

Masques perceptuels reposant sur l'interpolation en tatouage numérique

Vincent MARTIN, Marie CHABERT, Bernard LACAZE
ENSEEIH/IRIT, Institut National Polytechnique de Toulouse
2 Rue Camichel, BP 7122, 31071 Toulouse Cedex 7, France

vincent.martin, marie.chabert, bernard.lacaze@enseeiht.fr

Résumé – Cet article s'intéresse à la construction de tatouages numériques imperceptibles. Des masques psychovisuels, destinés à être combinés avec les algorithmes de tatouage classiques, sont proposés. Ils consistent à pondérer le tatouage par une erreur d'interpolation, générée heuristiquement ou par filtrage. Le masque spatial obtenu privilégie les textures et les contours du document. De nombreuses variantes, notamment préservant les contours, sont envisageables. D'autre part, nous proposons une comparaison des performances perceptuelles de masques psychovisuels classiques et des masques proposés, évaluées à l'aide de mesures de qualité objectives.

Abstract – This paper deals with the design of imperceptible digital watermarks. Psychovisual masks are proposed and combined with classical watermarking schemes. These masks weight the watermark by an interpolation error generated heuristically or by filtering. The obtained spatial mask favors textures and edges of the document. Numerous extensions are possible, such as edge-preserving interpolation. A comparison of the perceptual performance of classical and proposed masks, evaluated by objective quality metrics, is also provided.

1 Introduction

Le tatouage numérique consiste à insérer une information dans le contenu d'un document numérique, sous des contraintes d'imperceptibilité, de sécurité et de robustesse aux attaques. Ses applications incluent la protection des droits d'auteurs numériques, l'authentification de documents ou l'amélioration de contenu. L'insertion du tatouage peut s'effectuer dans n'importe quelle composante du document, comme le domaine spatial ou fréquentiel. Soit \mathbf{x} le document original, \mathbf{w} le tatouage, $\mathbf{y} = \mathbf{x} + \mathbf{w}$ le document tatoué. \mathbf{y} est transmis et éventuellement attaqué, conduisant à \mathbf{z} . Dans le tatouage additif à étalement de spectre par séquence directe (DS), \mathbf{w} résulte de la modulation d'un bit antipodal $b = \pm 1$ par un pseudo-bruit \mathbf{c} : $\mathbf{w} = b\mathbf{c}$. A la détection, on effectue une corrélation de \mathbf{z} par \mathbf{c} , afin d'obtenir une estimation \hat{b} du message : $\hat{b} = \text{signe}(\langle \mathbf{z}, \mathbf{c} \rangle)$. La transmission de plusieurs bits d'information est possible grâce à l'utilisation de mises en forme telles que la division temporelle, ou encore d'une division par codes. Le tatouage informé consiste à utiliser la connaissance du document hôte lors de l'insertion, afin d'améliorer les performances. Par exemple, l'étalement de spectre amélioré élimine une partie des interférences de l'hôte lors de l'insertion [22]. Les travaux récents se concentrent sur le codage avec information adjacente (*binning*) [17]. Un dictionnaire (éventuellement aléatoire) est divisé en sous-dictionnaires, chacun correspondant à un message possible. \mathbf{y} est l'élément du sous-dictionnaire approprié qui est le plus proche de \mathbf{x} . Le décodage identifie le sous-dictionnaire le plus proche du document reçu. En pratique, des dictionnaires sous-optimaux sont construits par quantification structurée, comme dans le Schéma de Costa Scalaire (SCS) et à étalement (ST-SCS) [5].

L'imperceptibilité est une préoccupation majeure en tatouage. C'est pourquoi la plupart des schémas de tatouage doivent être combinés avec des vecteurs Ψ , appelés masques perceptuels. Par exemple, Ψ peut pondérer les tatouages DS : $\mathbf{y} = \mathbf{x} + \Psi\mathbf{w}$.

Cette pondération peut être prise en compte dans les stratégies de tatouage informé comme l'étalement de spectre amélioré. Le code d'étalement dans ST-SCS peut également être pondéré par Ψ [14]. Dans les autres schémas fondés sur la quantification, le pas de quantification peut être adapté localement à l'hôte grâce à Ψ [3]. Ψ doit alors être recalculé au décodage à partir de \mathbf{z} . Cependant, dans le cas le plus fréquent de masques non stationnaires (car adaptés à une image, par exemple), une estimation précise de Ψ est difficile si le document a été altéré. En cas d'attaque, cette erreur d'estimation sur le pas de quantification nuit à la robustesse [13]. Enfin, les masques perceptuels jouent un rôle dans la robustesse et la sécurité. Par exemple, un tatouage respectant les propriétés spectrales de \mathbf{x} sera plus robuste à la compression [20] ou au débruitage [6].

Cet article s'intéresse à la construction et à l'évaluation des masques perceptuels dans le cas particulier des images numériques. Dans la partie 2, nous présentons les masques psychovisuels classiques utilisés en image, ainsi que plusieurs mesures perceptuelles objectives destinées à les évaluer. Dans la partie 3, nous proposons de nouveaux masques perceptuels fondés sur l'interpolation. Enfin, dans la partie 4, nous comparons les performances objectives des masques classiques et des masques proposés, combinés à la technique DS.

2 Masques et distances classiques en tatouage d'images

2.1 Masques psychovisuels

De nombreux masques psychovisuels ont été proposés pour le tatouage d'images. Ils utilisent des propriétés empiriques du Système Visuel Humain (SVH), combinées avec une analyse statistique. Une approche classique est d'utiliser des techniques de codage existant en compression d'image. On distingue deux types de masques. Les masques spatiaux sont calculés à par-

tir de la luminance et favorisent les contours et les régions de forte variance locale. Ces masques conduisent souvent à des tatouages passe-haut, ce qui peut nuire à la robustesse. Le masquage de contour n'est efficace que si le masque a la même orientation que le signal hôte. C'est pourquoi le masquage de texture est souvent privilégié. Les masques fréquentiels prennent en compte la sensibilité du SVH à une modification de chaque composante fréquentielle, à une modification de la luminance par rapport à un fond constant et réalisent un masquage de contraste prenant en compte la fréquence spatiale.

En tatouage d'images, les masques spatiaux les plus populaires pondèrent selon la variance locale calculée par une Fonction de Visibilité du Bruit (NVF) [20], par le résultat d'un filtrage des dérivées premières horizontales et verticales [11] ou soustraient une partie des dérivées secondes calculées par un filtre laplacien [4]. Dans le domaine de la Transformée en Cosinus Discrète (DCT), [10] préconise un masque de contraste et de texture inspiré des travaux de Ahumada, Watson et Peterson pour la compression JPEG.

2.2 Mesures de qualité perceptuelle

Les mesures de qualité perceptuelle permettent d'évaluer objectivement la distorsion introduite entre une image originale x et une image tatouée y . Le Pic du Rapport Signal à Bruit (PSNR) est défini par $PSNR = \frac{max^2}{E[(y-x)^2]}$. Pour une image codée sur 8 bits/pixel, $max=255$. Le PSNR est la mesure la plus basique et est directement lié à la puissance de w à l'insertion. Dans la plupart des méthodes de tatouage, on peut donc contrôler le PSNR *a priori*. SSIM est une mesure fondée sur les similarités structurelles qui évalue la corrélation entre les motifs des images, indépendamment de la luminance locale et du contraste [9]. C4 est une autre mesure fondée sur l'information structurelle [8]. La mesure du Gain Composé (CG) est une généralisation de la distance de Kullback-Leibler, calculée autour des points d'intérêt de l'image [18]. La distance de Watson [24], populaire en tatouage, repose sur les expériences psychovisuelles élaborées pour la compression JPEG. Enfin, Komparator est une mesure fréquentielle fondée sur le calcul d'une carte des différences visibles [16].

La pertinence de différentes mesures dans l'application spécifique au tatouage a été récemment évaluée dans [15], par comparaison avec une évaluation subjective rigoureuse. Les auteurs concluent que les mesures C4 et surtout Komparator sont plus fiables que SSIM et le PSNR pour évaluer l'imperceptibilité d'un tatouage.

3 Masques psychovisuels fondés sur l'interpolation

L'interpolation a jusqu'ici été considérée essentiellement comme une source de nuisance pour le tatouage. En effet, que ce soit dans la construction d'un domaine transformé ou lors d'une attaque, elle gêne le décodage du tatouage. L'interpolation n'a été utilisée comme un élément positif d'un schéma de tatouage que dans un but cryptographique, ou pour modifier des objets 3D. Dans cet article, nous proposons un ensemble de masques mettant à profit les propriétés perceptuelles de l'interpolation. Ce travail diffère de celui de [7], qui ne s'appuie pas sur les al-

gorithmes de tatouage classique et où le tatouage est construit à partir d'erreurs d'interpolation exactes, non modulées.

3.1 Techniques d'interpolation d'images

L'interpolation consiste à construire un signal continu à partir de données discrètes. Le signal continu doit correspondre exactement au signal connu sur les coordonnées discrètes. En traitement d'image, le spectre des images naturelles décroît rapidement et les techniques d'interpolation passe-bas sont souvent efficaces pour préserver la qualité visuelle. Des artefacts peuvent cependant apparaître : oscillations, effets de bloc, de flou ou repliement [23]. Les techniques d'interpolation incluent, par ordre croissant de performances, les plus proches voisins, l'interpolation linéaire, par splines convolutives cubiques et par B-splines [23]. Pour chacune de ces techniques, une extension bidimensionnelle est possible par filtrage séparable sur les lignes et les colonnes, si les données sont équiréparties. L'interpolation adaptative a pour but de préserver les objets d'une image, et notamment leurs contours.

3.2 Principe des masques proposés

Les techniques d'interpolation découlent d'études théoriques sur les signaux à bande limitée, destinées à garantir la qualité perceptuelle. L'erreur d'interpolation fait également ressortir les hautes fréquences d'un signal, ce qui correspond à la construction classique d'un masque spatial. Nous proposons donc de pondérer le tatouage par une erreur d'interpolation. Les masques proposés agissent comme une sorte de détecteur de singularité dans le domaine spatial. L'avantage des masques proposés est que leur conception repose sur les techniques d'interpolation déjà existantes. Cette proposition suppose que l'ajout d'une version pondérée et surtout modulée de l'erreur d'interpolation sera peu perceptible. Nous vérifions par la suite la validité de cette hypothèse par des expérimentations. Notons que de nombreux masques perceptuels s'appuient sur une analogie avec la compression ou, dans le cas de la NVF, une analogie avec la restauration d'images pour construire un masque de texture lié à l'erreur de débruitage. Il semble donc cohérent de chercher à construire des masques perceptuels liés à une erreur de reconstruction fondée sur l'interpolation.

Dans la suite, nous étudions une application à l'image et à l'interpolation bilinéaire et par spline bicubique, ainsi qu'une extension excluant les contours.

L'erreur d'interpolation bilinéaire est $\Psi = h_{\psi} * x$, où $h_{\psi} = \frac{1}{4} \begin{bmatrix} 0 & -1 & 0 & -1 \\ -1 & 4 & -1 & 0 \\ 0 & -1 & 4 & -1 \\ -1 & 0 & -1 & 4 \end{bmatrix}$ (masque *bilin*, cf Fig. 1). Ceci définit un filtre Laplacien qui calcule les dérivées secondes horizontale et verticale de l'image. Cette interprétation est intéressante puisque le masque laplacien [4] utilise également les dérivées secondes horizontale, verticale mais aussi diagonales : $h_{\psi} = \frac{1}{9} \begin{bmatrix} -1 & -1 & -1 & -1 \\ -1 & 8 & -1 & -1 \\ -1 & -1 & 8 & -1 \\ -1 & -1 & -1 & 8 \end{bmatrix}$. Il s'agit également d'une extension aux dérivées secondes du masque de [11].

Nous proposons plusieurs solutions pour générer une erreur d'interpolation par splines. Une première solution consiste à utiliser un voisinage en échiquier comme grille d'interpolation pour les B-splines cubiques (masque *bspline*) ou convolutives (masque *cspline*), à la manière de [12]. L'interpolation bénéficie des splines cubiques, mais les performances pâtissent du support en échiquier : certains points du voisinage sont omis,

alors que tout le voisinage est connu. Une autre implantation, plus heuristique, consiste à effectuer une transformation géométrique, puis à l'inverser. Nous prendrons l'exemple d'un sous-échantillonnage (changement d'échelle de facteur 1/2) suivi d'un suréchantillonnage par interpolation par B-spline cubique (masque *scaling*, cf Fig. 2).

Les techniques d'interpolation (en particulier la séquence sous-échantillonnage / suréchantillonnage) peuvent créer des effets de flou au niveau des contours d'une image. Cependant, lorsque le masque sert à pondérer un code antipodal pseudo-aléatoire, la modulation élimine ces artefacts¹. Les techniques d'interpolation basiques comme l'interpolation bilinéaire peuvent par ailleurs créer une pixellisation sur les contours. Des techniques d'interpolation adaptative préservant les contours sont donc à envisager.

Afin d'éliminer les artefacts sur les contours, une implantation simple consiste à détecter les contours de l'image, puis à les éliminer du masque (masque *contours*, cf Fig. 3). L'exclusion des contours d'un masque perceptuel classique a été proposée notamment dans [21]. Dans l'implantation utilisée dans la partie 4, les masques de Canny et Roberts [19] sont combinés au masque *bspline*. Une étude plus fine sera nécessaire pour déterminer les détecteurs de contours et les techniques d'interpolation les plus adaptés.

Comme pour les masques classiques, plusieurs implantations de la technique DS combinée à un masque sont possibles. Le masque peut être considéré comme un simple changement d'amplitude : $\Psi = |h_{\Psi} * x|$ et le décodage effectue une corrélation par c : $\hat{b} = \text{signe}(\langle z, c \rangle)$. Une autre solution est de moduler w par $\Psi = h_{\Psi} * x$. Soit $\Psi' = h_{\Psi} * z$ l'estimation de Ψ à la réception. n et w sont supposés suffisamment faibles sous l'hypothèse d'imperceptibilité pour que $\|\Psi - \Psi'\|$ soit faible. Deux stratégies de décodage sont alors possibles : $\hat{b} = \text{signe}(\langle z, c/\Psi' \rangle)$ ou $\hat{b} = \text{signe}(\langle z, \Psi'c \rangle)$. La première technique partage les propriétés du forçage à zéro en égalisation de canal (Ψ effectuant un filtrage sans mémoire et variant dans le temps de w) : instabilité (lorsque Ψ' prend de faibles valeurs) et sensibilité au bruit. Nous conseillons la seconde technique, plus robuste.



FIG. 1 – Masque utilisant l'interpolation bilinéaire, Lena



FIG. 2 – Masque *cspline*, Lena

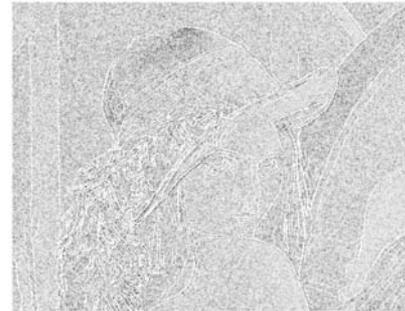


FIG. 3 – Masque *contours*, Lena

4 Etude expérimentale des performances perceptuelles

Nous comparons dans cette partie les performances perceptuelles des masques proposés avec celles des masques classiques, à l'aide de plusieurs mesures perceptuelles objectives. Les masques sont combinés à l'algorithme DS, pour un PSNR fixé *a priori* à 38 dB. Nous fournissons la moyenne des performances sur 56 images en niveaux de gris, de taille 256x256.

Les masques fondés sur l'interpolation, ainsi que les masques spatiaux [4],[11] sont d'implantation simple. Le masque NVF utilise un principe général inspiré du débruitage et pouvant être appliqué à divers domaines (spatial, Fourier, ondelettes...) ou modèles statistiques. Dans cet article, nous utilisons le masque NVF adapté au domaine spatial pour un modèle non-stationnaire gaussien, avec le paramètre $D = 100$ et un seuil $S_1 = 0.2$ [20]. Les performances perceptuelles de ce masque sont donc améliorables en utilisant d'autres domaines, ainsi qu'en adaptant empiriquement les paramètres à chaque image. Enfin, l'implantation utilisée pour le masque DCT est inspirée de celle de la distance de Watson dans le logiciel Checkmark [1].

Selon les résultats de la Table I, les masques fondés sur l'interpolation fournissent de bonnes performances en comparaison avec celles des masques spatiaux [4],[11]. Le masque *scaling* offre les meilleures performances si l'on se réfère aux mesures SSIM et CG. Le masque *contours* améliore les résultats du masque *bspline* pour toutes les mesures considérées, à l'exception de SSIM. Les masques NVF et DCT, et dans une moindre mesure [11], sont cependant les seuls à améliorer la méthode initiale (DS sans masque) quelle que soit la mesure. De plus, le masque DCT [10] présente les meilleurs résultats pour les mesures habituellement jugées les plus pertinentes

¹Il s'agit d'une propriété classique des masques perceptuels en tatouage, par exemple lors de l'utilisation du masque laplacien.

(Watson et Komparator). Les masques fondés sur l'interpolation doivent donc être améliorés, ce qui est possible en variant la technique d'interpolation. On remarquera que les mesures "spatiales" (SSIM, CG) avantagent les masques spatiaux par rapport aux masques fréquentiels et inversement. La présence de ces quelques contradictions entre les résultats des différentes mesures encourage la mise en œuvre d'un protocole d'évaluation subjective, ainsi que d'études comme celle de [15].

D'autre part, tous les masques étudiés peuvent générer des artefacts visuels, qui échappent aux mesures objectives. Ainsi, le masque DCT génère des effets de blocs gênants, par exemple au niveau d'un logo [2]. Si le seuil S_1 du masque NVF n'est pas réglé empiriquement pour chaque image, un effet de brouillage similaire à DS sans masque peut apparaître sur les zones d'intensité constante. Enfin, l'interpolation génère des artefacts au niveau des contours et des textures (la modulation corrige l'effet de flou). Les artefacts les plus visibles se situent sur des textures très régulières comme les toits d'une maison. Il serait donc intéressant de tenir compte de l'orientation des textures. Le masque *contours* améliore la qualité perceptuelle subjective, mais quelques artefacts peuvent subsister si des contours ont échappé au détecteur. Enfin, un effet "poivre et sel" peut apparaître si le masque répartit le tatouage sur un nombre trop faible de pixels dans l'image. Un seuil similaire à celui de la NVF pourrait donc être envisagé. Une évaluation subjective rigoureuse s'avère à nouveau nécessaire pour mettre au point les masques perceptuels.

	SSIM ¹	C4 ¹	CG ²	Komp. ²	D_W ²
DS	0.947	0.95	437	364	705
DS+NVF	0.976	0.954	393	286	616
DS+DCT	0.979	0.96	341	219	12
DS+[4]	0.990	0.95	316	354	715
DS+[11]	0.990	0.953	355	309	692
DS+ <i>bilin</i>	0.990	0.954	312	357	709
DS+ <i>bspline</i>	0.989	0.953	291	395	756
DS+ <i>cspline</i>	0.990	0.953	302	378	727
DS+ <i>scaling</i>	0.990	0.95	229	370	713
DS+ <i>contours</i>	0.986	0.956	237	387	708

TAB. 1 – Comparaison des mesures perceptuelles objectives, PSNR=38 dB¹ valeur optimale : 1, ² val. opt. : 0

5 Conclusion

Nous avons proposé dans cet article une nouvelle famille de masques perceptuels spatiaux pour le tatouage numérique. Ces masques pondèrent le tatouage par une erreur d'interpolation. Ils sont donc très simples de construction, et bénéficient des études sur l'interpolation. L'approche proposée est cohérente avec les masques spatiaux existants. De plus, la technique est indépendante du type de document. Il serait donc intéressant d'étudier les performances pratiques de tels masques pour le son et la vidéo. Cet article s'est concentré sur une application à l'image et à des techniques d'interpolation simples (interpolation bilinéaire et par spline bicubique). D'autre part, le principal défaut des masques spatiaux est de trop mettre en valeur les contours de l'image. Une application à une technique d'interpolation préservant les contours a donc été étudiée, qui amé-

liore la qualité perceptuelle subjective et objective. La seconde contribution de cet article est la mise en œuvre d'un protocole d'évaluation de la qualité perceptuelle des masques psychovisuels en tatouage, à l'aide de mesures objectives. Les performances des masques fondés sur l'interpolation restent du même ordre de grandeur que celles des masques spatiaux classiques. L'étude confirme également les bonnes performances des masques classiques DCT et NVF. Cette étude a cependant souligné les limitations des mesures de qualité perceptuelle objectives. Il serait intéressant de compléter l'évaluation des masques proposés et des masques classiques par une étude perceptuelle subjective, néanmoins coûteuse. Enfin, cette étude peut être étendue aux images couleur, pour lesquelles il existe des masques et des mesures spécifiques.

Références

- [1] CheckMark. <http://watermarking.unige.ch/Checkmark/>.
- [2] M.L. Miller, G.J. Doërr et I.J. Cox. Applying informed coding and embedding to design a robust high-capacity watermark. *IEEE Trans. on image processing*, 13(6) :792–807, 2004.
- [3] F. Pérez-González, C. Mosquera, M. Barni et A. Abrardo. Rational dither modulation : A high-rate data-hiding method invariant to gain attacks. *IEEE Trans. on Signal Processing*, 10(2) :3960–3975, 2005.
- [4] T. Kalker et A. Janssen. Analysis of SPOMF detection. *Proc. of IEEE conference on ICIP*, 1 :316–319, 1999.
- [5] J.J. Eggers, R. Bäuml, R. Tzschoppe et B. Girod. Scalar Costa Scheme for Information Embedding. *IEEE Trans. on Signal Proc.*, 51(4) :1003–1019, 2003.
- [6] J.K. Su et B. Girod. Power-spectrum condition for energy-efficient watermarking. *IEEE Trans. on Multimedia*, 4 :551–560, 2002.
- [7] V. Martin, M. Chabert et B. Lacaze. A novel watermarking scheme based on interpolation for digital images. *Proc. of ICASSP*, 2006.
- [8] M. Carnec, P. Le Callet et D. Barba. An image quality assessment method based on perception of structural information. *Proc. of ICIP*, 3 :185–188, 2003.
- [9] Z. Wang, A.C. Bovik, H.R. Sheikh et E.P. Simoncelli. Image quality assessment : From error visibility to structural similarity. *IEEE Trans. on Image Proc.*, 13 :600–612, 2004.
- [10] J.R. Hernández et F. Pérez-González. Statistical analysis of watermarking schemes for copyright protection of images. *IEEE Proc., Special Issue on Identification and Protection of Multimedia Information*, 87(7) :1142–1166, 1999.
- [11] M. Alvarez-Rodríguez et F. Pérez-González. Analysis of pilot-based synchronization algorithms for watermarking of still images. *Signal Processing : Image Communication*, 17(8) :611–633, 2002.
- [12] A.C. Popescu et H. Farid. Exposing digital forgeries in color filter array interpolated images. *IEEE Trans. on Signal Processing*, 53(10) :3948 – 3959, 2005.
- [13] Q. Li et I. J. Cox. Using perceptual models to improve fidelity and provide invariance to valumetric scaling for quantization index modulation watermarking. *IEEE ICASSP*, 2005.
- [14] Q. Li, G. Doërr et I. J. Cox. Spread Transform Dither Modulation using a Perceptual Model. *IEEE Workshop on Multimedia Signal Processing*, pages 98–102, 2006.
- [15] E. Marini, F. Autrusseau, P. Le Callet et P. Campisi. Evaluation of standard watermarking techniques. *SPIE Electronic Imaging, Security, Steganography, and Watermarking of Multimedia Contents IX*, 2007.
- [16] D. Barba et P. Le Callet. A robust quality metric for color image quality assessment. *Proc. of IEEE ICIP*, 1 :437–440, 2003.
- [17] P. Moulin et R. Koetter. Data-hiding codes. *Proc. of the IEEE*, 93(12) :2083–2127, 2005.
- [18] J.A. García, J. Fdez-Valdivia, X.R. Fdez-Vidal et R. Rodriguez-Sánchez. Information theoretic measure for visual target distinctness. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 23(4) :362–383, 2001.
- [19] R. Gonzalez et R. Woods. *Digital Image Processing*. Addison-Wesley Publishing Company, 1992.
- [20] S. Voloshynovskiy, A. Herrigel, N. Baumgartner et T. Pun. A stochastic approach to content adaptive digital image watermarking. *International Workshop on Information Hiding*, pages 212–236, 1999.
- [21] J.F. Delaigle, C. Devleeschouwer et B. Macq. Watermarking algorithm based on a human visual model. *Signal Processing*, 66(3) :319–336, 1998.
- [22] H.S. Malvar and D.A.F. Florêncio. Improved spread spectrum : a new modulation technique for robust watermarking. *IEEE Trans. on Signal Processing*, 51(4) :898–905, 2003.
- [23] M. Unser. Splines : A perfect fit for signal and image processing. *IEEE Signal Processing Magazine*, 16(6) :22–38, 1999.
- [24] A.B. Watson. Visually Optimal DCT Quantization Matrices for Individual Images. *Data Compression Conference*, pages 178–187, 1993.