

Agrandissement vidéo par exploitation de la prédiction de mouvement

Sylvain Fassino & Annick Montanvert
Laboratoire LIS – Fédération ELESAs
961, rue de la Houille Blanche – BP 46
38402 Saint Martin d'Hères cedex
s.fassino@ifrance.com annick.montanvert@lis.inpg.fr

Résumé

Cet article traite de l'agrandissement de séquences vidéo dans le but de diminuer les capacités de stockage et les débits de transmission. Ainsi, il définit un procédé permettant d'agrandir en taille les images d'une séquence qui peut ainsi être stockée sous forme de "petit format".

Dans un premier temps, il va traiter de l'agrandissement d'images par induction, une méthode de régularisation développée pour les images fixes.

Ensuite, il décrit les méthodes de compression utilisées en vidéo par l'exploitation de la redondance temporelle.

Enfin, il explique comment les vecteurs de mouvement et les images d'erreurs obtenus par un algorithme de Block-matching vont être utilisés avec l'induction pour réaliser un agrandissement de séquences vidéo de qualité.

Mots clef

Agrandissement, vidéo, compensation de mouvement, super-résolution.

1 Introduction

Avec l'essor du multimédia, l'utilisation de séquences vidéo, de DVD et d'Internet est devenue chose courante. Malheureusement, ces applications subissent des contraintes de capacité de stockage ou de débit de transmission.

L'agrandissement des séquences vidéo est alors une bonne alternative pour réduire ces contraintes en diminuant à la source le nombre d'informations à stocker et à transmettre.

Or, l'agrandissement créé par les lecteurs MPEG actuels reste décevant. En effet, il s'agit d'une simple duplication des pixels ou bien d'une interpolation linéaire dans le meilleur des cas.

Dans les paragraphes suivants, nous allons montrer comment un procédé original développé pour des images statiques (l'agrandissement par induction [1]), va être complété par l'utilisation/compensation de mouvement pour des vidéo.

Aussi, après avoir étudié les méthodes de compression vidéo (codage supprimant la redondance temporelle), nous proposons d'utiliser les vecteurs de mouvement [7]

[5] pour produire une séquence agrandie de bonne qualité avec un temps de calcul relativement réduit.

L'agrandissement par induction a été développé pour des images en niveaux de gris, aussi, le processus que nous avons étudié est appliqué uniquement pour des séquences en niveaux de gris, mais il reste tout à fait valable pour des séquences en couleurs.

2 Agrandissement par induction

Nous allons légèrement insister sur cette méthode d'induction car elle est la base de toute une partie du travail effectué par la suite.

L'agrandissement par induction [1] est une méthode de régularisation possédant un aspect original du fait qu'elle considère l'agrandissement comme étant un problème inverse de la réduction. Ainsi agrandir une image, c'est produire une image dont la réduction redonne l'image originale (contrainte de réduction).

Après un agrandissement réalisé par un procédé quelconque (nous avons choisi l'interpolation linéaire), nous appliquons l'induction de façon itérative pour faire converger l'image obtenue initialement, vers l'image qui satisfait au mieux la contrainte de réduction.

Ainsi, cette méthode restaure le contenu fréquentiel de l'image initiale dans la résolution supérieure en cohérence avec elle-même.

Cette méthode possède un aspect ensembliste car il existe un ensemble de solutions visuellement admissibles qui satisfait cette contrainte.

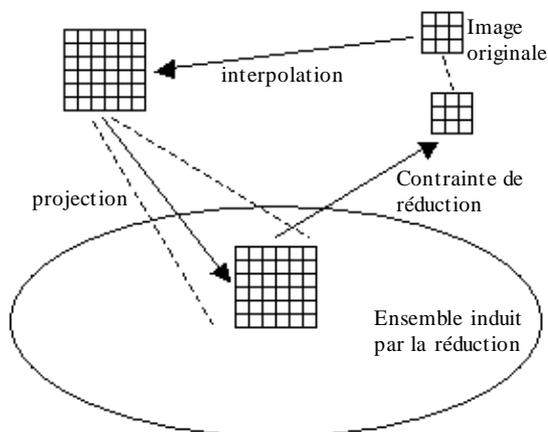


Figure 1 – Principe de l'agrandissement par induction.

Le schéma précédent (figure 1) illustre le principe énoncé. L'image originale est agrandie par interpolation linéaire pour donner une image appelée image inductrice (figure 2).

Cette image est ensuite projetée sur l'ensemble des solutions visuellement admissibles, c'est-à-dire l'ensemble des solutions qui satisfont la contrainte de réduction et qui donne une image proche de l'image originale (la figure 3 illustre le résultat obtenu après projection).

Les deux images ci-dessous comparent les résultats obtenus avec cette méthode.



Figure 2 – Agrandissement par interpolation linéaire.



Figure 3 – Agrandissement par induction.

Les résultats sont remarquables au niveau de zones texturées, où l'on peut voir une nette amélioration des détails ainsi que du contraste pour l'image qui a été induite.

Malheureusement, cette méthode a un temps de calcul relativement élevé, aussi elle est inutilisable telle qu'elle en vidéo si nous devons agrandir chaque image. Nous allons voir dans la suite, la manière dont nous allons l'employer pour réaliser des agrandissements vidéo rapides grâce notamment à la prédiction de mouvement.

3 Principe du codage vidéo

Entre deux images successives d'une séquence, un certain nombre d'éléments de la scène ne sont pas modifiés, ou se sont déplacés d'un mouvement résumé par une translation. L'idée qui est utilisée dans les différentes normes MPEG est de découper l'image en blocs, et de détecter les déplacements de ces derniers entre deux images successives [6]. Ensuite, il suffit de coder les vecteurs de déplacement détectés. C'est l'estimation et la prédiction de mouvement qui est couramment réalisée par des algorithmes de block-matching [6].

Malheureusement, les changements des différents blocs ne se limitent pas à des translations, il peut y avoir des variations de luminosité, des recouvrements, des déformations d'objets qui se traduisent par des différences de niveaux de gris.

Par conséquent, une image d'erreurs doit être créée : il s'agit de la différence entre l'image originale et l'image reconstruite par prédiction de mouvement. Cette image est ensuite codée car elle présente une faible entropie et donc elle admet une forte compression.

Dans les normes MPEG, trois types d'images sont définis :

- Les images I (*Intra*), codées indépendamment les unes des autres.
- Les images P (*Predicted*), prédites à partir d'une image de référence antérieure, qui peut être du type I ou P.
- Les images B (*Bidirectional interpolated*), prédites à partir de 2 images de références : une référence antérieure et une référence postérieure. Elles aussi peuvent être des images I ou P.

Ces images sont réparties en groupes. Un groupe d'images (Groupe Of Pictures : GOP) est composé d'un ensemble d'images P et B, entouré de deux images Intra (seule la première appartient au GOP).

Pour la suite, nous allons utiliser des paramètres standardisés pour du PAL : GOP initialisé par une image I, avec trois images P et deux images B intermédiaires [4].

Nous avons extrait des séquences les vecteurs de mouvement et les images d'erreurs par des algorithmes

de block-matching, puis nous les avons regroupés en GOP comme pour un flux MPEG.

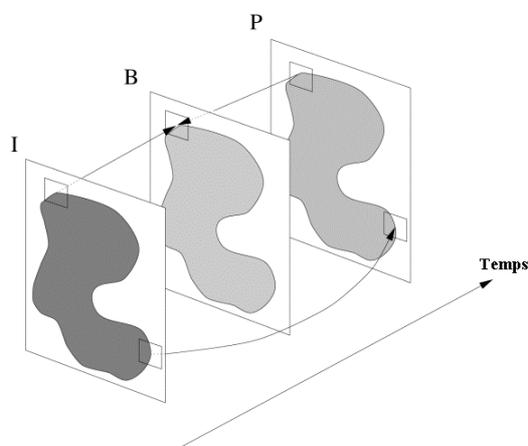


Figure 4 – Reconstruction à partir des vecteurs de mouvement.

4 Agrandissement en utilisant les vecteurs de mouvement

La méthode que nous avons mise en œuvre pour l'agrandissement d'une séquence vidéo est la suivante :

- La première image d'un groupe d'images (type I) est agrandie par induction, nous obtenons ainsi une image de bonne qualité visuelle.
- Les images suivantes du GOP sont reconstruites à partir :
 - des vecteurs de mouvement de taille double
 - de l'image d'erreurs agrandie par simple interpolation linéaire.
- Le procédé est itéré pour chaque groupe d'images (GOP) jusqu'à la fin de la séquence.

L'apport de cette méthode par rapport à une simple interpolation linéaire est l'induction de la première image d'un groupe (voir figure 8). Ainsi, les images suivantes qui se servent d'elle pour leur reconstruction voient leur netteté améliorée. De plus comme les images d'erreurs sont agrandies uniquement par interpolation, le temps de calcul moyen reste acceptable.

Le paragraphe suivant montre les différents résultats obtenus.

5 Résultats et interprétation

Voici, dans un premier temps, quelques images obtenues par notre méthode :



Figure 5 – Image I agrandie par induction.



Figure 6 – Image P reconstruite par compensation de mouvement.



Figure 7 – Image B reconstruite par bi-compensation de mouvement.

Voici, maintenant le graphe d'EQM d'une quarantaine d'images de la séquence :

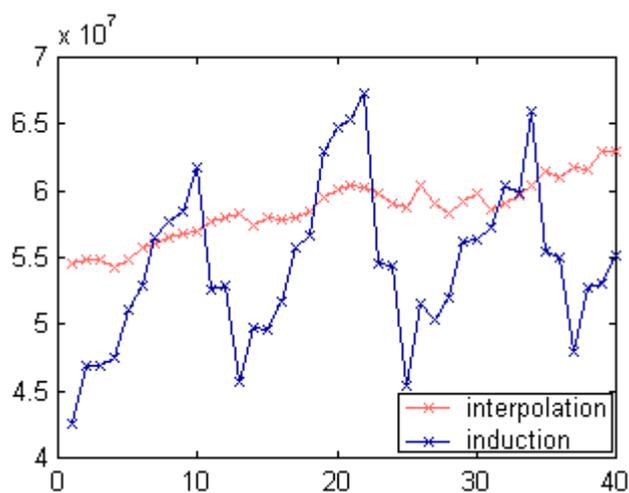


Figure 8 – Graphe d'EQM de la séquence, les images I assurent une erreur réduite, les plus grandes erreurs arrivent avec les images P, entre 2 images I nous avons la séquence I, B, B, P, B, B, P, B, B, P, B, B, I.

Les résultats d'ensemble sont corrects même avec une simple image d'erreurs agrandie par interpolation linéaire. Cependant quelques images induites ont une EQM supérieure aux images interpolées, ceci est dû à l'effet de blocs qui introduit des dégradations perçues par l'EQM alors que la qualité de l'image est bien meilleure visuellement.

Même si les résultats sont corrects, quelques problèmes apparaissent. Ainsi, lors d'un changement de plan, une simple image d'erreurs interpolée ne suffit pas pour corriger toutes les erreurs. Par conséquent, les images qui vont être prédites par la suite seront de mauvaise qualité. Il est possible de remédier à ce problème, en générant une nouvelle image I au moment du changement de scène.

L'autre effet gênant qui persiste et qui semble être délicat à supprimer, est un effet stroboscopique relativement visible. Ceci est dû à l'accumulation d'approximations et des erreurs. Ainsi, la netteté de la scène se dégrade peu à peu, puis lorsqu'un nouveau GOP apparaît, cette netteté est améliorée grâce à l'induction réalisée sur l'image I qui va suivre. Visuellement, l'œil est sensible à ces variations répétitives de la qualité lors de la séquence.

Parallèlement, des tests de robustesse à la dégradation de l'image d'erreurs ainsi qu'à la prédiction de mouvement ont été effectués.

Pour l'image d'erreurs, les expériences ont consisté à mettre les faibles valeurs de corrections à zéro.

Pour la prédiction de mouvement, deux tests ont été effectués, l'un à consister à rajouter une valeur aléatoire aux vecteurs de mouvement, l'autre à consister à utiliser des algorithmes de Block-matching sous-optimaux.

Dans tous les cas, la méthode donne des résultats corrects tant que les dégradations restent limitées.

Conclusion et travaux futurs

L'utilisation conjointe de l'agrandissement par induction, et des vecteurs de mouvement, offre un agrandissement de qualité supérieure à celui fourni par une duplication de pixels ou une simple interpolation linéaire.

Les avantages de notre méthode sont que nous sommes affranchis des différentes normes MPEG puisque nous avons utilisé uniquement les grands principes du codage. Mais, il est toujours possible de la greffer sur un flux MPEG existant.

Cependant, il reste quelques problèmes à traiter tels que l'apparition d'un effet de blocs. Ceci est dû au principe d'estimation de mouvement utilisé.

Il est bon également de remarquer, que pour des développements ou des études futures, il faudrait réfléchir sur la notion d'objet et en tirer partie pour définir des modèles adaptés ou participant au codage MPEG-4 [3] et remarquer que dans le cas d'un besoin en temps réel très contraignant, il est possible de diminuer le temps de l'induction en se focalisant sur des points importants de l'image [2]: textures, contours, zones particulières...

Il serait également intéressant de tester cette méthode dans un processus de réduction/agrandissement, dans le but de diminuer la bande-passante utilisée lors d'une transmission sur des réseaux de type Intranet/Internet.

Bibliographie

- [1] D.Calle and A. Montanvert. "Super resolution inducing of an image". Proceedings of IEEE International Conference on Image Processing ICIP-98, 3, October 1998.
- [2] C. Charrier et H. Cherifi. "Influence de la dégradation des contours sur la perception de la qualité des images couleur compressées". CORESA-98, Juin 1998.
- [3] T. Ebrahimi and C.Horne. "MPEG-4 natural video coding – An overview". http://www.cselt.it/leonardo/icjfiles/mpeg-4_si/7-natural_video_paper/7-natural_video_paper.htm.
- [4] "Generic coding of moving pictures and associated audio, Recommendation H.262". ISO/IEC 13818-2, Draft International Standard, March 1994.
- [5] N.R. Shah and A. Zakhor. "Resolution enhancement of color video sequences". IEEE Transactions on image processing, vol. 8, no. 6, pp. 879-885, June 1999.
- [6] Y.Q. Shi and H. Sun. "Image and video compression for multimedia engineering: fundamentals, algorithms and standards". CRC Press, 2000.
- [7] B.C. Tom. "Resolution enhancement of monochrome and color video using motion compensation". IEEE Transactions on image processing, vol. 10, pp. 278-287, February 2001.