

# Progrès récents dans le domaine de la compression vidéo

## 1.0 Introduction

**L**a compression vidéo est nécessaire pour la transmission des données vidéo numériques dans les réseaux à bande passante limitée d'aujourd'hui ainsi que pour les applications où le stockage constitue une limite. Imaginez, pour transmettre des données vidéo numériques à 24 bits par pixel brut échantillonné selon une résolution spatiale de 720 par 480 et une résolution temporelle de 30 images par seconde, il faudrait un débit binaire de 248 Mbps et plus!

La compression des données vidéo numériques sans dégradation significative de la qualité est possible lorsque les séquences vidéo affichent un degré élevé: 1) de redondance spatiale - corrélation entre les pixels voisins; 2) de redondance spectrale - corrélation entre les composants de la couleur; 3) de redondance temporelle - corrélation entre les images vidéo; 4) de redondance psycho-visuelle - propriétés de la vue humaine.

L'atténuation de la redondance temporelle dans le signal vidéo représente un pourcentage important de la compression obtenue. En conséquence, les techniques avancées de codage du signal résiduel ne permettent habituellement que peu de compression supplémentaire comparativement aux techniques classiques et, souvent, l'augmentation de la complexité ne justifie pas cette amélioration. Aussi, plutôt que d'améliorer le codage du signal résiduel lui-même, la plupart des techniques efficaces essayent de réduire le signal résiduel à coder en améliorant la prévision du mouvement dans une séquence vidéo.

Au cours des dernières années, l'intérêt pour le multimédia a entraîné l'investissement d'une somme importante d'efforts de recherche sur le codage vidéo dans les universités et l'industrie, efforts qui se sont traduits par l'élaboration de plusieurs normes (UIT-T H.261 [1], H.263 [2], ISO/CEI MPEG-1 [3], MPEG-2 [4] et MPEG-4 [5]). Ces normes visent un vaste éventail d'applications aux exigences variées quant au débit binaire, à la qualité d'image, à la complexité, à la résilience à l'erreur et au délai ainsi qu'à l'amélioration des rapports de compression.

Le présent article décrit en premier lieu le système de codage vidéo hybride par compensation du mouvement basée sur des blocs et par transformée qui est employé dans toutes les normes vidéo d'aujourd'hui. L'article donne aussi une brève description de chaque composant d'un tel système. On traite ensuite de la recommandation pour le codage vidéo H.26L [6]. Finalement, il est question des outils de codage de la recommandation H.26L, qui diffère passablement des normes de codage vidéo antérieures, puis on compare le rendement de cette recommandation à celui de normes antérieures (MPEG-2, MPEG-4, etc.).

## 2.0 Codage vidéo par compensation du mouvement basée sur des blocs et par transformée

Dans un codeur vidéo hybride à compensation du mouvement et à transformée, la prévision compensée du mouvement réduit en premier lieu les redondances temporelles. Un codage par transformée est ensuite appliqué à l'image différentielle correspondante pour réduire les redondances spatiales. Dans le cas des sources fortement corrélées, comme les images naturelles, la capacité de compression de la transformée en cosinus discrète, ou TCD (Discrete Cosine Transform, ou DCT), est très proche de celle de la transformée optimale, à savoir la transformée de Karhunen-Loève, ou TKL (Karhunen-Loève Transform, ou KLT). De plus, la TCD, à la différence de la TKL, n'est pas dépendante des données. Cela explique pourquoi la TCD est la transformée la plus populaire pour le codage d'images, comme le montre son utilisation dans la norme internationale pour les images fixes, à savoir la norme JPEG. En outre, bien que les images différentielles de prévision compensée du mouvement affichent une corrélation médiocre, la TCD demeure l'outil le plus populaire pour de telles images. En fait, la TCD

par Guy Côté et Lowell Winger,  
VideoLocus, Waterloo, Ontario

## Sommaire

La compression vidéo fait partie intégrante de plusieurs applications multimédias d'aujourd'hui. Certaines applications, par exemple les lecteurs DVD, la télévision numérique, la télévision par satellite, la transmission vidéo par l'Internet, la vidéo conférence, la sécurité vidéo et les caméras numériques, sont limitées par la capacité mémoire ou de bande passante pour la transmission, ce qui stimule la demande pour des rapports de compression vidéo plus élevés. Pour tenter de résoudre ce problème, on a adopté plusieurs normes de codage vidéo durant la dernière décennie. Le présent article traite en premier lieu de la structure générale et des composants d'un système de codage vidéo standard. L'article porte également sur la nouvelle recommandation H.26L, qui est présentement élaborée par les groupes de normalisation UIT et ISO, et souligne les principales différences avec les normes antérieures (MPEG-2, MPEG-4, H.263.20, etc.).

## Abstract

Video compression is a critical component of many multimedia applications available today. For applications such as DVD, digital television broadcasting, Satellite television, Internet video streaming, video conferencing, video security, and digital camcorders, limited transmission bandwidth or storage capacity stresses the demand for higher video compression ratios. To address these different scenarios, many video compression standards have been ratified over the past decade. This article first discusses the general structure and components of a standards-based video coding system. An overview of the emerging video coding standard H.26L, currently being developed jointly by the ITU and ISO standard bodies, is then presented, highlighting key differences with its predecessor standards, such as MPEG-2, MPEG-4, and H.263.

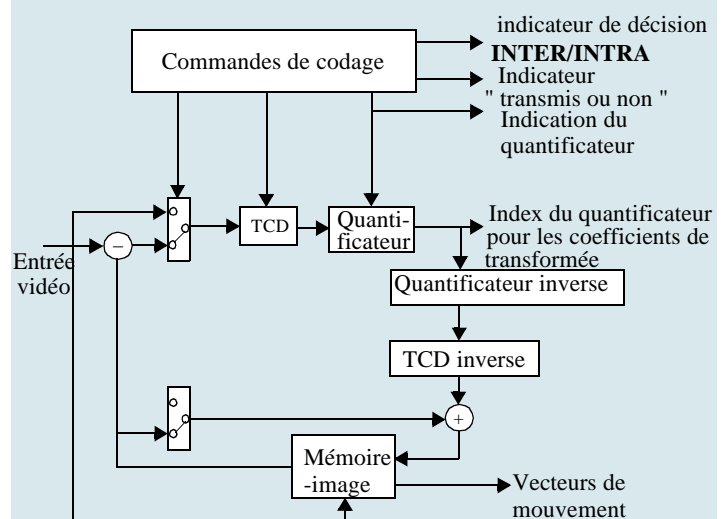
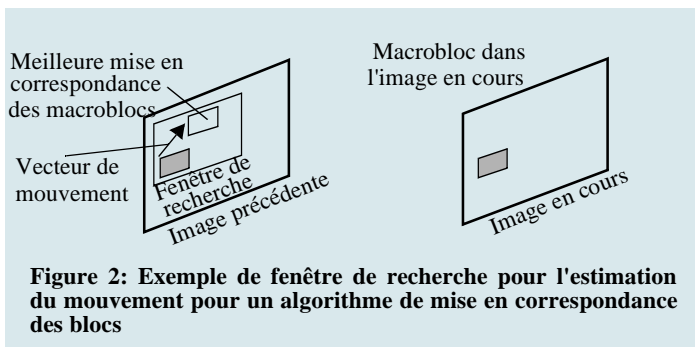


Figure 1: Schéma fonctionnel d'un codeur vidéo à compensation du mouvement basée sur des blocs et à transformée.



**Figure 2: Exemple de fenêtre de recherche pour l'estimation du mouvement pour un algorithme de mise en correspondance des blocs**

est employée dans toutes les normes courantes de codage vidéo.

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. Les images vidéo comprennent trois matrices rectangulaires de données de pixels, lesquelles représentent le signal de luminance (luma 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. Le format le plus commun rencontré dans les normes de compression vidéo est huit bits avec sous-échantillonnage 4:2:0: les deux composants de chrominance sont réduits à la moitié de la résolution verticale et horizontale du composant luminance.

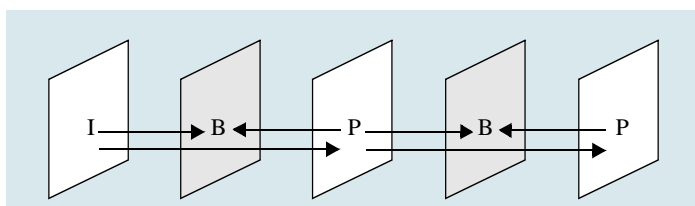
La figure 1 présente un exemple type de codeur vidéo à compensation du mouvement basée sur des blocs et à transformée. Dans les prochaines sections, nous décrirons les éléments constitutifs de ce codeur vidéo.

## 2.1 Estimation et compensation du mouvement

Chaque image vidéo est divisée en macroblocs de taille égale. La prévision du mouvement par compensation repose sur l'hypothèse qu'un bloc de pixels de l'image en cours peut être modélisé en tant que translation d'un bloc de l'image précédente, comme on peut le voir à la figure 2.

Chaque bloc est normalement prévu à partir de l'image précédente. On suppose donc que chaque pixel du bloc subit le même degré de mouvement translationnel. Les vecteurs de déplacement bidimensionnels, ou vecteurs de mouvement, représentent cette information de mouvement. Comme la représentation de l'image est fondée sur des blocs, beaucoup d'algorithmes d'estimation du mouvement utilisent des techniques de mise en correspondance des blocs dans lesquelles le vecteur de mouvement est obtenu par la réduction d'une fonction de coût mesurant la disparité entre un bloc candidat et le bloc en cours.

Bien que plusieurs fonctions de coût aient été proposées, celle la plus largement répandue dans les algorithmes d'estimation du mouvement est la somme des différences absolues (SDA), qui est définie par la somme des différences de pixels entre le bloc candidat de référence et le bloc original. Pour trouver le bloc produisant l'erreur de disparité la moins importante, il faut calculer la SDA à plusieurs endroits dans une fenêtre de recherche. La méthode de recherche la plus simple, mais aussi la plus



**Figure 3: Agencement type d'un groupe d'images.**

exigeante en matière de calculs, est la recherche complète, ou recherche approfondie, qui évalue la SDA à chaque emplacement de pixel possible dans le secteur de recherche. Pour réduire la complexité des calculs, on a proposé plusieurs algorithmes qui limitent la recherche à quelques points [7].

On attribue habituellement un vecteur de mouvement par bloc pour la compensation du mouvement. Les algorithmes d'estimation du mouvement à des valeurs inférieures au pixel peuvent améliorer de façon substantielle la qualité de reproduction. Les normes les plus récentes de codage vidéo prévoient des composants horizontaux et verticaux pour les vecteurs de mouvement d'une précision d'un demi-pixel. La gamme des vecteurs de mouvement permis limite souvent la fenêtre de recherche utilisée dans l'estimation du mouvement. Une valeur positive pour le composant horizontal ou vertical du vecteur de mouvement représente respectivement un bloc dans l'espace vers la droite ou en-dessous du bloc prévu.

Les macroblocs peuvent être prévus soit à partir des images antérieures (macroblocs P), soit à partir des images antérieures et/ou postérieures (macroblocs B). Le rendement en compression des macroblocs B est supérieur à celui des macroblocs P étant donné les options de codage additionnelles. Cependant, un délai de décodage additionnel est produit, puisque les images P postérieures doivent être décodées avant que les images B antérieures puissent être décodées. Un groupe type d'images est montré à la figure 3.

Selon les normes en vigueur, la compensation du mouvement est habituellement exécutée sur des blocs 16x16 ou 8x8 pour les macroblocs B et P, et est suivie d'une transformée, comme on l'explique dans la prochaine section.

## 2.2 Transformée

Le but de la TCD par blocs 8x8 utilisée dans toutes les normes actuelles de codage vidéo est la décorrélation des blocs 8x8 de pixels originaux ou de pixels différentiels à mouvement compensé et d'en compresser l'énergie dans le moins de coefficients possible. Outre ses capacités relativement élevées de décorrélation et de compression d'énergie, la TCD est efficace et applicable aux logiciels et au matériel. L'algorithme le plus commun pour mettre en application la TCD par blocs 8x8 est la TCD sur 8 points pour chaque rangée, suivie de la TCD sur 8 points pour chaque colonne.

Bien que la reconstitution exacte des données originales puisse être réalisée en théorie (inversion de la TCD), il est souvent impossible de la faire avec une arithmétique à précision finie. Même si les erreurs de TCD en aval peuvent être tolérées, les erreurs liées à la TCD inverse doivent respecter un minimum de précision si l'on veut éviter une disparité due à la TCD inverse entre les images reconstituées et les images originales.

La TCD par blocs 8x8 donne un coefficient DC et 63 coefficients AC. Le coefficient DC est la moyenne des échantillons transformés et représente les détails les plus bruts du bloc d'image (plus basse fréquence spatiale). Les coefficients AC représentent les détails les plus fins de l'image (fréquences spatiales plus élevées).

Dans le codeur, la transformée est suivie de la quantification des coefficients, étape à laquelle la perte des détails vidéo est justifiée par le rapport de compression vidéo.

## 2.3 Quantification

L'œil humain est plus sensible aux erreurs de reconstitution liées aux basses fréquences spatiales qu'à celles liées aux hautes fréquences. Les changements linéaires lents de l'intensité ou de la couleur (information de basse fréquence) sont importants pour l'œil. Les changements rapides dans les hautes fréquences peuvent souvent être imperceptibles, c'est pourquoi on peut les éliminer. Pour chaque coefficient dans la matrice TCD, il faut calculer une valeur quantifiée correspondante en divisant chaque coefficient TCD par un paramètre de quantification. Le paramètre de quantification est le principal moyen par lequel on peut régir le degré de compression et de réduction correspondante de la fidélité.

lité de l'information vidéo compressée. L'effet net est habituellement une variance réduite entre les coefficients quantifiés par rapport à la variance entre les coefficients TCD originaux de même qu'une réduction du nombre de coefficients non nuls, ce qui améliore l'efficacité du codage entropique, sujet dont nous allons maintenant traiter.

## 2.4 Codage entropique

Le codage entropique réduit le nombre moyen de bits employés pour représenter l'information vidéo compressée par l'utilisation de moyens tels que les codes de longueur variable, ou CLV (variable length codes, ou VLC). Les CLV sont souvent produits avec des codes de Huffman, ce qui permet l'utilisation de mots codes plus courts pour représenter les symboles ayant la plus grande occurrence (faibles valeurs de coefficient, etc.). Le codage arithmétique peut également être employé pour le codage entropique. D'autres informations, comme les types de prévisions et les indications de quantificateur, font aussi l'objet d'un codage entropique au moyen de CLV ou de codes arithmétiques.

Avant l'exécution du codage entropique, les vecteurs de mouvement sont habituellement extrapolés à la valeur médiane de trois vecteurs de mouvement voisins déjà transmis, à savoir les vecteurs de mouvement des blocs vers la gauche, vers le haut et vers le haut à droite du bloc en cours. Les vecteurs de mouvement différentiels font ensuite l'objet d'un codage entropique.

Également avant l'exécution du codage entropique, les coefficients TCD quantifiés sont agencés en un ensemble unidimensionnel par balayage en zigzag. Ce réagencement place le coefficient DC en premier dans l'ensemble, et les coefficients AC restants sont classés de la basse fréquence à la haute fréquence. L'ensemble réagencé est codé à l'aide d'une table CLV tridimensionnelle par longueur de plage selon le triplets DERNIER, PLAGE, NIVEAU. Le symbole PLAGE correspond à la distance entre deux coefficients non nuls dans l'ensemble. Le symbole NIVEAU correspond à la valeur non nulle venant tout juste après une séquence de valeurs nulles. Finalement, le symbole DERNIER est l'équivalent du code de fin de bloc utilisé dans le codage bidimensionnel par longueur de ligne, où "DERNIER = 1" signifie que le code courant correspond au dernier coefficient du bloc codé. Cette méthode de codage produit une représentation compressée des coefficients TCD du fait qu'un grand nombre de coefficients sont normalement quantifiés à zéro. Théoriquement, le réagencement entraîne la création de longues séquences de valeurs nulles consécutives.

## 2.5 Commandes de codage

Les deux commutateurs apparaissant dans la figure 1 représentent le choix de mode (intra ou inter).

Le choix du mode de codage se fait habituellement au niveau des macroblocs. Le rendement du processus d'estimation du mouvement, habituellement mesuré en fonction de valeurs de distorsion connexes, peut être employé pour choisir le mode de codage. On opte pour un codage en fonction des prévisions temporelles (mode inter) si le processus de compensation du mouvement est efficace, et seulement si le macrobloc d'erreur de prévision - la différence entre le macrobloc original et le macrobloc prévu avec compensation du mouvement - doit être codé. Si la prévision temporelle n'est pas employée, on utilise le codage en mode intra. Si un macrobloc ne change pas de manière significative par rapport à l'image de référence, le codeur peut choisir de ne pas le coder, et le décodeur répète simplement le macrobloc situé à l'emplacement indiqué dans l'image de référence. Ce mode de codage est appelé saut. Des algorithmes de choix de mode de codage plus complexes basés sur des méthodes d'optimisation débit-distorsion peuvent également être employés, comme on le verra dans la prochaine section.

## 2.6 Codage vidéo optimisé au sens débit-distorsion

L'un des volets clés du codage vidéo à compression élevée avec perte est la commande opérationnelle de l'encodeur effectuée à partir du processus d'estimation du mouvement, du choix des paramètres de quantification et du choix du mode de codage vidéo. Le processus de choix des représentations avec diverses efficacités débit-distorsion peut

être optimisé à l'aide des techniques de minimisation de Lagrange fondées sur la théorie du débit-distorsion [8]; ces techniques sont brièvement décrites dans la présente section. Au moment du codage à la source, la théorie du débit-distorsion fixe des limites pour la distorsion à la sortie que l'on peut atteindre avec un débit de sortie au codeur donné ou, réciproquement, fixe des limites pour le débit de sortie que l'on peut atteindre avec une distorsion à la sortie donnée.

Dans le codage vidéo, les modes de fonctionnement sont généralement associés à des caractéristiques débit-distorsion dépendantes du signal; les choix débit-distorsion font quant à eux partie du processus de choix des paramètres de codage. L'optimisation consiste donc à choisir, pour chaque bloc d'image, la représentation codée la plus efficace au sens débit-distorsion. Cette opération devient cependant plus complexe lorsque les diverses options de codage varient en efficacité selon le débit binaire et le contenu de la scène. Ainsi, le codage inter est efficace pour représenter un contenu évolutif clé dans des séquences d'images. Par contre, le codage intra peut être plus efficace lorsque le modèle de mouvement translationnel fondé sur des blocs ne peut représenter avec exactitude les changements dans la séquence d'images. Pour les zones de faible activité de la séquence vidéo, l'utilisation d'un codage par sauts peut être préférable. En permettant le recours à de multiples modes de fonctionnement, nous pouvons espérer un meilleur rendement au sens débit-distorsion si la méthode de choix du mode est appliquée judicieusement à différentes zones spatio-temporelles de la séquence vidéo.

Le but du système de compression vidéo est d'obtenir la meilleure fidélité (ou la plus faible distorsion  $D$ ) compte tenu de la capacité du canal de transmission déterminant de la contrainte de débit  $R(D)$ . L'optimisation peut être atteinte avec la méthode des multiplicateurs de Lagrange, où le terme de distorsion est pondéré en fonction d'un terme de débit. Selon la formulation de Lagrange, nous minimisons  $J = D + 8R$ , pour un certain paramètre de Lagrange  $\lambda$ . Pour un paramètre de Lagrange  $\lambda$  donné, la solution minimise la distorsion pour une certaine contrainte de débit. Une valeur donnée de  $\lambda$  représente un point spécifique sur la courbe opérationnelle de débit-distorsion. Il est possible d'obtenir une relation approximative entre la valeur du paramètre de quantification  $Q$ , qui commande le débit binaire de sortie, et la valeur optimale de  $\lambda$ . Cela est particulièrement utile quand une méthode de commande du débit est employée pour produire un débit binaire d'encodeur vidéo particulier.

## 3.0 Éléments communs de la nouvelle recommandation H.26L

Les éléments communs à toutes les normes de codage vidéo dont on a traité dans les sections précédentes sont également compris dans la nouvelle recommandation H.26L, qui devrait devenir la norme internationale de codage vidéo au début de 2003. Voici en résumé les éléments que comprend la recommandation H.26L: macroblocs de 16 lignes par 16 pixels; représentation de la luminance avec une résolution plus élevée que la chrominance et avec un sous-échantillonnage de 4:2:0; compensation du mouvement et transformée des blocs suivies d'une quantification scalaire et d'un codage entropique; vecteurs de mouvement extrapolés à la valeur médiane des vecteurs de mouvement des blocs voisins; images bi-directionnelles de type B supportées et pouvant faire l'objet d'une compensation de mouvement à partir des images précédentes et suivantes; mode direct pour les images de type B permettant la dérivation des vecteurs de mouvement vers l'avant et vers l'arrière à partir du vecteur de mouvement d'un macrobloc cositué dans une image de référence. Dans les sections suivantes, les blocs de codage

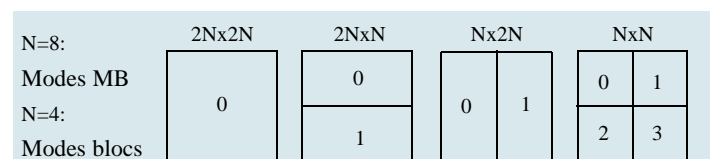


Figure 4: Division des macroblocs disponibles avec la recommandation H.26L

de la nouvelle recommandation H.26L sont comparés à d'autres normes récentes.

### 3.1 Préviation en mode intra

La recommandation H.26L utilise des techniques d'extrapolation spatiale pour les macroblocs codés en mode intra. Grâce à ces techniques avancées d'extrapolation, le rendement de la compression des images en mode intra avec la recommandation H.26L est semblable à celui obtenu avec la norme récente de compression d'images fixes JPEG-2000. La recommandation H.263 et la norme MPEG-4 permettent également l'extrapolation en mode intra. Les différences entre les recommandations H.26L et H.263 (et la norme MPEG-4) résident dans le fait que l'extrapolation est dans le domaine du pixel, par opposition au domaine de la fréquence, et que des modes d'extrapolation au niveau des sous-blocs sont disponibles, contrairement aux modes macroblocs uniquement.

Les macroblocs codés en mode intra (dans les images codées intra ou inter) peuvent employer des modes d'extrapolation spatiale par blocs 16x16 ou 4x4 pour la luminance. Trois sous-modèles sont disponibles pour l'extrapolation par blocs 16x16. Un macrobloc 16x16 peut être extrapolé à partir des pixels adjacents décodés précédemment qui sont disponibles en raison du décodage ligne par ligne des macroblocs (du haut à gauche avec des passages de gauche à droite): extrapolation verticale des pixels du haut, extrapolation horizontale des pixels vers la gauche et extrapolation en plan par interpolation spatiale entre les deux ensembles de pixels.

Neuf sous-modes peuvent être utilisés avec l'extrapolation par blocs 4x4. Un sous-bloc 4x4 secondaire peut être extrapolé à partir des pixels adjacents décodés précédemment qui sont disponibles en raison du décodage ligne par ligne de chaque bloc 8x8 d'un macrobloc, et le décodage ligne par ligne imbriqué de chaque sous-bloc 4x4 avec chaque bloc 8x8. En raison de cet ordre de décodage, on ne disposera pas pour tous les modes d'extrapolation 4x4 de données de pixels décodés dans la direction d'extrapolation souhaitée. En pareille situation, les données de pixels décodés les plus rapprochées sont employées. Les modes d'extrapolation intra sont les suivants: extrapolation DC à partir de la moyenne des pixels adjacents du haut et vers la gauche; extrapolation verticale (vers le bas) à partir des pixels du haut; extrapolation horizontale (vers la gauche) à partir des pixels de droite; extrapolation diagonale (vers le bas et la gauche) à partir des pixels du haut et de gauche; extrapolation diagonale (vers le bas et la gauche) à partir des pixels du haut et de gauche. On dispose aussi de quatre modes hors-diagonale (extrapolations de +/- 22,5 degrés: à gauche de la verticale; à droite de la verticale; en haut de l'horizontale; en bas de l'horizontale).

### 3.2 Estimation et compensation du mouvement

La recommandation H.26L soutient l'utilisation d'images de référence multiples pour la préviation des macroblocs et des blocs inter. Les images de référence multiples peuvent aider à l'extrapolation d'un fond couvert de façon transitoire et à la préviation d'un mouvement non translationnel périodique.

Comme avec la norme MPEG-4, une compensation du mouvement de  $\frac{1}{4}$  pixel est employée pour la préviation temporelle. Un filtrage par interpolation à six voisins pour les positions au  $\frac{1}{2}$  pixel est suivi d'une interpolation bilinéaire permettant d'obtenir par dérivation les positions au  $\frac{1}{4}$  pixel. On a rajouté une position factice, qui est interpolée plus localement pour les situations où seuls les détails bruts, et non les fréquences spatiales élevées, de l'image en cours sont prévus avec précision à partir de l'image de référence. En outre, un filtrage par interpolation à huit voisins pour les positions au  $\frac{1}{8}$  pixel est disponible.

Comme pour la recommandation H.263 et la norme MPEG-4, le modèle pour la compensation du mouvement est une translation de blocs de taille variable; les vecteurs de mouvement peuvent se prolonger au delà des limites de l'image par le prolongement des valeurs des pixels limitrophes à l'extérieur de l'image. Cependant, on dispose d'un plus vaste éventail de longueurs de bloc pour la compensation du mouvement.

Chaque macrobloc 16x16 peut être divisé horizontalement et/ou verticalement aux fins de la compensation du mouvement. Si un macrobloc est divisé horizontalement et verticalement en quatre blocs 8x8, chacun des blocs 8x8 peut également être divisé horizontalement et/ou verticalement. Ainsi, jusqu'à 16 vecteurs de mouvement peuvent être transmis pour un macrobloc. La figure 4 présente la division courante d'un macrobloc 16x16 et d'un bloc 8x8.

Les macroblocs des images inter (images P ou images B) peuvent être codés en mode par saut et direct (images B seulement), par préviation spatiale intra 4x4 ou par mouvement compensé avec jusqu'à 16 vecteurs de mouvement et la possibilité de coder sur option chaque sous-bloc 8x8 avec préviation intra 4x4.

Chaque sous-bloc 8x8 peut servir à effectuer une préviation à partir d'une image de référence différente.

Le post-filtrage pour l'élimination d'artéfacts de treillisage et la distorsion des hautes fréquences est considéré comme un élément critique pour obtenir une qualité perceptuelle élevée avec les normes actuelles. Avec la recommandation H.26L, la distorsion des hautes fréquences est diminuée par l'utilisation d'une transformée sur seulement quatre points. Le filtrage de détreillisage adaptatif est déplacé dans la boucle de codage de façon que la préviation temporelle soit basée sur des images (décodées) reconstituées et filtrées de qualité supérieure.

Le signal de télévision classique est diffusé dans un format entrelacé dans lequel l'image est divisée en deux trames qui sont montrées à intervalle différent. La première trame dans le temps (appelée trame supérieure) apparaît sur chaque ligne impaire de l'image, tandis que la deuxième trame d'image (appelée trame inférieure) apparaît sur chaque ligne paire de l'image. Les deux trames forment l'image. La diffusion de données entrelacées permet la séparation de la compensation du mouvement et de la transformée en deux trames quand des mouvements soudains sont présents. Avec la recommandation H.26L, le transfert adaptatif entre les trames distinctes et combinées est soutenu au niveau de l'image plutôt qu'au niveau du macrobloc, comme le permet la norme MPEG-2.

### 3.3 Transformée et quantification

Une transformée en nombres entiers pseudo-TCD par blocs 4x4 remplace la TCD par blocs 8x8 usuelle pour la décorrélation des résidus associés à la préviation des pixels. La transformée plus courte devient plus concurrentielle face aux transformées plus longues en raison des techniques améliorées de préviation pour les pixels de luminance. Les avantages de cette nouvelle transformée sont l'élimination complète des disparités associées à la transformée inverse, qui occasionnaient une disparité codeur/décodeur dans les normes précédentes, une qualité perceptuelle améliorée et une complexité réduite.

Les transformées 4x4 passent à des transformées 8x8 plus habituelles pour la chrominance et à des blocs 16x16 prévus pour la luminance par l'utilisation d'une deuxième transformée 2x2 agissant sur le coefficient DC des blocs 4x4 transformés dans un macrobloc de 4:2:0. Comme pour l'annexe T de la recommandation H.263, un plus petit paramètre de quantification est employé avec la chrominance pour en améliorer la fidélité.

### 3.4 Codage entropique

Contrairement à d'autres normes de codage vidéo récentes, une table universelle des codes de longueur variable (CLV) est utilisée pour le codage entropique CLV. On obtient la simplification en faisant correspondre chaque symbole au mot-code CLV qui convient compte tenu de sa fréquence dans le flux binaire plutôt qu'en adoptant l'approche plus courante qui consiste à élaborer des tables CLV distinctes pour chacun des ensembles de symboles (résidus de préviation des vecteurs de mouvement, longueur de plage et niveaux, modes macrobloc, etc.).

Comme solution de rechange plus complexe et plus efficace au codage entropique CLV, on peut utiliser le codage arithmétique adaptatif fondé sur le contexte (CABAC). Un code arithmétique est plus efficace qu'un

CLV pour les probabilités de symbole supérieures à 50%, puisqu'il permet à un symbole d'être représenté avec moins d'un bit. Les codes adaptatifs réduisent l'inefficacité des statistiques non stationnaires de symbole provoquées par la disparité entre les longueurs statiques de mot-code et les probabilités de symbole qui changent en raison du débit binaire, du type de mouvement à la source et d'autres facteurs. La modélisation contextuelle fournit des estimations des probabilités conditionnelles relatives aux symboles. La complexité des contextes définis avec le CABAC amène une amélioration sensible par rapport au codage arithmétique fondé sur la syntaxe, une option disponible avec la recommandation H.263.

### 3.5 Commandes de codage

L'amélioration de la compression vidéo avec les normes récentes a souvent été réalisée de façon assez prévisible par l'utilisation d'un plus grand nombre de choix. Or, un éventail de modes de codage encore plus vaste est offert par la recommandation H.26L. Aussi, lorsque le nombre de choix de codage augmente, la recherche et l'optimisation du débit-distorsion en tant qu'outils pour la prise de décisions dans le processus de codage, comme on en a parlé à la section 2.5, deviennent de plus en plus importants.

### 4.0 Rendement relatif et conclusion

Cet aperçu de la nouvelle recommandation H.26L nous a permis de comparer les différences principales qui amènent un accroissement du rendement en compression par rapport aux normes antérieures. L'amélioration de la compression pouvant atteindre 50% par rapport aux meilleures normes antérieures est la principale motivation pour mettre de l'avant la nouvelle recommandation H.26L. Bien que le stockage et la largeur de bande soient des domaines en constante évolution, la demande croissante pour des résolutions plus élevées et des flux simultanés avec les voies de transmission et les supports de stockage en place et à venir continuera à stimuler la demande pour un meilleur rendement en compression.

### 5.0 Références

- [1]. UIT-T, Recommandation H.261 : Codec vidéo pour services audiovisuels à p x 64 kbit/s, Genève, 1990.
- [2]. UIT-T, Recommandation H.263, version 2 : Codage vidéo pour communications à faible débit, Genève, 1998.
- [3]. ISO/CEI, 11172-2:1993, Technologies de l'information - Codage de l'image animée et du son associé pour les supports de stockage numérique jusqu'à environ 1,5 Mbit/s - Partie 2.
- [4]. ISO/CEI, 13818-2:2000, Technologies de l'information - Codage générique des images animées et des informations sonores associées - Partie 2 : Vidéo.
- [5]. ISO/CEI, 14496-2:2001, Technologies de l'information - Codage des objets audiovisuels - Partie 2 : Codage visuel.
- [6]. ISO/CEI JTC1/SC29/WG11, UIT-T VCEG: Working Draft Number 2 of Joint Video Team Standard, document le plus récent disponible pour le public: [ftp://ftp.imtc-files.org/jvt-experts/draft\\_standard/](ftp://ftp.imtc-files.org/jvt-experts/draft_standard/) et [ftp://standards.pictel.com/video-site/0201\\_Gen/JVT-B118r2.zip](ftp://standards.pictel.com/video-site/0201_Gen/JVT-B118r2.zip)
- [7]. Peter Kuhn, Algorithms, Complexity Analysis and VLSI Architectures for MPEG-4 Motion Estimation, Kluwer Academic Publications, 1999.
- [8]. T.Berger, Rate Distortion Theory, NJ : Prentice Hall, Inc., 1971.

## À propos de l'auteurs

**M. Guy Côté** détient un Ph.D. en génie électrique et informatique de l'University of British Columbia et un B.Sc.A. en génie électrique du Collège militaire royal du Canada. M. Côté est co-fondateur et vice-président R. et D. de VideoLocus. Avant VideoLocus, il a développé les algorithmes de codage vidéo pour PixStream Inc., entreprise acquise par Cisco Systems en décembre 2000. Chez PixStream/Cisco, il a notamment mené des recherches sur différents aspects des normes de codage vidéo, dont MPEG-1, MPEG-2, MPEG-4 et H.263. M. Côté participe activement aux travaux du groupe d'experts en codage vidéo (VCEG) de l'UIT-T et est membre votant du Comité international de normalisation ISO/CEI/JTC1/SC29 (JBIG, JPEG, MPEG). M. Côté a publié plus de 25 articles et contributions dans le domaine de la normalisation et a quatre brevets en instance dans le secteur de la compression vidéo.



**M. Lowell Winger** détient un Ph.D. en génie électrique et informatique de l'University of Toronto ainsi qu'une M.Sc.A. et un B.Sc.A. en génie d'étude de systèmes de l'University of Waterloo et fait partie d'un ordre d'ingénieurs. M. Winger est co-fondateur et présentement directeur en chef des technologies de VideoLocus Inc. Avant de fonder VideoLocus, il a assuré la supervision et l'orientation techniques pour l'élaboration d'une plate-forme de codage MPEG-2 multipasses de deuxième génération et d'une architecture pour une plate-forme de traitement vidéo flexible chez PixStream Inc. Il est également professeur adjoint au département des études de systèmes à l'University of Waterloo, un participant actif de l'IEEE (Institute of Electrical and Electronics Engineers) et membre votant du Comité international de normalisation ISO/CEI/JTC1/SC29 (JBIG, JPEG, MPEG).

