Amélioration du modèle statistique de bruit pour le codage vidéo distribué

Jérôme GAUTHIER[1], Thomas MAUGEY[1], Béatrice PESQUET-POPESCU[1], Christine GUILLEMOT[2]

[1]**Telecom ParisTech**, Département TSI, 46 rue Barrault, 75634 Paris Cedex 13, France

[2]**IRISA**, Campus Universitaire de Beaulieu, 35042 Rennes Cedex, France

{gauthier,maugey,pesquet}@telecom-paristech.fr, Christine.Guillemot@inria.fr

Résumé – Nous nous intéressons, dans cet article, au modèle statistique du bruit de corrélation utilisé dans le turbo décodeur pour le codage distribué. Des tests réalisés sur différentes séquences vidéo montrent que le modèle standard reposant sur l'utilisation de la distribution Laplacienne n'est pas toujours satisfaisant. La principale contribution de cet article est de raffiner le modèle Laplacien en considérant, à la place, une distribution Gaussienne Généralisée. Le modèle Laplacien étant un cas particulier du modèle Gaussien Généralisé, il est alors possible de traiter certains cas pour lesquels le modèle Laplacien n'était pas suffisant, tout en conservant de bons résultats dans les sous-bandes où ce modèle est bien adapté aux données. Les paramètres des distributions sont estimés par deux méthodes et selon deux modes (en-ligne et hors-ligne). Afin d'illustrer l'intérêt de ce changement de modèle, nous proposons des résultats expérimentaux sur des vidéos CIF et QCIF. Les résultats de ces comparaisons indiquent que le modèle Gaussien Généralisé semble plus efficace que le modèle Laplacien.

Abstract – The statistical model used to model the correlation noise at the turbo decoder for distributed video coding is studied in this paper. After working with several video sequences, we have shown that the classical model, based on Laplacian distribution, is not always producing accurate results. We propose to replace this model by a more general one, namely the Generalized Gaussian model. Two methods are then used to estimate the parameters of this model: the moment method and the maximum likelihood method. Tests are carried on in different settings: in on-line or off-line mode and with CIF or QCIF video, demonstrating that the Generalized Gaussian model seems more efficient than the Laplacian one.

1 Introduction

Le codage vidéo distribué (CVD) est une technique récente permettant de transmettre des vidéos avec une complexité réduite à l'encodage, et ce, sans affecter les performances globales de débit/distorsion. Cette propriété est particulièrement importante pour les équipements disposant d'une faible autonomie comme les téléphones portables ou les réseaux de capteurs. Le CVD est fondé sur un résultat fondamental de théorie de l'information montrant que deux sources corrélées pouvaient être encodées indépendamment sans affecter les performances de transmission, à condition de les décoder conjointement. En particulier, dans le cadre de la transmission vidéo cela signifie que l'étude de la corrélation des trames peut être évitée à l'encodeur sans réduire la qualité visuelle de la reconstruction ou augmenter le taux de transmission.

Dans ce travail nous considérons l'approche dite de Stanford [1] qui consiste à séparer la séquence vidéo en deux ensembles de trames : les trames clefs (TC) et les trames de Wyner-Ziv (TWZ). Les TC sont transmises à l'aide d'un codec intra comme JPEG2000 ou H.264, et sont utilisées par le décodeur pour calculer une estimation des TWZ, appelée information adjacente (IA). À l'encodeur, après transformation (DCT ou ondelettes) et quantification, les TWZ sont encodées par un turbo codeur. Le décodeur corrige les IA avec les bits de parité envoyés par le codeur WZ. Ce processus est effectué sous l'hypothèse que l'erreur d'estimation peut être considérée comme une erreur de canal. Pour fonctionner, le turbo décodeur a besoin d'un modèle pour le bruit de corrélation entre la TWZ et son IA associée et les performances du codec dépendent en partie de la qualité de ce modèle. Le bruit de corrélation est généralement estimé à l'aide d'un modèle Laplacien [2, 3].

Nous proposons dans cet article de remplacer le modèle Laplacien par un modèle Gaussien Généralisé (GG) recouvrant une large classe de distributions classiques comme les Gaussiennes ou les Laplaciennes. Il a été montré que ce modèle est bien adapté pour la représentation des coefficients d'ondelettes de signaux ou d'images [4]. Notons enfin qu'il a été montré que cette distribution offre un bon modèle pour les coefficients de DCT d'images naturelles [5]. Ces propriétés peuvent conduire à appliquer ce modèle aux deux transformations usuelles en compression d'images et de vidéos que sont la DCT et les ondelettes.

Dans la section 2 nous introduisons tout d'abord les notations et les hypothèses utilisées dans ce travail avant de rappeler la méthode d'estimation de paramètre dans le cas d'une distribution Laplacienne. La section 3 est consacrée à l'estimation des paramètres pour une Gaussienne Généralisée. Après avoir illustré par des résultats de simulations la pertinence du modèle Gaussien Généralisé, des résultats de décodage sont enfin présentés à la section 4 dans différents contextes expérimentaux : en faisant varier les séquences, les fréquences d'échantillonnage, les modes (en-ligne ou hors-ligne) et les modèles.

2 Estimation de paramètres dans le cas d'une distribution Laplacienne

2.1 Représentation de l'erreur

Soit X la TWZ originale et soient I_p et I_s les trames de références construites à partir des TC précédente et suivante. Au décodeur, l'IA est notée Y et le résidu R est défini comme la différence entre les trames I_p et I_s compensées. Soit $\mathbf{s} = (x, y)$ un pixel et en notant les champs de vecteurs de mouvement précédent et suivant par MV_p et MV_s, alors Y et R s'expriment de la manière suivante :

$$Y(\mathbf{s}) = \frac{I_p(\mathbf{s} + MV_p(\mathbf{s})) + I_s(\mathbf{s} + MV_s(\mathbf{s}))}{2}, \qquad (1)$$

$$R(\mathbf{s}) = \frac{I_p(\mathbf{s} + MV_p(\mathbf{s})) - I_s(\mathbf{s} + MV_s(\mathbf{s}))}{2}. \qquad (2)$$

X, Y et R peuvent être transformés à l'aide d'une DCT entière 4×4 ou à l'aide d'une transformée en ondelette biorthogonale de type 9/7 (sur 3 niveaux de décomposition). Nous notons ainsi par $x_{k,i}$, $y_{k,i}$ et $r_{k,i}$ les $i^{\text{ème}}$ coefficients de la $k^{\text{ème}}$ sous-bande ($k \in [1, ..., K]$ et $i \in [1, ..., N_k]$), résultant de la décomposition de X, Y et R.

Une hypothèse classique en CVD est de considérer que la corrélation dépend uniquement de la sous-bande et que le bruit est modélisé par une distribution Laplacienne. Dans les premiers travaux sur le CVD [2], l'estimation des coefficients se faisait hors-ligne. Autrement dit, il était supposé que les paramètres $(\alpha_k)_{k=1}^{K}$ de chaque sous-bande sont connus par le décodeur. Cette hypothèse, peu réaliste puisqu'elle suppose connue l'erreur $x_{k,i} - y_{k,i}$ au décodeur, a ensuite été remplacée par une solution en-ligne [3] qui consiste à estimer l'erreur à l'aide des coefficients du résidu $r_{k,i}$.
La figure 1 nous montre, pour une bande i, la distribution des $(x_{k,i} - y_{k,i})_{k \in [1,...,K]}$, erreur hors-ligne, et des $(r_{k,i})_{k \in [1,...,K]}$, erreur en-ligne. L'exemple indiqué dans cette figure est représentatif d'un phénomène observé sur la plupart des sous-bandes, et pour plusieurs trames de plusieurs séquences vidéo : le résidu possède une distribution qui est plutôt proche de la distribution réelle, avec cependant des différences remarquables comme une valeur en 0 plus élevée et une plus faible variance. Cette différence est la cause d'une perte de performance entre le cas en-ligne et hors-ligne.

2.2 Estimation de paramètres

En mode en-ligne ou hors-ligne, la procédure d'estimation des $(\alpha_k)_{k=1}^{K}$ est identique. On note $\boldsymbol{\xi} = (\xi_i)_{1 \leq i \leq N_k}$ le vecteur

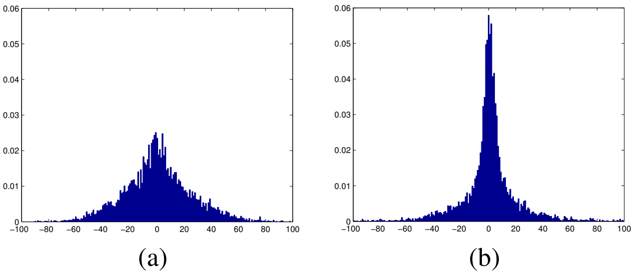

FIG. 1 – Distributions d'erreurs dans les cas hors-ligne (a) et en-ligne (b) pour une sous bande après transformation DCT 4×4.

d'observations (de moyenne nulle) dans une sous-bande. À partir du calcul de la variance σ^2, le coefficient α_k correspondant est alors obtenu par : $\widehat{\alpha_k} = \sqrt{\frac{\sigma^2}{2}}$.
Il est possible de raffiner ce modèle [3], en modifiant la valeur du coefficient α au sein d'une même bande : si $(|\xi_i| - \mu_1)^2 > \sigma^2$ alors $\widehat{\alpha}_{k,i} = \sqrt{(|\xi_i| - \mu_1)^2/2}$, où μ_1 est le moment d'ordre 1 de $|\xi_i|_{1 \leq i \leq N_k}$.

3 Estimation des paramètres d'une Gaussienne Généralisée

Rappelons tout d'abord la densité de probabilité d'une GG de moyenne nulle et de paramètres $(\alpha, \beta) \in \mathbb{R}_+^{*2}$:

$$f_{\alpha,\beta}(x) = \frac{\beta}{2\alpha\Gamma\left(\frac{1}{\beta}\right)} e^{-\left(\frac{|x|}{\alpha}\right)^{\beta}},$$

où $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt$ est la fonction Gamma d'Euler (si $\beta = 1$ on retrouve la densité d'une Laplacienne).

3.1 Méthode des moments

La première méthode d'estimation d'(α, β) est la **méthode des moments**. En combinant les expressions des moments d'ordre 2 et 4 on peut exprimer le kurtosis κ comme une fonction de β : $\kappa = \frac{\mu_4}{\sigma^4} = \frac{\Gamma\left(\frac{5}{\beta}\right)\Gamma\left(\frac{1}{\beta}\right)}{\Gamma\left(\frac{3}{\beta}\right)^2} = g(\beta)$. Les paramètres α et β sont alors estimés par :

$$\widehat{\beta} = g^{-1}(\kappa) \quad \text{et} \quad \widehat{\alpha} = \sqrt{\frac{\Gamma\left(\frac{1}{\beta}\right)}{\Gamma\left(\frac{3}{\beta}\right)}\sigma^2}.$$

Cette méthode repose donc sur l'estimation de la variance et du kurtosis du vecteur observé et sur l'inversion de la fonction g, strictement décroissante.

3.2 Maximum de vraisemblance

La seconde méthode envisagée est la **méthode du maximum de vraisemblance**. Dans cette section, nous cherchons à nouveau à estimer les paramètres α et β étant donné un ensemble d'observations indépendantes $\boldsymbol{\xi} = (\xi_i)_{1\leq i\leq N}$. La densité de probabilité de la distribution conjointe s'écrit alors :

$$F_{\alpha,\beta}(\boldsymbol{\xi}) = \left(\frac{\beta}{2\alpha\Gamma\left(\frac{1}{\beta}\right)}\right)^N e^{-\sum_{i=1}^N \left(\frac{|\xi_i|}{\alpha}\right)^\beta}.$$

On exprime alors l'anti log-vraisemblance sous la forme :

$$\begin{aligned} p(\alpha,\beta|\boldsymbol{\xi}) &= -\ln(F_{\alpha,\beta}(\boldsymbol{\xi})) \\ &= \sum_{i=1}^N \left(\frac{|\xi_i|}{\alpha}\right)^\beta + N\left(\ln(\alpha) - \ln\left(\frac{\beta}{2\Gamma\left(\frac{1}{\beta}\right)}\right)\right). \end{aligned} \tag{3}$$

Afin de minimiser l'anti log-vraisemblance, ce qui revient à maximiser la vraisemblance, on dérive tout d'abord $p(\alpha,\beta|\boldsymbol{\xi})$ par rapport à α :

$$\frac{\partial p(\alpha,\beta|\boldsymbol{\xi})}{\partial\alpha} = -\frac{\beta}{\alpha^{\beta+1}}\sum_{i=1}^N |\xi_i|^\beta + \frac{N}{\alpha}.$$

En étudiant les zéros de cette différentielle partielle, on peut exprimer α_{min} comme une fonction de β :

$$\alpha_{min} = \left(\frac{\beta}{N}\sum_{i=1}^N |\xi_i|\right)^{\frac{1}{\beta}}. \tag{4}$$

En combinant les équations (3) et (4), nous obtenons :

$$p(\widehat{\alpha},\beta|\boldsymbol{\xi}) = \frac{1}{\beta} - \ln\left(\frac{\beta}{\Gamma\left(\frac{1}{\beta}\right)}\right) + \frac{1}{\beta}\ln\left(\frac{\beta}{N}\sum_{i=1}^N |\xi_i|^\beta\right) = h(\beta).$$

Finalement, $\widehat{\beta}$ est calculé comme l'argmin de h et on obtient $\widehat{\alpha}$ en remplaçant β par $\widehat{\beta}$ dans (4).

4 Résultats expérimentaux

4.1 Tests préliminaires

Les deux méthodes d'estimation ont d'abord été testées sur des vecteurs de longueur 6336 (correspondant au nombre de coefficients d'une sous bande d'une DCT 4×4 pour une image CIF) générés avec différentes valeurs de (α,β). Les résultats obtenus sur 100 réalisations montrent que les deux méthodes évaluent bien les paramètres (α,β) (l'erreur quadratique moyenne étant inférieure à 10^{-3}).
Un deuxième test a été mené en mode hors-ligne sur les différentes sous bandes d'une image de la vidéo Football. Les résultats dans le cas d'une DCT 4×4 sont présentés dans le tableau 1 pour les méthodes d'estimation par moment et par maximum de vraisemblance. On peut remarquer, dans un premier temps, que les paramètres $\widehat{\beta}$ estimés avec les deux méthodes sont éloignés de 1 pour presque toutes les sous-bandes, justifiant ainsi *a posteriori* le choix d'une distribution GG à la place d'une Laplacienne. On constate également des différences dans les résultats d'estimation entre les deux méthodes même si ils sont cohérents entre eux pour la plupart des sous-bandes.

TAB. 1 – Paramètres (α,β) estimés dans les différentes sous bandes en mode hors-ligne sur une image de la séquence Football.

Bande	(α,β) MV	(α,β) Mom
1	$(7,26;0,50)$	$(11,07;0,54)$
2	$(7,22;0,58)$	$(7,02;0,57)$
3	$(4,14;0,57)$	$(2,68;0,50)$
4	$(2,49;0,63)$	$(1,58;0,54)$
5	$(9,44;0,63)$	$(9,59;0,63)$
6	$(6,94;0,61)$	$(6,26;0,59)$
7	$(3,74;0,58)$	$(3,33;0,56)$
8	$(2,07;0,62)$	$(1,72;0,59)$
9	$(6,97;0,68)$	$(6,36;0,66)$
10	$(5,47;0,67)$	$(4,51;0,62)$
11	$(3,91;0,68)$	$(3,19;0,63)$
12	$(1,65;0,67)$	$(1,09;0,58)$
13	$(5,17;0,85)$	$(4,74;0,82)$
14	$(2,36;0,69)$	$(2,61;0,72)$
15	$(1,29;0,64)$	$(1,32;0,65)$
16	$(0,63;0,27)$	$(0,42;0,47)$

La figure 2 présente, dans les cas en-ligne et hors-ligne, les densités de probabilité Laplacienne et Gaussienne Généralisée obtenues après estimation des paramètres sur deux sous-bandes d'une image de la séquence Foreman. Les images (a) et (b) représentent la distribution d'erreur (cas hors-ligne) et les distributions Laplacienne (en rouge) et Gaussienne Généralisée (en vert), les paramètres de cette dernière étant estimés avec la méthode des moment. Les images (c) et (d) représentent la distribution du résidu (cas en-ligne) et à nouveau les distributions Laplacienne (en rouge) et Gaussienne Généralisée (en vert). On constate que dans les deux cas présentés la distribution Gaussienne Généralisée estime de façon plus précise la distribution de l'erreur ou du résidu. Elle prédit une valeur plus juste du pic en zéro et approxime mieux la distribution pour des erreurs comprises entre 0 et 20.

4.2 Résultats débit-distortion

Nous nous sommes ensuite intéressés à différents cadres expérimentaux : en considérant des vidéos de différentes résolutions (CIF : 288×352 et QCIF : 144×176), différentes fréquences d'images par secondes (15Hz et 30 Hz) et en se plaçant en mode en-ligne ou hors-ligne. Les tests ont également été effectués à quatre niveaux de quantification (Q-Index des TWZ | Q-Step du codage intra des TC) : $1|42$, $4|34$, $6|31$ et $8|28$, afin de couvrir une large plage de débits. Enfin, les simulations

TAB. 2 – Gains en débit (%) de la méthode 2 par rapport à la méthode 1 sur différentes séquences.

Méthode 1	Méthode 2	City (CIF, 30Hz)	Football (CIF, 30Hz)	Foreman (QCIF, 15Hz)
Laplacien hors-ligne	GG hors-ligne MV	-0,96	-3,73	-1,78
Laplacien hors-ligne	GG hors-ligne Mom	1,21	-3,61	-1,52
Laplacien en-ligne	GG en-ligne MV	0,36	-3,29	-0,90
Laplacien en-ligne	GG en-ligne Mom	-1,30	-4,30	-1,88
Laplacien hors-ligne	Laplacien en-ligne	1,73	2,67	1,53
GG hors-ligne MV	GG en-ligne Mom	1,40	2,10	1,39
Laplacien hors-ligne	GG en-ligne Mom	0,44	-1,64	-0,38

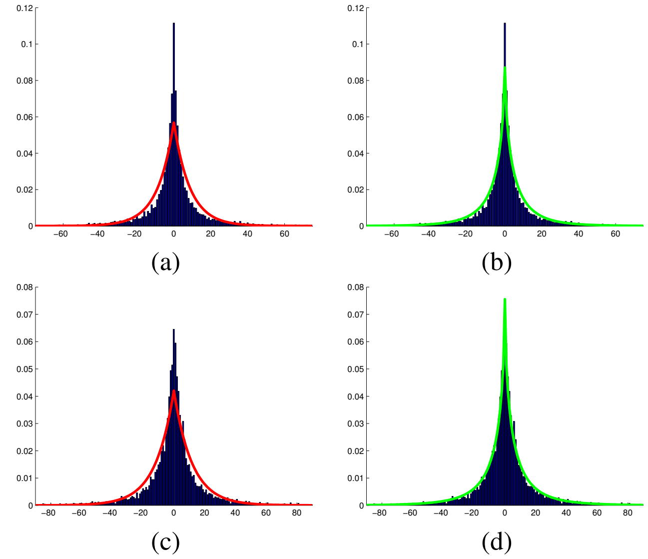

FIG. 2 – Exemples de distributions d'erreur (cas hors-ligne : (a) et (b)) et du résidu (cas en-ligne : (c) et (d)) pour deux sous-bandes de la séquence Foreman ainsi que des densités Laplacienne (en rouge) et Gaussienne Généralisée (en vert) dont les paramètres sont évalués par la méthode des moments.

ont été réalisées en employant les modèles Laplacien (Lap) et Gaussien Généralisé (GG), dont les paramètres sont estimés avec : le maximum de vraisemblance (MV) et la méthode des moments (Mom).

Le tableau 2 montre un exemple de résultat que l'on obtient en changeant le modèle Laplacien par un modèle GG dans le cas d'une transformée DCT entière 4×4. Les gains en débit sont calculés à l'aide de la « métrique » de Bjontegaard [6]. On constate que sur toutes les séquences testées la méthode GG permet de diminuer le débit aussi bien en mode hors-ligne (jusqu'à $3,73\%$ sur Football CIF et $1,78\%$ sur Foreman QCIF) qu'en mode en-ligne. Pour un PSNR de $38,38$dB sur la séquence Football cela correspond à une réduction de 194kbs hors-ligne et 128kbs en-ligne, et sur Foreman à $39,94$dB les différences sont de 44kbs hors-ligne et 46kbs en-ligne. On peut noter que la méthode MV semble plus performante en hors-ligne et que la méthode des moments donne de meilleurs résultats en mode en-ligne. Finalement, on peut remarquer qu'en utilisant GG Mom en mode en-ligne on peut, sur certaines séquences, obtenir des gains par rapport aux résultats avec le modèle Laplacien hors-ligne.

5 Conclusion

Dans cet article, nous avons mis en évidence l'intérêt d'utiliser un modèle plus général que le modèle Laplacien au niveau du turbo-décodeur dans le cadre du codage vidéo distribué. Les résultats expérimentaux montrent ainsi qu'en utilisant une transformée de type DCT 4×4 on peut améliorer les résultats débit-distortion grâce à ce changement de modèle. Une étude préliminaire montre que les résultats obtenus à la section 4.1 se vérifient également en utilisant un codeur de type JPEG2000, une perspective intéressante serait alors de poursuivre ces travaux en utilisant un codeur WZ employant une transformée en ondelettes.

Références

[1] B. Girod, A. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed video coding," *Proc. IEEE*, vol. 93, pp. 71–83, Jan. 2005.

[2] A. Aaron, R. Zhang, and B. Girod, "Wyner-Ziv coding of motion video," in *Proc. Asilomar Conference on Signals, Systems and Computers*, vol. 1, pp. 240–244, Nov. 2002.

[3] C. Brites and F. Pereira, "Correlation noise modeling for efficient pixel and transform domain Wyner–Ziv video coding," *IEEE Trans. on Circ. and Syst. for Video Technology*, vol. 18, pp. 1177–1190, Sep. 2008.

[4] S. Mallat, "A theory for multiresolution signal decomposition : The wavelet representation," *IEEE Trans. on Pattern Anal. and Match. Int.*, vol. 11, pp. 674–693, July 1989.

[5] F. Müller, "Distribution shape of two-dimensional dct coefficients of natural images," *Electronics Letters*, vol. 29, pp. 1935–1936, Oct. 1993.

[6] G. Bjontegaard, "Calculation of average PSNR differences between RD curves," tech. rep., 13th VCEG-M33 Meeting, Austin, TX, USA, Apr. 2001.