

Mesure objective de la perception des effets de blocs dans le cas de séquence d'images codées par transformée

Christian Derviaux, François-Xavier Coudoux, Marc Gazalet, Laurent Grimbert et Francis Haine

Université de Valenciennes, Institut d'Electronique et de Micro-électronique du Nord
Département OAE, UMR CNRS 9929
Le Mont Houy, BP 311
F-59304 Valenciennes Cedex, France

RÉSUMÉ

L'un des inconvénients majeurs des techniques de codage par Transformée en Cosinus Discrète (TCD) est l'apparition de discontinuités aux frontières entre les blocs adjacents, dues à la quantification grossière des coefficients transformés. Un modèle de perception des effets de blocs est défini qui prédit l'importance de cette dégradation dans chaque image de la séquence vidéo. Celui-ci intègre les propriétés essentielles de la vision humaine telles que la sensibilité spatio-temporelle au contraste et les phénomènes de masquage spatial et temporel. Cette mesure est ensuite utilisée dans un post-traitement adaptatif pour la réduction des effets de blocs dans les images. Les résultats obtenus montrent que les images corrigées présentent une amélioration sensible de la qualité visuelle.

ABSTRACT

One of the main drawbacks of the Discrete Cosine Transform (DCT) is visible block boundaries due to coarse quantization of the transform coefficients. A visual model is then defined for predicting the block edge visibility across each picture of the coded sequence. This model incorporates essential properties of human vision such as contrast spatio-temporal sensitivity, as well as spatial and temporal masking phenomenon. Then an adaptive postprocessing algorithm, for the reduction of blocking effect in video coded sequence, is developed based on this distortion measure. Experimental results show that the postfiltering operation yields significant results with enhanced visual quality.

1 Introduction

Les méthodes de codage avec perte basées sur la Transformée en Cosinus Discrète (DCT) constituent la base de la plupart des normes actuelles de compression d'images numériques telles que JPEG, MPEG ou H26x. Pour des taux de compression élevés, des discontinuités apparaissent aux frontières entre les blocs. Ce phénomène, couramment appelé effet de blocs, est particulièrement gênant et détériore gravement la qualité visuelle de l'image affichée.

Il est reconnu que les critères mathématiques, tels que le SNR et la MSE, sont généralement peu corrélés avec l'appréciation du résultat final par un observateur humain qui constitue le dernier maillon de la chaîne de réception. La perception des défauts dans une image est un phénomène complexe, qui dépend de plusieurs variables telles que le contenu de l'image et le type de la distorsion, que de tels critères ne prennent pas en compte. Les tests subjectifs sont, quant à eux, longs et coûteux à mettre en oeuvre. L'idéal serait de disposer d'un critère mathématique, qui traduise la qualité visuelle d'une image. Plusieurs travaux ont porté sur le développement d'une telle mesure de qualité, basée généralement sur un modèle plus ou moins complexe du système visuel humain [1]. En particulier, des mesures de la distorsion introduite par les effets de blocs ont été proposées

[2]-[3] dans le cas du codage d'images fixes suivant l'algorithme JPEG.

Dans ce papier, nous proposons une extension de ces travaux pour définir un nouveau critère de la perception des artefacts de blocs dans le cas de séquences d'images codées par transformée. Nous commençons par décrire les différentes caractéristiques mises en oeuvre dans notre modèle, avant d'aborder le problème de l'amélioration de la qualité des images, où nous proposons un algorithme adaptatif de post-traitement. Celui-ci permet d'éliminer les effets de blocs sans modifier les standards existants ni introduire de flou excessif dans la séquence reconstruite.

2 Description du modèle de perception

L'effet de blocs se présente comme un motif géométrique fortement corrélé horizontalement et verticalement, dont la position est connue *a priori* dans l'image.

La visibilité locale dépend de l'importance du défaut détecté, de ses variations spatio-temporelles, et également des caractéristiques non linéaires de la vision humaine. Notre modèle permet la prédiction de ces faux contours et comprend différentes étapes (Fig. 1). Nous déterminons, pour chaque frontière de blocs, si l'artefact est visible ou non.

La mesure de distorsion est détaillée dans le cas d'un artefact vertical ; elle peut être facilement étendue aux

contours horizontaux, puisque les effets de blocs et verticaux correspondent uniquement à des faux contours horizontaux

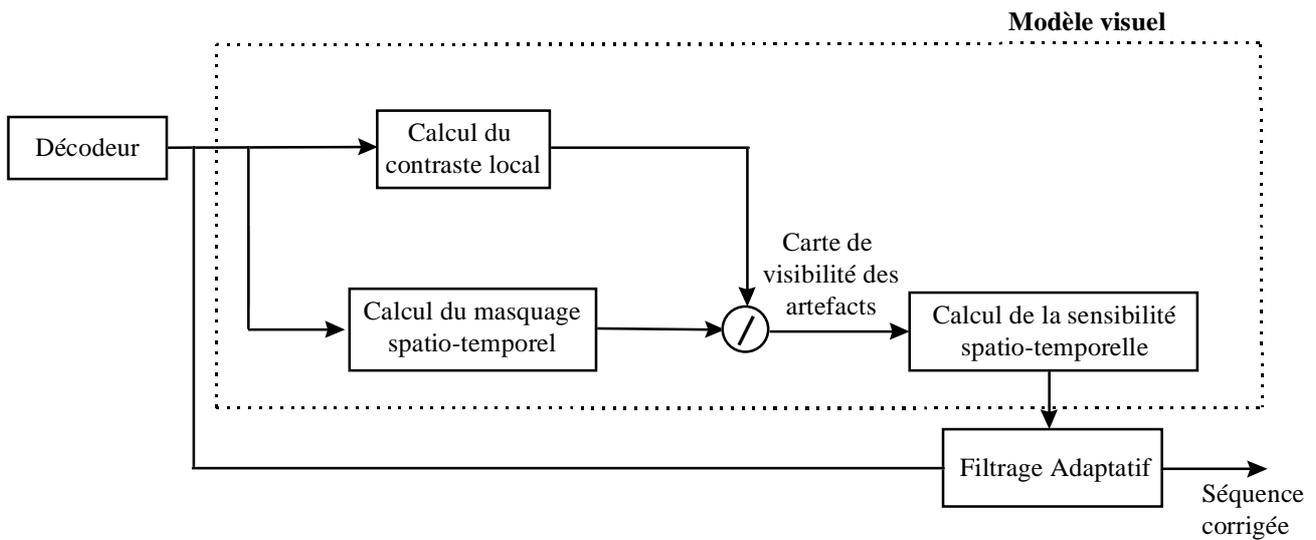


Figure 1 : Schéma du modèle visuel

Pour permettre de quantifier l'importance de chaque phénomène sur la visibilité des effets de blocs, la mesure développée s'appuie sur des résultats obtenus au cours de tests subjectifs. La sensibilité visuelle est obtenue en mesurant le seuil de visibilité, à un artefact modélisé, dans différentes configurations expérimentales. Ces tests sont basés sur une procédure de test à choix forcé comme décrite dans [4]. Le sujet voit deux images simultanément. L'artefact est superposé sur l'une d'elles. Une fois l'artefact détecté, le sujet répond en indiquant dans quelle image l'artefact est présent. L'expérimentation est répétée pour mesurer la sensibilité à des artefacts verticaux en fonction de l'amplitude de l'artefact, de la luminance et des fréquences sur lequel il est superposé.

2.1 Calcul du contraste local

L'œil est sensible aux variations d'intensité lumineuse à l'intérieur de l'image. En particulier, la visibilité d'un stimulus visuel (ici un artefact) diminue avec l'accroissement de la luminance de fond en accord avec la loi de Weber. Les tests montrent que la visibilité de ce défaut est proportionnelle à son amplitude et à la luminance sur laquelle il est superposé (Fig. 2).

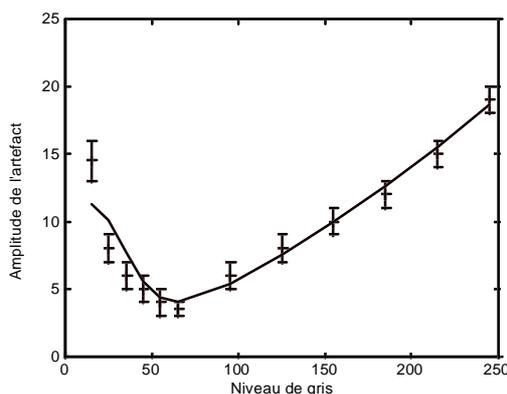


Figure 2 : Sensibilité à un artefact vertical en fonction de la luminance de fond

Notre premier traitement consiste à isoler les artefacts de blocs par un filtrage passe haut et à calculer le contraste local de l'artefact de bloc. Ce contraste local noté ΔC est calculé pour chaque artefact à partir des valeurs de luminance extrapolés de part et d'autre de la frontière inter-blocs. Celui-ci est modélisé, conformément aux résultats des tests subjectifs, par le rapport de l'amplitude de l'artefact détecté dans l'image sur la luminance moyenne :

$$\Delta C(i,k) = \frac{\frac{1}{N} \sum_{j=0}^N \left[\left(\frac{3}{2} x_{l-1,j}^\gamma - \frac{1}{2} x_{l-2,j}^\gamma \right) - \left(\frac{3}{2} x_{l,j}^\gamma - \frac{1}{2} x_{l+1,j}^\gamma \right) \right]}{\frac{1}{2} (L_{moy}(i,k) + L_{moy}(i+1,k))} \quad (1)$$

où x représente la valeur des pixels dans l'image. Ce calcul tient compte de la précorrection de Gamma, qui relie la valeur numérique du pixel à la luminance effectivement affichée sur l'écran, et de la luminance moyenne L_{moy} des deux blocs voisins sur lesquels l'artefact est superposé.

2.2 Le phénomène de masquage

La visibilité d'un contour sera affectée par les fréquences présentes dans son voisinage immédiat. Ce phénomène, appelé masquage, se traduit par la diminution de la sensibilité de l'oeil du fait de l'activité de son entourage. Typiquement, des séquences peuvent contenir des zones uniformes, mais aussi contenir des détails et des mouvements qui affectent la détection des défauts. L'effet de masquage est maximal [5] quand le stimulus et le masque sont accordés (en termes d'orientation et de fréquences spatiale et temporelle). Afin de prendre en compte le phénomène de masquage nous déterminons deux termes d'activité : l'un spatial et l'autre temporel.

Les expériences montrent que les fréquences horizontales masquent plus un contour vertical que les fréquences verticales (Fig. 3). Nous calculons un index d'activité en

fonction du contenu fréquentiel de chacun des deux blocs situés de part et d'autre de chaque frontière inter-blocs :

$$Act_{spat}(i, k) = \sum_{u=0}^7 \sum_{v=0}^7 W(u, v) X_{i, k}(u, v) + \sum_{u=0}^7 \sum_{v=0}^7 W(u, v) X_{i+1, k}(u, v) \quad (2)$$

Dans l'équation (2), u et v désignent les fréquences spatiales, $X(u, v)$ les coefficients du bloc transformé. Les $W(u, v)$ sont des facteurs de pondération visuelle obtenus à partir des résultats des courbes de masquage pour différentes fréquences et orientations.

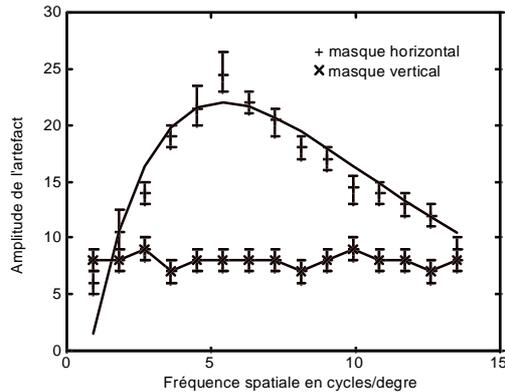


Figure 3 : Résultats du masquage d'un artefact vertical

Dans le cadre de séquence d'images, il convient également de tenir compte du phénomène de masquage temporel pour la mesure de qualité. S'il y a du mouvement dans une scène, les erreurs seront moins perçues. Pour détecter les variations temporelles entre deux images successives, nous utilisons la différence entre celles-ci. Et nous supposons que le masquage temporel est proportionnel à l'amplitude de l'activité. Celle-ci est décrite par :

$$Act_{temp}(i, k) = \sum_{l=1}^{16} \sum_{j=1}^8 |x(l, j, t) - x(l, j, t-1)| \quad (3)$$

2.3 La sensibilité spatio-temporelle

Nous obtenons donc, pour chaque image de la séquence, une carte de détection. Celle-ci s'apparente à une mire qui évolue aussi bien spatialement que temporellement. Elle est fonction de la détection ou non d'un artefact suivant qu'il est masqué ou pas. La sensibilité au contraste dépend à la fois de la fréquence spatiale et de la fréquence temporelle du stimulus [6]. Nous calculons pour chaque artefact les fréquences spatiale et temporelle associées :

La visibilité d'un artefact aura d'autant plus d'importance qu'il y aura d'autres artefacts présents dans le voisinage. Pour la fréquence spatiale, le calcul de la transformée de Fourier est réalisé sur une fenêtre gaussienne qui prend en compte la sensibilité fovéale de l'œil. Cette sensibilité est maximale sur 2° d'angle visuel et décroît avec l'excentricité. A une distance de $6H$, 8 blocs sont vus sous un angle de 2° . Le calcul réalisé, nous cherchons la valeur prédominante qui représente la fréquence spatiale de répétition de l'artefact (F_{spat}).

La répétition locale de l'artefact d'une image à l'autre est une manifestation de sa évolution temporelle. Nous procédons de la même façon que précédemment, une transformée de Fourier est faite sur plusieurs cartes de visibilité afin d'obtenir les variations temporelles de l'artefact. Ensuite, nous évaluons la principale fréquence temporelle d'apparition de l'artefact (F_{temp}).

Les deux valeurs obtenues, nous cherchons dans la courbe de sensibilité spatio-temporelle au contraste, modélisée par Kelly [7], un terme de pondération locale $S(i, k)$ de chaque artefact (Fig. 4).

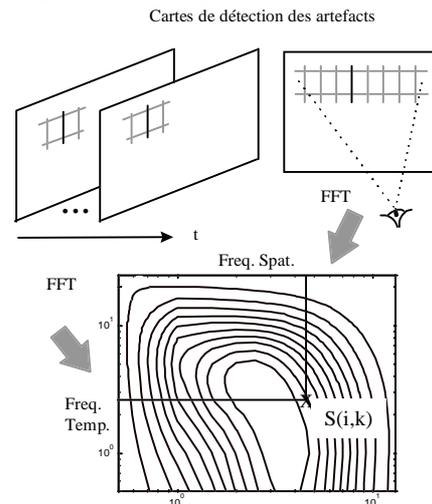


Figure 4 : Calcul de la sensibilité spatio-temporelle

2.4 La mesure de visibilité

Finalement, nous définissons, pour chaque artefact de bloc présent dans l'image, un paramètre de visibilité. Celui-ci est une fonction décroissante des fonctions de masquage. En première approximation, nous supposons que le masquage est séparable. La mesure de distorsion que nous proposons est basée sur une loi simple :

$$V(i, k) = \frac{\Delta C(i, k)}{1 + (\alpha Act_{spat}(i, k) + \beta Act_{temp}(i, k))^\mu} \mu S(i, k) \quad (4)$$

Actuellement, nous comparons les mesures de qualité calculées par notre méthode avec des notes obtenues par des expériences subjectives au CCETT afin de préciser les paramètres (α , β et μ) de notre modèle.

3 Application au post-traitement de séquence d'images

Le critère présenté dans le paragraphe précédent permet de mesurer l'importance perceptuelle des effets de blocs. Une utilisation intéressante d'une telle mesure est la réalisation d'un traitement adaptatif des images à la réception. Un précédent papier [8] a montré l'intérêt d'une telle approche en post-traitement d'images.

Les artefacts de blocs sont d'abord détectés, et leurs visibilités sont calculées conformément à la méthode décrite précédemment. Afin d'éliminer les distorsions de blocs sans

modifier le contenu de l'image originale, en évitant notamment d'introduire un flou excessif dans l'image, il est important de faire varier l'opération de filtrage proportionnellement à la visibilité de la dégradation.

Notre approche a donc consisté à employer un filtre dont la réponse varie en fonction de la mesure de visibilité des artefacts de blocs ; le filtrage est d'autant plus important que la distorsion est gênante (Fig. 5).

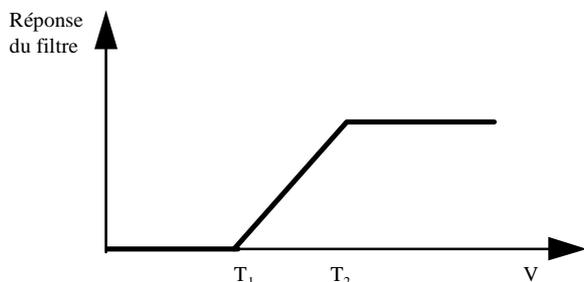


Figure 5: Relation non linéaire entre la réponse du filtre et la visibilité des défauts

Les tests de simulation ont été effectués sur plusieurs séquences codées par l'algorithme MPEG1. Les séquences ont été codées à différents débits (1.15, 0.75 et 0.5 Mbits/s). Plusieurs observateurs ont participé à un test informel sur diverses séquences. Les scènes post-traitées ont été jugées de meilleure qualité dans la majorité des cas.

La Figure 6 représente une image codée de la séquence claire. La même image par lissage adaptatif est montrée à la Figure 7. On constate que les effets de blocs sont éliminés de manière efficace sur les parties les plus dégradées situées au centre de l'image. Toutefois, on peut noter que les zones détaillées (lèvres, yeux) ne sont pas rendues floues, et que la finesse de l'image a été convenablement préservée. Cette caractéristique de la méthode de post-traitement est en particulier remarquable lors de la visualisation de la séquence complète.

4 Conclusion

Nous avons proposé un algorithme pour la réduction des effets de blocs dans le cas de séquences codées par transformée. Celui-ci s'appuie sur un calcul local de l'importance du défaut dans l'image. Les résultats obtenus montrent que le traitement proposé permet de corriger localement les défauts les plus visibles, sans introduire de flou excessif sur les images. Il permet d'améliorer sensiblement la qualité visuelle des images lors de l'affichage.

D'autres phénomènes visuels sont encore à considérer dans notre modèle. Certaines caractéristiques, notamment sur la vision des couleurs, devraient être prises en compte. La suite de nos travaux consiste à introduire les propriétés chromatiques de la vision dans notre critère de visibilité. Notre mesure permet d'évaluer la qualité des images mais peut tout aussi bien permettre une optimisation du codage à l'émission.

Remerciements : nous remercions le CCETT pour les tests subjectifs réalisés sur les séquences d'images codées.

5 Références

- [1] C. J. van den Branden Lambrecht, "Perceptual models and architecture for video coding applications," PhD thesis, EPFL Lausanne, 1996.
- [2] S. A. Karunasekera and N. G. Kingsbury, "A distortion measure for blocking artifacts in images based on human visual sensitivity," in IEEE Transactions on Image Processing, vol. 4, N° 6, pp. 713-724, June 95.
- [3] F. X. Coudoux, "Evaluation de la visibilité des effets de blocs dans les images codées par transformée - Application à l'amélioration d'images," Thèse de doctorat, Janvier 94.
- [4] F. Kretz, J. L. Boudeville et P. Sallio, "Critères objectifs de qualité et exemple d'optimisation d'un codage numériques," Revue de radiodiffusion-télévision, N° 52, pp. 20-25, 1978.
- [5]: G. Tziritas and C. Labit, "Motion analysis for image sequence coding," in series: Advances in Image Communication, Chapter 2, pp. 37-68, 1994.
- [6]: J. G. Robson, "Spatial and temporal contrast sensitivity functions of the visual system," Journal of Optical Society of America, vol. 56, pp. 1141-1142, August 66.
- [7]: D. H. Kelly, "Motion and vision. II. Stabilized spatio-temporal threshold surface," Journal of Optical Society of America, vol. 69, N° 10, pp. 1340-1349, October 79.
- [8]: C. Derviaux, F. X. Coudoux, M. Gazalet and P. Corlay, "Blocking artifact reduction of DCT coded image sequences using a visually adaptive postprocessing," in proceedings IEEE ICIP-96, vol II, pp. 5-8, Sept. 96.



Figure 6 : Image reconstruite non corrigée

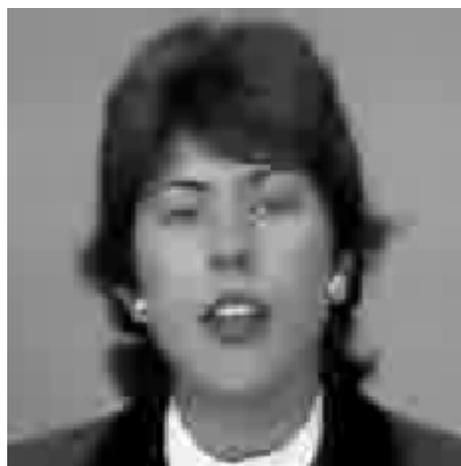


Figure 7 : Image après correction adaptative