

VERS UN TATOUAGE À ÉTALEMENT DE SPECTRE OPTIMAL UTILISANT LE SYSTÈME VISUEL HUMAIN

Stefan Winkler, Martin Kutter

Laboratoire de Traitement des Signaux
École Polytechnique Fédérale de Lausanne
1015 Lausanne, Suisse

1. INTRODUCTION

La protection des droits d'auteurs des données numériques devient de plus en plus importante suite à une distribution toujours moins contrôlable sur des réseaux tels que l'Internet. Aussi, il devient intéressant de pouvoir suivre la distribution de ces données. Une solution à ce problème sont les *tatouages numériques*. Le concept est basé sur l'idée d'insérer une information, sous forme de bits, dans les données en les modifiant de telle sorte que ces modifications ne soient pas perceptibles. Une grande collection de techniques ont déjà été proposées dans la littérature pour différents types de données, tels les images, les vidéos où même l'audio. Pour une introduction à ces sujets le lecteur peut consulter Hartung et Kutter [5].

Le fait que les tatouages numériques doivent être invisibles indique qu'il faut exploiter les caractéristiques du système visuel humain. Malheureusement ce fait a été largement négligé dans le passé. Kutter et al. [7] ont proposé de marquer la composante bleue de l'image, de sorte que la modification soit proportionnelle à la luminance locale. Des approches plus sophistiquées ont été proposées par Podilchuk et Zeng [13], basées sur la matrice de quantification de JPEG, ou par Delaigle et al. [4], basées sur le masquage d'un bruit limité en bande spectrale dans les régions texturées et contenant des contours. Malgré que toutes ces approches augmentent la performance du système, elles n'exploitent pas toutes les propriétés du système visuel humain et négligent souvent les caractéristiques du tatouage numérique.

Nous proposons une approche théorique aux problèmes de la visibilité du tatouage numérique en développant un masque de pondération basé sur les caractéristiques du système visuel humain ainsi que les propriétés statistiques du tatouage numérique. Dans la section 2, le système de tatouage numérique est introduit. Ensuite la section 3 présente les caractéristiques du système visuel humain exploitées dans ce travail. Les résultats de ce travail sont présentés dans la section 4, et les conclusions dans la section 5.

2. TATOUAGE NUMÉRIQUE

L'approche choisie pour le processus de tatouage numérique est basée sur la modulation à spectre étalé [12, 6]. Dans les deux sections qui suivent, les processus d'insertion et de détection du tatouage numérique sont présentés. Pour le moment, le système visuel humain n'est pas encore considéré.

2.1. Insertion du tatouage numérique

Notre but est d'insérer une signature binaire $B = \{b_0, \dots, b_N\}$ d'une longueur de N -bits dans une image I . Pour le moment nous ne considérons que les images comportant 256 ni-

veaux de gris différents. Le tatouage w est défini par une superposition linéaire de N fonctions bi-dimensionnelles $p_i(x,y)$, chacune représentant un bit :

$$w(x,y) = \sum_{i=1}^N p_i(x,y) \quad (1)$$

L'image tatouée \hat{I} est générée en ajoutant le tatouage w à l'image :

$$\hat{I}(x,y) = I(x,y) + w(x,y) \quad (2)$$

Les fonctions bi-dimensionnelles p_i sont définies par :

$$p_i(x,y) = b_i \alpha(x,y) \phi_i(x,y) \quad (3)$$

où b_i définit la valeur du bit i projetée de $\{0,1\}$ à $\{-1,1\}$, et $\phi_i(x,y)$ est une fonction de modulation bi-dimensionnelle pseudo-aléatoire. $\alpha(x,y)$ est une fonction de pondération ayant pour but d'adapter le tatouage numérique à l'image de telle sorte que l'énergie du tatouage soit maximisée sous la contrainte que la qualité de l'image tatouée ne soit pas inférieure à un certain seuil. Le processus de génération de α est le sujet de la section 3.

Les fonctions de modulation ϕ_i sont orthogonales, c'est-à-dire :

$$\langle \phi_i, \phi_j \rangle = \delta_{ij} \|\phi_i\|^2 \quad (4)$$

où $\langle \cdot, \cdot \rangle$ est le produit scalaire, δ_{ij} la fonction *Delta* de *Kronecker*, et $\|\cdot\|^2$ représente l'énergie de la fonction de modulation.

Il y a plusieurs moyens de générer ces fonctions de modulation. Dans les applications de tatouages numériques, il est important d'avoir un contrôle maximal sur les artéfacts. En principe, les fonctions de modulation sont définies de sorte que l'intersection des ensembles de positions (x,y) ayant une valeur $\phi_i(x,y) \neq 0$ soit vide. Cette technique nous assure que chaque valeur de l'image originale est modifiée par une fonction seulement. Les fonctions sont générées en utilisant des ensembles de positions S_i dépendant d'une clé k . Comme mentionné précédemment, l'intersection des ensembles doit être nulle, c'est-à-dire $S_i \cap S_j = \emptyset, \forall i \neq j$. Les fonctions de modulation peuvent donc être définies ainsi :

$$\phi_i(x,y) = \begin{cases} s_i(x,y) & \text{si } (x,y) \in S_i \\ 0 & \text{ailleurs.} \end{cases} \quad (5)$$

Dans notre cas, les fonctions s_i sont des fonctions pseudo-aléatoires avec une distribution bi-modale de $\{-1,1\}$, mais d'autres distributions sont aussi possibles.

Pour avoir un contrôle maximal sur les artéfacts, nous introduisons la densité D qui définit la fraction des pixels de l'image qui sont modifiés par le processus de tatouage numérique. Cette densité D est donnée par :

$$D = \frac{|\{\cup_{i=1}^N S_i\}|}{|\{S\}|} \quad (6)$$

où $|\{\cdot\}|$ est le cardinal de l'ensemble et $\{S\}$ l'ensemble universel. La probabilité de distribution des positions dans l'image est uniforme, ce qui veut dire que la probabilité d'une position de faire partie de l'ensemble S_i est $\frac{D}{N}$.

Comme déjà remarqué par plusieurs auteurs, cette méthode de tatouage numérique peut être considérée comme une modulation par étalement de spectre [14].

2.2. Détection du tatouage numérique

Pour démoduler l'information insérée dans l'image, on utilise un corrélateur linéaire qui calcule la corrélation entre l'image tatouée et la fonction de modulation. Suite au fait que les propriétés statistiques de l'image ne sont pas stationnaires, et que l'espérance n'est pas égale à zéro, nous proposons un processus de traitement préalable de l'image. Ce processus a pour but de diminuer la variance de l'image, résultant en une augmentation de la performance du système. La statistique du détecteur est donnée par :

$$r_i = \langle \epsilon(\hat{I}), \phi_i \rangle \quad (7)$$

où ϵ est la fonction de traitement préalable de l'image, et \hat{I} l'image tatouée. Dans notre cas, le processus de traitement préalable est une convolution avec un filtre non-adaptatif qui calcule une prédiction du tatouage inséré. Le filtre $H_+^{\mathcal{W}}$ d'une taille de $\mathcal{W} \times \mathcal{W}$ est défini par :

$$\mathcal{H}_+^{\mathcal{W}} : \mathbf{h}_+(x,y) = \begin{cases} -\frac{1}{2\mathcal{W}-2} & \text{si } x = 0 \text{ et } 0 < |y| \leq \frac{\mathcal{W}-1}{2} \\ -\frac{1}{2\mathcal{W}-2} & \text{si } y = 0 \text{ et } 0 < |x| \leq \frac{\mathcal{W}-1}{2} \\ 1 & \text{si } x = 0 \text{ et } y = 0 \\ 0 & \text{ailleurs.} \end{cases} \quad (8)$$

En utilisant ce filtre, qui a la forme d'un croix, la statistique de détection peut être exprimée ainsi :

$$\begin{aligned} r_i &= \langle \hat{I} * H_+^{\mathcal{W}}, \phi_i \rangle \\ &= \langle (I + w) * H_+^{\mathcal{W}}, \phi_i \rangle \\ &= \langle w * H_+^{\mathcal{W}}, \phi_i \rangle + \langle I * H_+^{\mathcal{W}}, \phi_i \rangle \end{aligned} \quad (9)$$

La valeur d'un bit caché est maintenant trouvée en observant le signe de la statistique de détection :

$$\tilde{b}_i = \begin{cases} 1 & \text{si } r_i \geq 0 \\ 0 & \text{si } r_i < 0 \end{cases} \quad (10)$$

Cette approche, basée sur le filtre de prédiction présenté ci-dessus, peut être améliorée en utilisant des filtres plus performants et adaptatifs, tels les filtres de *Wiener*.

3. MODÈLE DU SYSTÈME VISUEL

Pour bien cacher le tatouage dans l'image, il est utile d'exploiter les faiblesses du système visuel humain. Cependant, le système visuel humain est extrêmement complexe, et plusieurs de ses propriétés ne sont pas bien comprises aujourd'hui encore (c.f. [16] pour plus de détails). Nous nous concentrons ici sur deux aspects importants, qui sont la variation de la sensibilité au contraste et le phénomène de masquage.

3.1. Sensibilité au contraste

Le contraste est une mesure de la variation relative de la luminance. Malheureusement, il n’y a pas de définition du contraste qui soit appropriée pour tous les genres de stimuli visuels. Dans les expériences psycho-visuelles, on utilise souvent des motifs périodiques, par exemple des sinusoides, dont la luminance varie entre L_{\min} et L_{\max} . Le contraste de ces “gratings” est donné par la définition de Michelson :

$$C_M = \frac{L_{\max} - L_{\min}}{L_{\max} + L_{\min}}.$$

Par contre, le contraste d’un stimulus qui se compose d’un incrément ou décrement ΔL sur un fond uniforme d’une luminance L est mieux décrit par la définition de Weber :

$$C_W = \frac{\Delta L}{L}.$$

Ces deux définitions ne sont pas équivalentes : le contraste selon Michelson varie entre 0 et 1, tandis que le contraste selon Weber varie entre -1 et ∞ . Il existe d’autres définitions similaires du contraste [11], mais aucune d’entre elles ne convient à la description du contraste d’images complexes, car pour la plupart de ces définitions, il suffit de quelques points très foncés ou très clairs pour déterminer le contraste de l’image entière.

En utilisant le fait que la sensibilité au contraste du système visuel humain varie avec l’adaptation à la luminance moyenne locale, Peli [10] a proposé une mesure de contraste local limité bande en bande spectrale,

$$C_P(x,y) = \frac{\text{bp}(x,y)}{\text{lp}(x,y)},$$

où $\text{bp}(x,y)$ est la réponse passe-bande, et $\text{lp}(x,y)$ est la réponse passe-bas qui contient l’énergie au-dessous de la bande passante. Des versions modifiées de cette définition du contraste local ont été utilisées avec succès dans certains modèles de vision [3, 9, 11].

Cependant, pour une bonne concordance avec le contraste de Michelson pour le cas d’un grating sinusoidal (c’est-à-dire une réponse constante partout), le contraste doit tenir compte de la composante en-phase ainsi que de la quadrature du contraste. L’énergie des filtres analytique représente un bon moyen pour ce faire. Or, des filtres analytiques généraux bi-dimensionnels ne sont pas bien définis à cause de l’absence d’une transformation de Hilbert bi-dimensionnelle [1]. Winkler et Vandergheynst [17] décrivent la construction d’une mesure de contraste *non-orienté* en combinant les réponses de filtres analytiques passe-bande orientés $\text{bop}_i(x,y)$. La propriété de non-orientation est importante pour insérer un tatouage en forme de spectre étalé, parce que son bruit blanc est non-orienté lui-même. Le contraste résultant correspond à la somme des énergies de ces filtres, normalisée par l’image passe-bas :

$$C_E(x,y) = \frac{\sqrt{2 \sum_i |\text{bop}_i(x,y)|^2}}{\text{lp}(x,y)}. \quad (11)$$

La sensibilité au contraste est définie comme l’inverse du seuil de la visibilité, c’est-à-dire le contraste minimal nécessaire pour qu’on puisse percevoir le stimulus. Elle dépend non seulement fortement de la fréquence spatiale des stimuli (voir figure 1), mais aussi de leur couleur et de leur intensité. En général, notre sensibilité est la plus haute pour des stimuli lumineux d’une fréquence basse à moyenne.

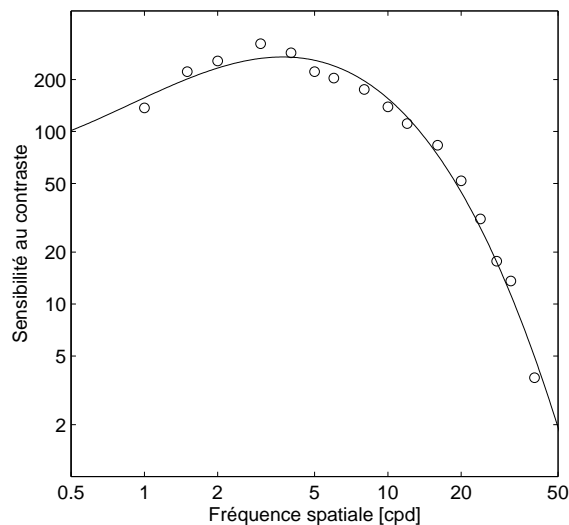


FIG. 1 – La sensibilité au contraste en fonction de la fréquence spatiale selon [15].

3.2. Masquage

Le masquage est un phénomène visuel très important, parce qu’il décrit les interactions entre stimuli. On parle de masquage quand un stimulus, qui est visible seul, ne peut pas être perçu à cause de la présence d’un autre stimulus.* Dans le contexte du tatouage numérique, nous considérons le tatouage comme le stimulus qui est masqué par l’image originale, qui fonctionne donc comme fond. Ce masquage explique pourquoi le bruit du tatouage est désagréable dans certaines régions d’une image tandis qu’il n’est guère perceptible ailleurs.

En d’autres termes, le masquage augmente le seuil de la visibilité en fonction du contraste du masque C_M . Un modèle simple pour le masquage peut être formulé comme suit (voir figure 2):

$$k(C_M) = \begin{cases} 1 & C_M < C_S \\ (C_M/C_S)^\epsilon & \text{ailleurs.} \end{cases} \quad (12)$$

Le seuil de la visibilité du tatouage sans aucun masque, C_0 , est multiplié par ce facteur k . Normalement, C_S est proche du seuil du contraste du masque lui-même. ϵ dépend du type de stimuli et de leurs similarités; en général, $0.6 \leq \epsilon \leq 1$.

En raison de la structure multi-canaux du système visuel humain, le masquage est le plus fort quand les stimuli ont des caractéristiques similaires, c’est-à-dire des fréquences, des orientations et des couleurs semblables. En général, le masquage entre des stimuli occupants des canaux différents est plus faible [8].

3.3. Application au tatouage

Pour intégrer les deux phénomènes mentionnés ci-dessus dans un système de tatouage, il faut d’abord calculer le contraste local de l’image. La modélisation se complique parce que le tatouage en question représente un bruit blanc, et par conséquent comprend toutes les fréquences. Vu que la sensibilité au contraste est la plus haute pour les hautes fréquences, nous concentrons nos efforts sur cette région, car c’est là où l’on a le plus de marge pour cacher des informations dans l’image.

* Parfois l’effet opposé, la facilitation, se produit : un stimulus qui n’est pas visible seul peut être perçu à cause de la présence d’un autre.

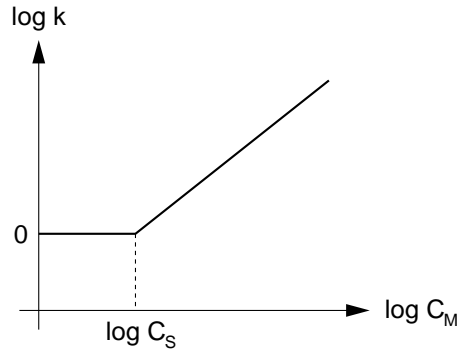


FIG. 2 – Le modèle de masquage selon l'éq. 12.

Pour cette raison, le contraste local limité en bande C_E est calculé à partir de la bande passe-haut selon l'éq. 11. Les filtres spécifiques employés dans notre l'implémentation s'appuient sur la transformée de Cortex de Watson, modifiée par Daly [3]; le filtre passe-bas est un filtre gaussien, et les autres filtres sont formés par des fenêtres de Hanning dans les directions radiales et angulaires (voir figure 3).

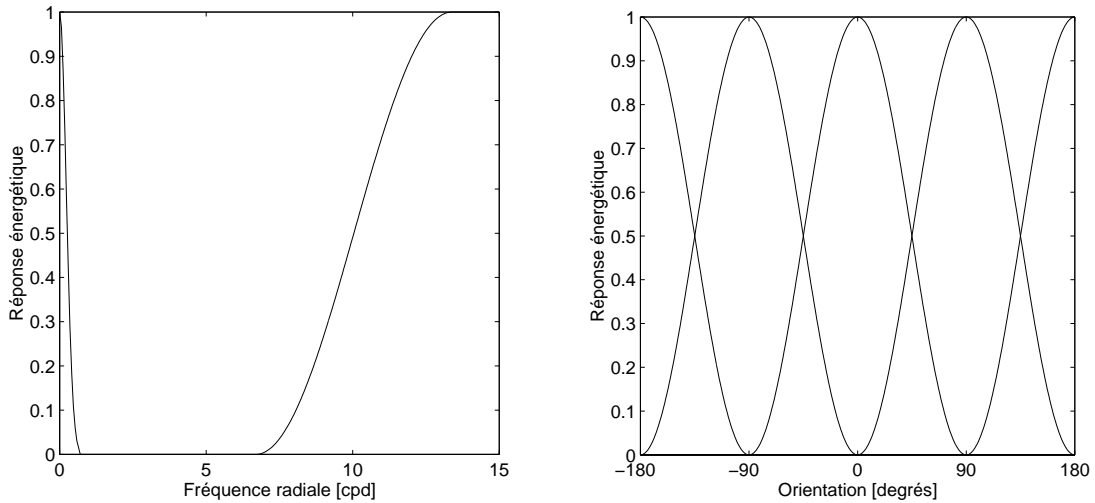


FIG. 3 – Les filtres passe-bas et passe-haut radiaux (à gauche) et la sélectivité en orientation de ce dernier (à droite).

Ensuite, l'augmentation du seuil de la visibilité $k(x,y)$ est calculée à l'aide du contraste C_E selon l'éq. 12. Finalement, le tatouage (jusqu'à maintenant représenté par une série de $-1, 0$ et 1) est modulé par la fonction $\alpha(x,y)$, qui tient compte du contraste local du bruit :

$$\alpha(x,y) = C_0 \cdot k(x,y) \cdot lp(x,y). \quad (13)$$

Cette fonction est illustrée dans la figure 4 pour l'image *bike*.

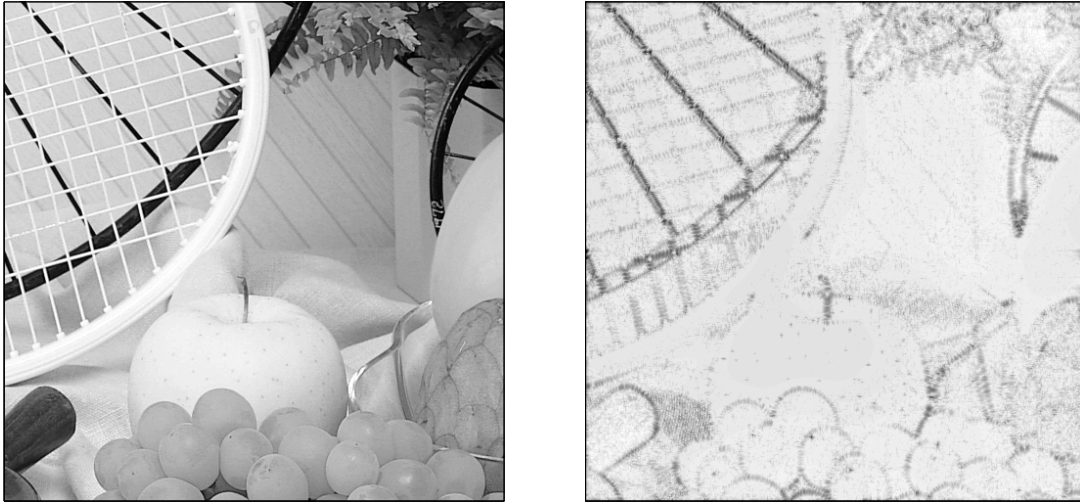


FIG. 4 – Une image de test et le masque de pondération $\alpha(x,y)$ de son tatouage selon l'éq. 13 (le noir représente des amplitudes fortes).

4. RÉSULTATS

4.1. Paramètres

Le modèle du système visuel humain comme décrit ci-dessus contient quelques paramètres qui doivent être choisis en effectuant des expériences psycho-visuelles afin d'obtenir une bonne modélisation. Malheureusement, il n'y a pas de données expérimentales pour la sensibilité au bruit blanc et son masquage dans la littérature scientifique. Nous sommes en train de nous livrer aux expériences correspondantes. Pour cette raison, dans l'implémentation présente, tous les paramètres ont été choisis en extrapolant des données psycho-visuelles comparable. Pour ne pas surmener le modèle, cette extrapolation est plutôt conservatrice. À cause de cela, nous sommes convaincus que la performance de la méthode décrite ci-après peut être encore améliorée.

Le seuil du contraste du masque a été fixé à $C_S = 0.01$, ce qui correspond au seuil approximatif pour des fréquences vers 15 cpd^\dagger (voir figure 1), c'est-à-dire au début de la bande des filtres utilisés. Pour estimer le seuil du contraste de bruit blanc, nous avons effectué quelques tests visuels, selon lesquels $C_0 = 0.015$. Bâtissant sur des expériences avec un bruit en bande-passante [2], l'exposant de la fonction de masquage a été fixé à $\epsilon = 0.6$. Naturellement, ces valeurs peuvent être augmentées pour produire un tatouage qui n'est plus invisible, mais qui sera plus robuste en contre-partie. Cependant, le modèle de la sensibilité du masquage est basé sur des expériences proche du seuil de visibilité, donc c'est là où il est le plus fiable.

Pour le choix de la densité, nous avons pris en compte la visibilité du tatouage, la longueur en bits du tatouage, et la robustesse à la compression JPEG. Pour une longueur de 256-bits, une densité de $D = 0.8$ est appropriée pour une robustesse à la compression JPEG acceptable. La taille du filtre pour le traitement préalable dans la détection du tatouage a été fixée à $\mathcal{W} = 5$.

4.2. Évaluation et comparaison

Pour évaluer le système proposé ci-dessus, nous avons procédé à deux séries de tests. Dans la première série, nous avons tatoué l'image originale en utilisant le nouveau masque de pondération.

\dagger cycles par degrés

Masque	<i>lena</i>	<i>bike</i>
Constant	36.64dB	35.16dB
Linéaire	37.52dB	35.61dB
Linéaire passe-bas	37.96dB	35.64dB
Proposé	38.97dB	36.11dB

TAB. 1 – Valeurs de PSNR pour différents masques de pondération.

Ensuite la robustesse de l’image été testée à l’aide de la compression JPEG. Un fois que la valeur de la compression maximale a été trouvée, l’image originale a été à nouveau tatouée en utilisant trois masques de pondération différents, en ajustant l’échelle du masque pour obtenir la même robustesse. Les trois masques de pondération sont les suivants :

1. constant : $\alpha_c(x,y) = \pm\beta$
2. linéaire : $\alpha_l(x,y) = \beta I(x,y)$
3. linéaire passe-bas : $\alpha_{lp}(x,y) = \beta lp(x,y)$

où β définit l’amplitude du tatouage et $lp(x,y)$ est la version filtrée passe-bas de l’image. Les tests ont été répétés sur deux images, *bike* et *lena*, chacune d’une taille de 512×512 . Pour *bike*, la qualité minimale de compression JPEG pour une détection sans faute est 70%, pour *lena* elle est de 68%. Les valeurs de PSNR pour une même robustesse avec les autres masques sont indiquées dans le tableau 1. Ces valeurs indiquent que le nouveau masque est très efficace, car la qualité est supérieure dans tous les cas, ce qui veut dire que le masque de pondération proposé exploite les régions où le tatouage est moins visible. Des tests visuels sur les images tatouée ont confirmé ces résultats.

Dans la deuxième série des tests, les images ont été tatouées de sorte que la distorsion dans les images soit la même pour les quatre masques (32.44 dB pour l’image *lena*). Ensuite les images ont été présentées à des sujets pour évaluer la qualité des images. Ces tests subjectifs ont clairement montré que le masque de pondération produit des images tatouées de meilleure qualité. Ce résultat montre que le masque de pondération permet l’insertion d’un tatouage avec plus d’énergie en préservant la qualité visuelle de l’image, entraînant un tatouage plus robuste.

5. CONCLUSIONS

Un nouveau masque de pondération a été présenté, basé sur un modèle du masquage et de la sensibilité du système visuel humain. Ce masque permet d’adapter un tatouage numérique, sous forme de spectre étalé, aux images, avec pour but de minimiser la distorsion visuelle. Les premiers résultats ont clairement démontrés que le masque satisfait à la théorie, et permet l’insertion d’un tatouage d’une énergie plus élevée dans une image ayant la même distorsion visuelle de l’image que pour d’autres techniques. L’augmentation de l’énergie du tatouage augmente ainsi sa robustesse aux divers traitements de l’image. Pour la recherche à venir, nous avons l’intention d’effectuer des tests plus extensifs sur la qualité et la robustesse de cette méthode, et d’essayer de l’améliorer comme indiqué.

Remerciements

Nous tenons spécialement à remercier notre collègue Pascal Fleury pour avoir soigneusement amélioré le français de cet article, ainsi que pour ses apports techniques et ses commentaires constructifs.

Références

- [1] J.-P. Antoine et al.: “Directional wavelets revisited: Cauchy wavelets and symmetry detection in patterns.” *Appl. Comp. Harm. Anal.* 1999, accepté pour publication.
- [2] S. Comes: *Les Traitements Perceptifs des Images Numérisées*. Thèse de doctorat, Université Catholique de Louvain, Belgique, 1995.
- [3] S. Daly: “The visible differences predictor: An algorithm for the assessment of image fidelity.” dans *Digital Images and Human Vision*, ed. A. B. Watson, 179–206, MIT Press, 1993.
- [4] J. F. Delaigle et al.: “Watermarking algorithm based on a human visual system.” *Signal Processing* **66**, 319–335, 1998.
- [5] F. Hartung, M. Kutter: “Multimedia watermarking techniques.” *Proc. IEEE* 1999, accepté pour publication dans: Special Issue on Identification and Protection of Multimedia Information.
- [6] M. Kutter: “Watermarking resisting to translation, rotation, and scaling.” dans *Multimedia Systems and Applications*, eds. A. G. Tescher, B. Vasudev, V. M. Bove, B. Derryberry, vol. 3528, 523–431, SPIE, San Jose, CA, USA, 1998.
- [7] M. Kutter et al.: “Digital watermarking of color images using amplitude modulation.” *Journal of Electronic Imaging* **7**(2), 326–332, 1998.
- [8] G. E. Legge, J. M. Foley: “Contrast masking in human vision.” *J. Opt. Soc. Am.* **70**(12), 1458–1471, 1980.
- [9] J. Lubin: “A visual discrimination model for imaging system design and evaluation.” dans *Vision Models for Target Detection and Recognition*, ed. E. Peli, 245–283, World Scientific Publishing, 1995.
- [10] E. Peli: “Contrast in complex images.” *J. Opt. Soc. Am. A* **7**(10), 2032–2040, 1990.
- [11] E. Peli: “In search of a contrast metric: Matching the perceived contrast of Gabor patches at different phases and bandwidths.” *Vision Res.* **37**(23), 3217–3224, 1997.
- [12] R. L. Pickholz et al.: “Theory of spread-spectrum communications—a tutorial.” *IEEE Trans. Communications* **30**(5), 855–884, 1982.
- [13] C. Podilchuk, W. Zeng: “Perceptual watermarking of still images.” dans *Proc. Workshop on Multimedia Signal Processing*, Princeton, New Jersey, USA, 1997.
- [14] J. Smith, B. Comiskey: “Modulation and information hiding in images.” dans *Lecture Notes in Computer Science: Information Hiding*, vol. 1174, 207–226, Springer, 1996.
- [15] F. L. van Nes, M. A. Bouman: “Spatial modulation transfer in the human eye.” *J. Opt. Soc. Am.* **57**(3), 401–406, 1967.
- [16] S. Winkler: “Issues in vision modeling for perceptual video quality assessment.” *Signal Processing* **78**(2), 1999, accepté pour publication.
- [17] S. Winkler, P. Vandergheynst: “Band-limited local contrast.” dans *Proc. ICIP*, Kyoto, Japon, 1999, accepté pour publication.