

**JOURNÉES
D'ÉTUDES
ET
D'ÉCHANGES**

**“Nouvelles
techniques
pour la compression
et la représentation
des signaux
audiovisuels”**

12-13 janvier 1995

 **France Telecom
CNET**

Centre National d'Études
des Télécommunications
38-40, rue du Général Leclerc
92131 Issy-les-Moulineaux Cédex

CORESA'95

CCETT

Centre Commun d'Études
de Télédiffusion
et Télécommunications
4, rue du Clos Courtel

35512 Cesson-Sévigné Cédex

DÉCOMPOSITION PSYCHOVISUELLE ET TRANSMISSION RAPIDE D'IMAGES

Jean-Pierre Guédon & Dominique Barba

Laboratoire d'Analyse et de Traitement d'Image, SEI, IRESTE

Atlanpole, route de la chantrerie, CP 3003, 44087 Nantes

e-mail : jguedon@lati.ireste.fr, dbarba@lati.ireste.fr

I. INTRODUCTION

I.1 Importance d'une décomposition psychovisuelle

La description du contenu d'une image en termes de canaux visuels [1] est la base d'une décomposition psychovisuelle de l'image qui permet de ne retenir de l'image que ce que l'oeil humain est capable de percevoir. Ce schéma de décomposition permet donc de réduire avant toute quantification la masse de données à transmettre. Un premier ensemble de tests a montré la validité d'un découpage radial/angularaire des canaux dans le domaine de Fourier. Le système de perception visuelle peut être représenté par un ensemble de 5 couronnes (avec un découpage radial quasi-dyadique), chacune de ces couronnes étant elle-même divisée en secteurs angulaires dont le nombre dépend de la couronne. La décomposition d'image selon cet ensemble de bandes radiales/angularaires appelés patches a été obtenue dans un second temps en modifiant la méthode proposée par Watson [2]. L'image originale est divisée en sous-images correspondant à chacun des patches en utilisant une matrice de sous-échantillonnage propre à chacun d'eux. Dans chaque sous-image, on transforme le signal obtenu en un signal de contraste local [3] donné par le rapport entre la luminance du pixel dans la sous-image et la luminance du pixel reconstruite à partir des couronnes inférieures (luminance moyenne locale). Lorsque l'on conçoit des quantificateurs psychovisuels optimaux pour quantifier ce contraste local, on obtient des quantificateurs qui sont toujours linéaires [4]. Les résultats pratiques donnent une excellente qualité visuelle des images codées par ce système (erreurs non visibles).

I.2 Transmission rapide d'images

Les méthodes de codage d'image sont tributaires des modes et des canaux de transmission à débit élevé comme par exemple la transmission ATM (Asynchronous

Transfer Mode) [5]. Dans ce type de technique, le coût pour transmettre avec des impératifs de temps-réel risque plus d'être dans la retransmission de cellules corrompues que dans la transmission elle-même. Ceci est d'autant plus vrai que les mécanismes de retransmission sont rejetées aux extrémités. Il existe deux mécanismes pour trouver une solution à ce problème. Le premier est de corrélérer les données aux niveaux du protocole qui peut ainsi corriger des cellules endommagées voire non reçues (mécanismes d'entrelaçage, de découpage de l'information au niveau de l'AAL pour mise en cellule). La seconde est au niveau du codage : on utilise une transformation qui modifie la corrélation de l'information image. Dans l'espace transformé, on obtient des groupes d'information décorrélés les uns des autres mais avec une forte corrélation des éléments d'un groupe (Fourier, ondelettes). Cependant ces deux niveaux de corrélation sont indépendants et il est donc difficile de tirer parti des deux à la fois (pour éviter des retransmissions).

La transformée de Radon que l'on propose, basée sur des projections de l'image, permet de corrélérer les pixels de l'image de façon indépendante sur chaque projection. En découpant ensuite chaque projection avec une analyse multirésolution, on obtient une façon de produire des données très indépendantes supportées par un petit nombre de cellules ATM. Le gros avantage ici est de pouvoir générer plus de projections (ou parties de projections) que nécessaire pour assurer la bonne reconstruction même si une partie de l'information transmise n'est pas arrivée. Sur le plan mathématique, on passe donc d'un formalisme ondelette qui supprime toute redondance à une structure de frame : l'ensemble des vecteurs projections est sur-complète pour reconstruire l'information.

II. TRANSFORMÉE DISCRÈTE DE RADON MULTIRÉSOLUTION

La transformée que l'on propose est basée sur une représentation exacte de l'image, des critères psychovisuels et de la redondance. On définit ici une transformée discrète de Radon spécifique et sa transformée inverse.

II.1 Transformée de Radon et Fourier

La transformée de Radon [6] d'une image $f(x,y)$, notée $Rf(x,y)$, définit l'ensemble des projections de $f(x,y)$ pour tout angle de projection $\theta \in [0,\pi[$, par

$$Rf(x,y) = p_{\theta}(t) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x,y) \delta(t - x\sin\theta + y\cos\theta) dx dy . \quad (1)$$

La transformée inverse existe mais fait généralement partie de la classe des problèmes inverses mal-posés. Nous évitons ce problème en définissant une transformée discrète adaptée aux images numériques. Le lien entre transformées de Fourier et de Radon est donné par le théorème de la tranche centrale. Soit $P_\theta(v)$ la transformée de Fourier 1D de la projection $p_\theta(t)$ et $F(\lambda, \mu)$ la transformée de Fourier 2D de $f(x, y)$, on a alors : $P_\theta(v) = F(v \sin\theta, -v \cos\theta)$. Ce théorème explique pourquoi l'espace de Radon est bien adapté à la décomposition angulaire psychovisuelle puisque chaque rayon du domaine de Fourier est directement relié à une projection dans l'image. En pratique, une image $N \times N$ possède une transformée de Fourier sur une grille cartésienne: c'est en choisissant des angles de projections qui respectent cette grille que l'on va définir une transformée exacte et réversible.

On définit donc des angles de projections de la forme $\tan \theta = \frac{p}{q}$ où $p, q \in \mathbb{Z}$ et sont premier entre eux comme dans Katz [7]. La transformée est exacte et réversible si le pas d'échantillonnage sur chaque projection dépend de l'angle de projection (et que le nombre de projections est suffisant). La transformée de Radon discrète est donc définie par :

$$Rf(k, l) = p_\theta(m) = \sum_k \sum_l f(k, l) \Delta(m - k \sin\theta + l \cos\theta); \quad (\tan \theta = \frac{p}{q})$$

avec $\Delta(m - k \sin\theta + l \cos\theta) = 1$ si $m = \frac{p(kp + lq)}{q\sqrt{p^2 + q^2}}$ et 0 ailleurs.

Pour un pas d'échantillonnage de Δ dans l'image, une projection d'angle θ est échantillonnée à $m\Delta \sin\theta$ si $\theta \in [0, \frac{\pi}{2}[$, et $m\Delta \cos\theta$ si $\theta \in [\frac{\pi}{2}, \pi[$.

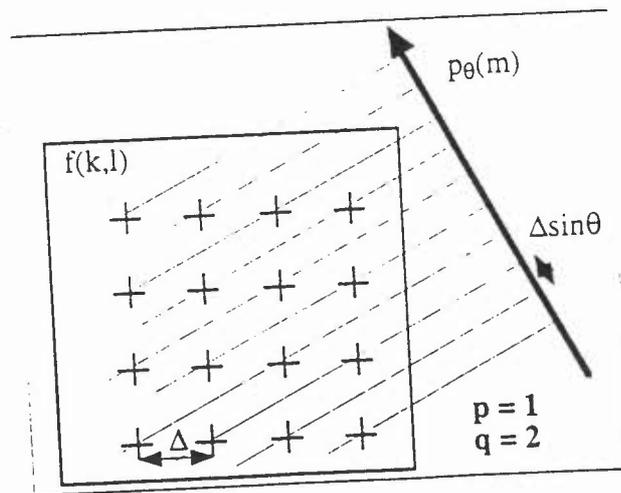


Figure 1 : La transformée discrète de Radon pour $\theta = \arctan(1/2)$

Le premier avantage de cette définition est que la valeur d'une cellule de projection est obtenue par simple addition de valeurs de pixels. Les deux grandes différences avec une transformée de Radon classique résident dans le pas d'échantillonnage de la projection qui devient dépendant de l'angle et la valeur de ces angles qui ne seront pas régulièrement espacés sur $[0, \pi]$. L'algorithme correspondant est trivial. Pour terminer la définition de cette transformée il nous reste à définir le nombre de projections et leur valeurs. Katz a montré que l'on peut trouver une borne supérieure pour le nombre M de projections nécessaire pour assurer une transformation réversible. Ainsi, pour un ensemble de M projections définies par (p_i, q_i) , (tel que $\tan \theta_i = \frac{p_i}{q_i}$) on peut retrouver

l'image $N \times N$ à partir des projections si : $N < 1 + \text{Max} \left(\sum_{i=1}^M |p_i|, \sum_{i=1}^M |q_i| \right)$. Cette

inéquation ne donne toutefois pas le nombre minimal d'angles M en fonction de N . La transformée de Radon est l'outil le plus adapté pour décrire l'aspect angulaire des patches psychovisuels. Concernant le découpage radial, il est bien connu qu'une analyse multirésolution résoud exactement cela sur des grilles cartésiennes. On approximera donc ici le découpage de couronnes par des carrés dyadiques successifs. Ceci est aussi cohérent avec l'échantillonnage dépendant des projections que nous avons choisi.

II.2 Transformée directe

L'algorithme calcule une projection d'angle $\theta = \arctan\left(\frac{p}{q}\right)$ à partir d'une image N^2 ; le nombre de points sur cette projection est alors $m_\theta = 1 + (N-1)(|p|+|q|)$. En fait les petites valeurs de (p, q) produisent des cellules très corrélées mais peu de cellules (pour $\theta=0$: $m_\theta = N$) alors que pour des hautes valeurs de (p, q) ($|p|+|q| \cong N$) m_θ est de l'ordre de N^2 et il n'y a plus de corrélation entre pixels. Les différents degrés de corrélation constituent le point clé pour la transformée inverse.

Le second choix concerne la structure formée par l'ensemble des angles de projection. Pour obtenir un algorithme efficace pour la transformée inverse et pour respecter les caractéristiques psychovisuelles, les projections sont choisies à partir des suites de Farey. Une suite de Farey d'ordre N est l'ensemble F_N de rationnels réduits $\frac{p}{q}$ dans $[0, 1]$ avec $q \leq N$, arrangés en ordre croissant. Par exemple, $F_6 = \left(\frac{0}{1}, \frac{1}{6}, \frac{1}{5}, \frac{1}{4}, \frac{1}{3}, \frac{2}{5}, \frac{1}{2}, \frac{3}{5}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \frac{5}{6}, \frac{1}{1} \right)$. Les propriétés de ces suites (exemple : 2 éléments consécutifs $\frac{p}{q}$ et $\frac{p'}{q'}$ vérifient $p'q - pq' = 1$) permettent de mettre en oeuvre des algorithmes efficaces en prenant $\theta = \arctan\left(\frac{p}{q}\right)$. Enfin, ces projections sont suffisamment régulièrement espacées pour permettre une quantification psychovisuelle.

II.3 La transformée inverse en mono-résolution

Le calcul de la transformée inverse peut être fait en résolvant un système linéaire $Ax=b$ où la matrice A est de taille $N^2 \times N^2$. Pour éviter cela, nous utilisons le degré de corrélation relatif des pixels dans les projections. Lorsque les angles de projections sont des éléments d'une série de Farey on peut reconstruire l'image en partant des bords et en allant vers le centre de l'image. Les coins sont reconstruits en utilisant des cellules non corrélées. On augmente ensuite le degré de corrélation dans les cellules et on utilise les pixels déjà reconstruits pour en déduire les autres. Cette opération s'effectue en ne résolvant que des matrices triangulaires et un seul point de projection par pixel à reconstruire. L'ordre de l'algorithme est entre N^2 et N^3 . A titre d'exemple, on a montré que pour $N=32$, les angles de Farey $(\frac{-3}{4}, \frac{-2}{3}, \frac{-1}{2}, \frac{-1}{3}, \frac{-1}{4}, \frac{1}{4}, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{1}{1})$ donnent une transformée réversible. Cependant, le nombre de cellules à coder reste trois fois plus élevé que le nombre de pixels. L'analyse multirésolution, indispensable pour la quantification psychovisuelle du contraste local permet de réduire considérablement cette proportion pour atteindre un échantillonnage quasi-critique.

II.4 La transformée inverse en multirésolution

A partir de l'image pleine résolution, les N^2 pixels sont transformés en projections par la méthode précédente. Avant l'étape de quantification et de transmission, chaque projection est découpée de façon dyadique en partie basse et haute fréquences en utilisant une fonction d'échelle et son ondelette associée, ici une ondelette cubique spline. La partie basse-fréquence des projections correspond à une image de résolution $\frac{N}{2} \times \frac{N}{2}$ en utilisant la transformée inverse.

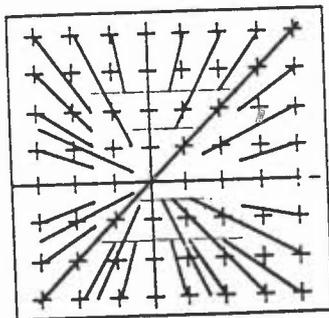


Figure 3 : Transformée de Fourier des projections à trois résolutions différentes (2, 4, 8)

Cependant, on dispose de trop de projections pour reconstruire l'image basse-fréquences et on peut en supprimer tout en gardant la transformée réversible à cette résolution. La clé de l'algorithme est que le décodeur peut reconstruire cette partie basse-fréquence non transmise en utilisant la transformée directe. C'est en itérant ce processus (destruction de plus en plus d'angle lorsque l'on réduit la résolution), que l'on obtient un échantillonnage quasi-critique comme représenté schématiquement dans la Figure 3. On peut noter que le niveau de redondance peut donc être choisi en fonction des qualités du réseau.

III. CONCLUSION

Cette transformée permet de bien décrire le plan psychovisuel, mais aussi d'utiliser directement un ensemble d'information très légèrement supérieur à celui de départ (par la redondance) pour éviter des mécanismes de retransmission sur des réseaux rapides. Il faut noter que la transformée est réversible et de mise en oeuvre possible en multirésolution à cause du pas d'échantillonnage qui dépend de la projection. La seule modification faite à l'exactitude du schéma psychovisuel est de remplacer le support fréquentiel de forme disque par un support carré. C'est la seule façon d'obtenir un schéma multirésolution qui reconstruise exactement l'image si aucune quantification de la décomposition ondelettes n'est effectuée.

BIBLIOGRAPHIE

- [1] A. Saadane, H. Senane, D. Barba, "Masking and quantization laws in a visual subband coding scheme," IEEE Conf. on image processing, Vol. II, pp.869-873, November 1994.
- [2] E. Peli, "Contrast in complex images," JOSA, Vol. 7, n°10, pp. 2032-2040, October 1990.
- [3] A. B. Watson, "Efficiency of a model human image code," JOSA, Vol. 4, pp. 2401-2417, 1987.
- [4] A. Saadane, D. Barba, "Quantificateurs psychovisuels et schéma de codage associé," Journées sur les "nouvelles techniques pour la compression et la représentation des signaux audiovisuels", CNET, Janvier 1995.
- [5] J.P. Leduc, "Multiplexing digital television sources on ATM networks," Signal Processing: Image Communication, Vol. 6, N°5, Octobre 1994.
- [6] J. Radon, "Über die Bestimmung von Functionen durch ihre Integralwerte langs gewisser Mannigfaltigkeiten." Berichte Sachsische Academie der Wissenschaften, Leipzig, *Math.-Phys. Kl.*, Vol 69 pp. 262-267, 1917.
- [7] M. Katz, Questions of uniqueness and resolution in reconstruction from projections. Lectures Notes in Biomathematics 26, S. Levin Ed, Springer-Verlag, Berlin, 1978.

Le laboratoire LATI/SEI fait partie du Groupe De Recherche 134 Traitement Du Signal et Images