retenue.

Avec ceci, on peut sortir l'image prédictive.



Figure 7: Image prédictive

2.4 Différence

En calculant la différence entre l'image originale et l'image prédictive (et en recentrant en y ajoutant 128), voici ce que l'on observe.

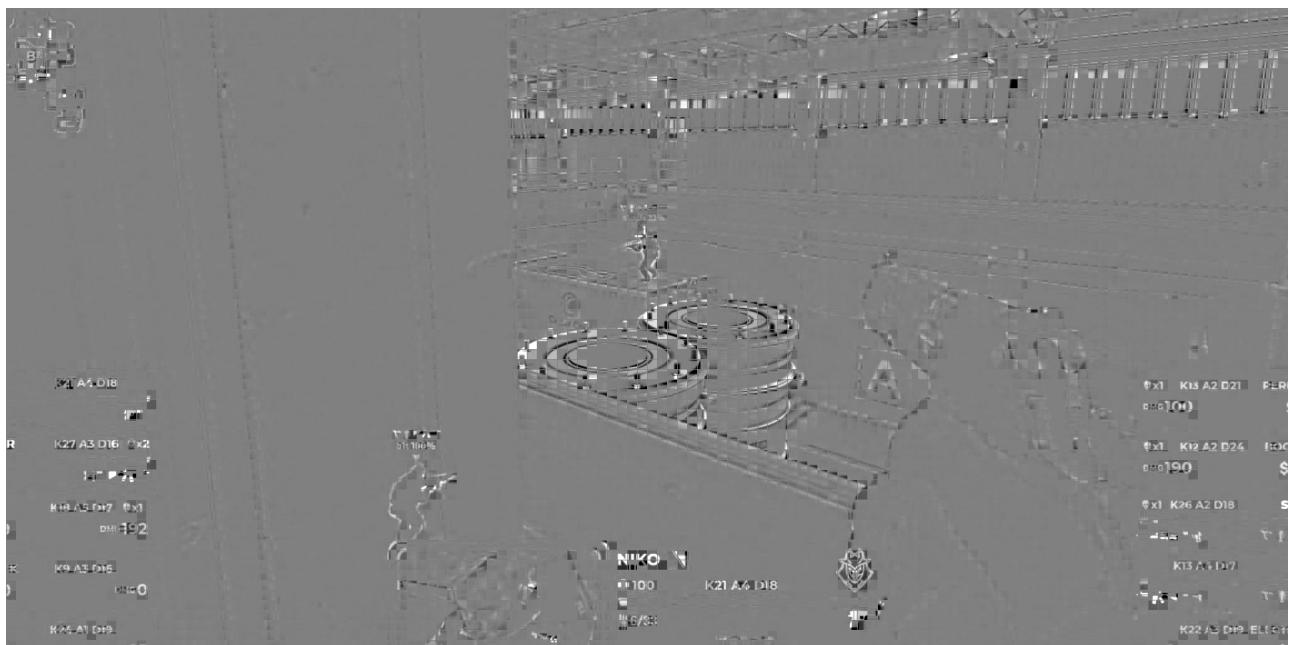


Figure 8: Image de la différence

3 Transformation par DCT entière

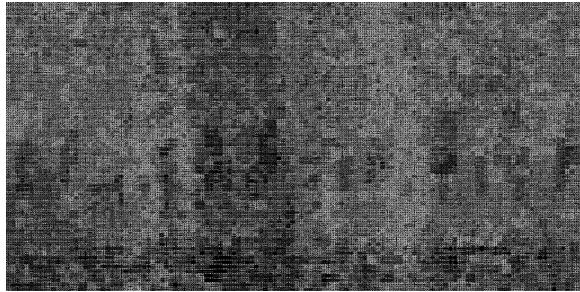
3.1 Transformation sur chaque sous-macrobloc

Au sein d'un macrobloc, on applique la DCT entière à chaque sous-macrobloc de taille 4×4 . Pour ce faire, on utilise l'approximation suivante :

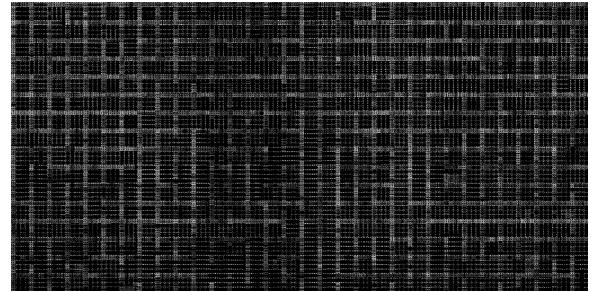
$$\left(\begin{pmatrix} 1 & 1 & 1 & 1 \\ 2 & 1 & -1 & -2 \\ 1 & -1 & -1 & 1 \\ 1 & -2 & 2 & -1 \end{pmatrix} \begin{pmatrix} X \end{pmatrix} \begin{pmatrix} 1 & 2 & 1 & 1 \\ 1 & 1 & -1 & -2 \\ 1 & -1 & -1 & 2 \\ 1 & -2 & 1 & -1 \end{pmatrix} \right) \otimes \begin{pmatrix} a^2 & \frac{ab}{2} & a^2 & \frac{ab}{2} \\ \frac{ab}{2} & \frac{b^2}{4} & \frac{ab}{2} & \frac{b^2}{4} \\ a^2 & \frac{ab}{2} & a^2 & \frac{ab}{2} \\ \frac{ab}{2} & \frac{b^2}{4} & \frac{ab}{2} & \frac{b^2}{4} \end{pmatrix}$$

avec $a = \frac{1}{2}$ et $b = \sqrt{\frac{2}{5}}$ et X le sous-macrobloc considéré.

Voici ce que l'on obtient lorsque l'on applique la DCT entière sur l'image de la luminance et sur l'image prédictive.



(a) Image de la luminance après DCT entière

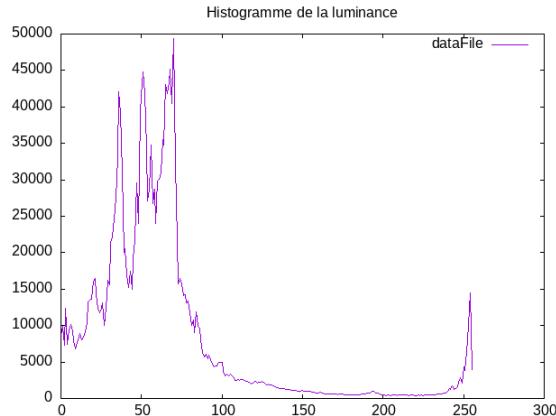


(b) Image prédictive après DCT entière

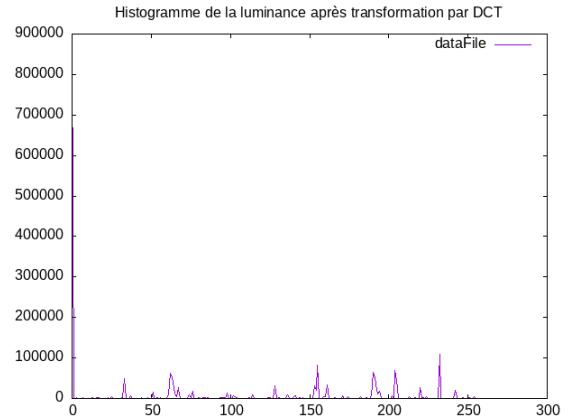
Figure 9: Application de la DCT entière sur l'image de la luminance et sur l'image prédictive

3.2 Comparaison des histogrammes pré et post transformation

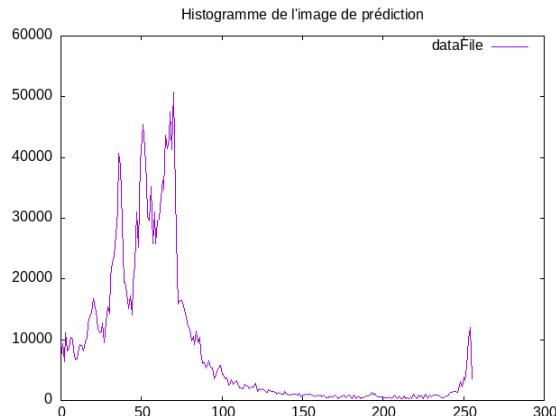
Ce qui est intéressant, c'est de comparer les histogrammes des deux images avant et après transformation.



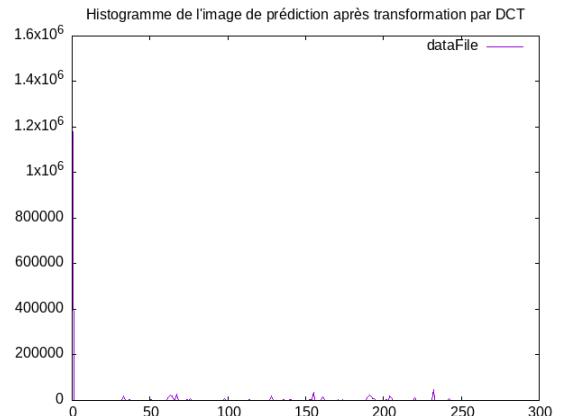
(a) Histogramme de l'image de la luminance avant transformation



(b) Histogramme de l'image de la luminance après transformation



(c) Histogramme de l'image prédictive avant transformation



(d) Histogramme de l'image prédictive après transformation

Figure 10: Comparaison des histogrammes des images de la luminance et de la prédiction avant et après transformation par DCT entière

On remarque que l'application de la DCT concentre la majorité des informations dans beaucoup moins de coefficients différents. La DCT a réduit le redondance des données et à mis l'accent sur les composantes les plus significatives de l'image. Ainsi, peu de coefficients de haute énergie sont nécessaires pour reconstruire l'image, tandis que la plupart des autres coefficients peuvent être soit quantifiés soit omis si l'on accepte la perte de qualité.

3.3 Reconstruction de l'image par DCT entière inverse

Il est possible de reconstruire l'image à partir des coefficients calculés par la DCT, car cette transformation est exactement inversible mathématiquement parlant. La fonction inverse est donnée par :

$$\begin{pmatrix} 1 & 1 & 1 & \frac{1}{2} \\ 1 & \frac{1}{2} & -1 & -1 \\ 1 & -\frac{1}{2} & -1 & 1 \\ 1 & -1 & 1 & -\frac{1}{2} \end{pmatrix} \left(\begin{pmatrix} X \end{pmatrix} \otimes \begin{pmatrix} a^2 & ab & a^2 & ab \\ ab & \frac{b^2}{2} & ab & \frac{b^2}{2} \\ a^2 & ab & a^2 & ab \\ ab & \frac{b^2}{2} & ab & \frac{b^2}{2} \end{pmatrix} \right) \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & \frac{1}{2} & -\frac{1}{2} & -1 \\ 1 & -1 & -1 & 1 \\ \frac{1}{2} & -1 & 1 & -\frac{1}{2} \end{pmatrix}$$

avec $a = \frac{1}{2}$ et $b = \sqrt{\frac{2}{5}}$ et X les coefficients DCT du sous-macroBloc considéré.

Après application de l'IDCT entière, on retombe exactement sur les mêmes images qu'avant application de la DCT entière.



(a) Image de la luminance après DCT entière puis IDCT entière



(b) Image prédictive après DCT entière puis IDCT entière

Figure 11: Application de la DCT entière puis de l'IDCT entière sur l'image de la luminance et sur l'image prédictive

3.4 Comparaison des histogrammes pré et post reconstruction

Au niveau des histogrammes, on constate qu'ils sont très très identiques, et je pense que les seules différences (minimes) sont dues à des erreurs de précision dans les calculs flottants.

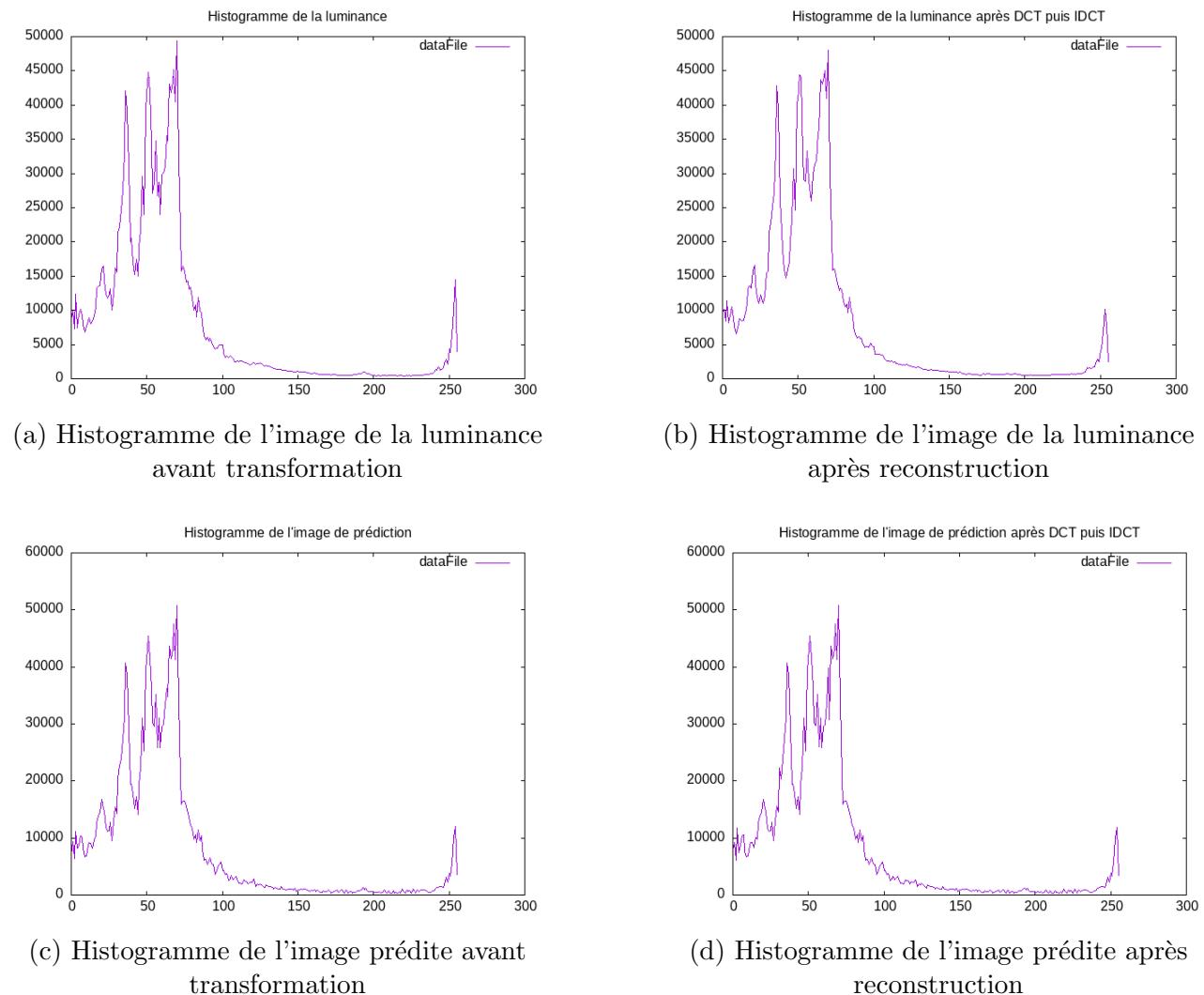


Figure 12: Comparaison des histogrammes des images de la luminance et de la prédiction avant et après transformation par DCT entière puis reconstruction par IDCT entière

4 Quantification par QP

Le processus de quantification des coefficients de la DCT que nous utilisons ici repose sur l'utilisation d'un paramètre de quantification, connu sous l'acronyme QP, qui permet de contrôler le niveau de compression d'une image ou d'une vidéo. Le QP est un facteur critique dans le compromis entre la qualité de l'image et l'efficacité de la compression. Une valeur de QP faible entraîne une compression légère, et inversement, une valeur de QP élevée augmente le niveau de compression, ce qui réduit la taille du fichier mais peut également entraîner une perte de qualité plus marquée. QP peut prendre 52 valeurs, comprises entre 0 et 51. Pour utiliser QP, on définit un pas de quantification Qstep :

$$Qstep(0) = 0.625$$
$$Qstep(QP) = Qstep(QP - 1) \cdot 2^{\frac{1}{6}}$$

La quantification et la déquantification sont réalisées de la manière suivante :

- On divise chaque coefficient de la matrice de scaling de la DCT entière par Qstep.
- Ensuite, on applique la DCT entière sur les sous-macroblocs.
- Une fois les coefficients obtenus, on les arrondit à l'entier le plus proche. C'est **ici** que se produit la perte d'information.
- Pour reconstruire l'image, on multiplie chaque coefficient de la matrice de scaling de l'IDCT entière par Qstep, puis on applique l'IDCT entière sur chaque sous-macrobloc.

4.1 Comparaison des histogrammes pré et post quantification

En ce qui concerne les histogrammes, on remarque que la quantification réduit énormément la précision des coefficients, ce qui facilite grandement la compression.

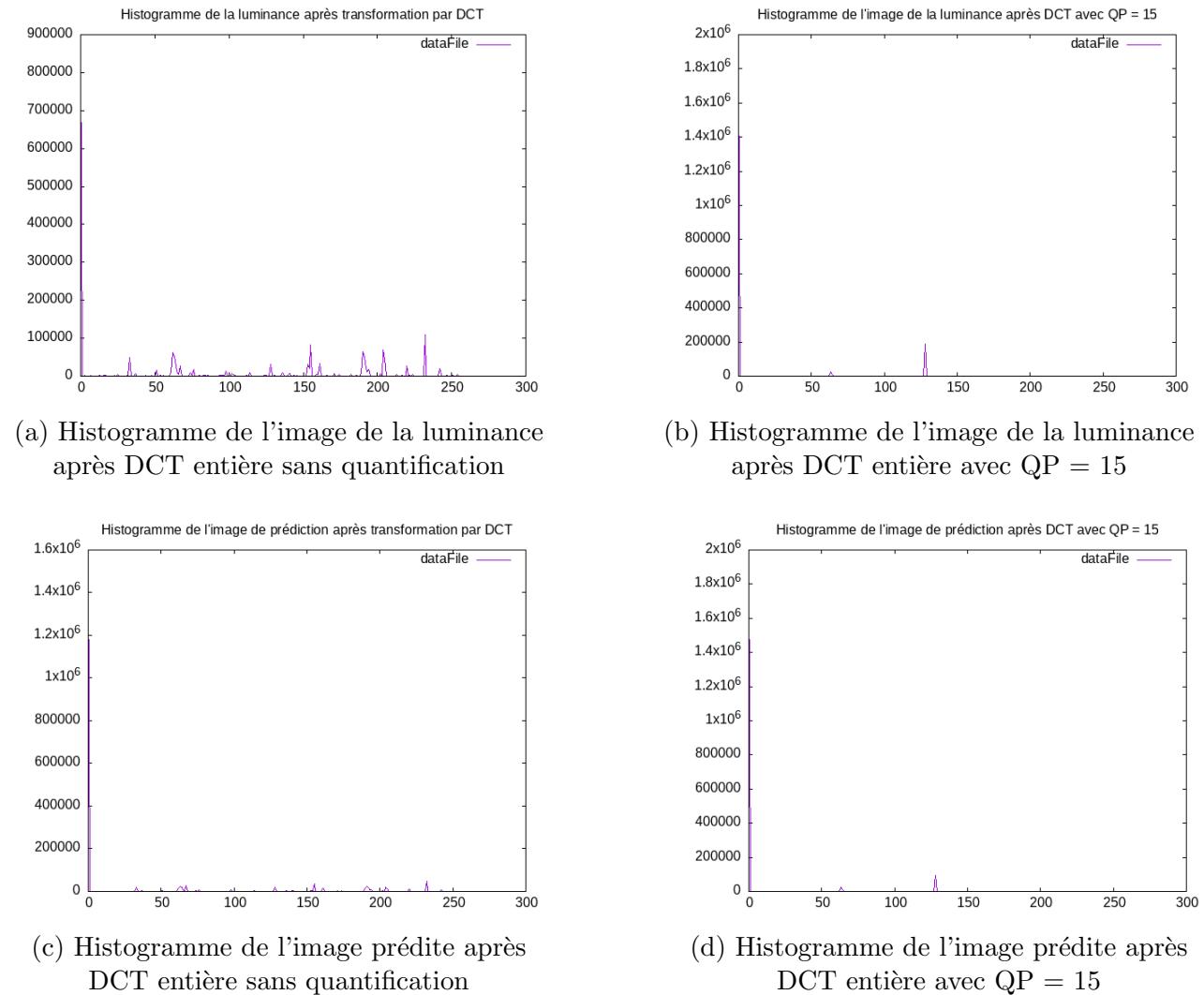


Figure 13: Comparaison des histogrammes des images de la luminance et de la prédiction après transformation par DCT sans et avec quantification

4.2 Reconstruction après quantification

Pour la reconstruction, quand $QP = 5$, on maintient une qualité très acceptable. On commence à observer quelques artefacts (quoiqu'acceptables) à partir de $QP = 15$. Au-delà, la qualité semble devenir trop médiocre (voir défauts sur les fenêtres pour $QP = 25$).



(a) $QP = 5$



(b) $QP = 15$



(c) $QP = 25$



(d) $QP = 51$

Figure 14: Reconstruction de l'image prédictive après quantification, pour plusieurs valeurs de QP différentes

5 Prédiction temporelle

5.1 Choix d'une seconde image

Pour cette partie, j'ai du choisir une image assez proche de *nikomiss1.ppm*. C'est donc tout naturellement que j'ai choisi une image qui se passe temporellement quelques frames après cette dernière. J'y ai appliqué les mêmes traitements (redimensionnement, conversion au format *ppm* et passage en espace YCbCr) et l'ai nommée *nikomiss2.ppm*.



Figure 15: *nikomiss2.ppm*



Figure 16: Canal Y de *nikomiss2.ppm*



Figure 17: Image prédictive depuis *nikomiss2.ppm*

5.2 Méthode de prédiction temporelle

La prédiction temporelle joue un rôle crucial dans la compression vidéo. Elle exploite la redondance entre les frames consécutives, reposant sur l'hypothèse que la différence des valeurs des pixels entre deux frames qui se suivent est estimable.

Pour chaque macrobloc de la seconde image, il faut examiner non seulement le macrobloc correspondant dans la première image mais aussi d'autres macroblocs à proximité dans un rayon de k pixels. Ce processus implique la recherche du macrobloc le plus similaire, qui sert alors de prédicteur. Le meilleur candidat est celui qui minimise la différence avec le bloc actuel, ici évaluée à l'aide de la somme des différences absolues (SAD). En optimisant la sélection des macroblocs prédicteurs, le but est de réduire la quantité d'informations nécessaires pour représenter la vidéo.



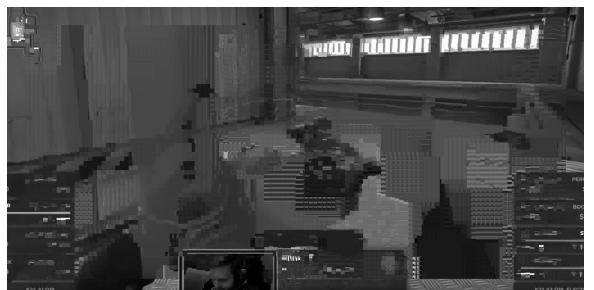
(a) $k = 4$



(b) $k = 16$



(c) $k = 32$



(d) $k = 64$

Figure 18: Prédiction temporelle de *nikomiss2.ppm* à partir des macroblocs prédicts de *nikomiss1.ppm*, pour plusieurs valeurs de k différentes

Je ne suis pas certain si c'est parce que mes images sont trop éloignées, mais je n'obtiens pas de résultat satisfaisant, à moins de choisir une grande valeur de k . Le problème est que plus k est grand, plus les calculs sont longs. **N.B** : Pour les bords, je n'ai pas réussi à trouver une solution, et ai donc recopié les valeurs de l'image d'origine.

6 Conclusion

En conclusion, le travail effectué dans ce TP a permis de mettre en lumière la complexité et l'efficacité des techniques de compression qui sont au cœur du traitement des vidéos. En comprenant la prédition spatiale et temporelle, j'ai pu constater comment la compression exploite la redondance dans les images pour réduire la quantité de données nécessaires à leur représentation. J'ai pris du plaisir à le réaliser même si je n'ai pas fait l'entièreté du sujet.

Je tiens à m'excuser pour tout retard dans la soumission de ce travail.

Merci pour le temps et l'attention que vous avez consacrés à la lecture de ce compte-rendu.