Attaque de tatouage d'image

fondée sur estimation bayésienne non-linéaire non-paramétrique dans le domaine des ondelettes

Larbi Boubchir, Ayoub Otmani et Nadia Zerida

GREYC UMR 6072 CNRS - ENSICAEN/Université de Caen
larbi.boubchir@greyc.ensicaen.fr
www.greyc.ensicaen.fr/∼boubchir/

Plan

Plan

Plan

Plan

Plan

Plan

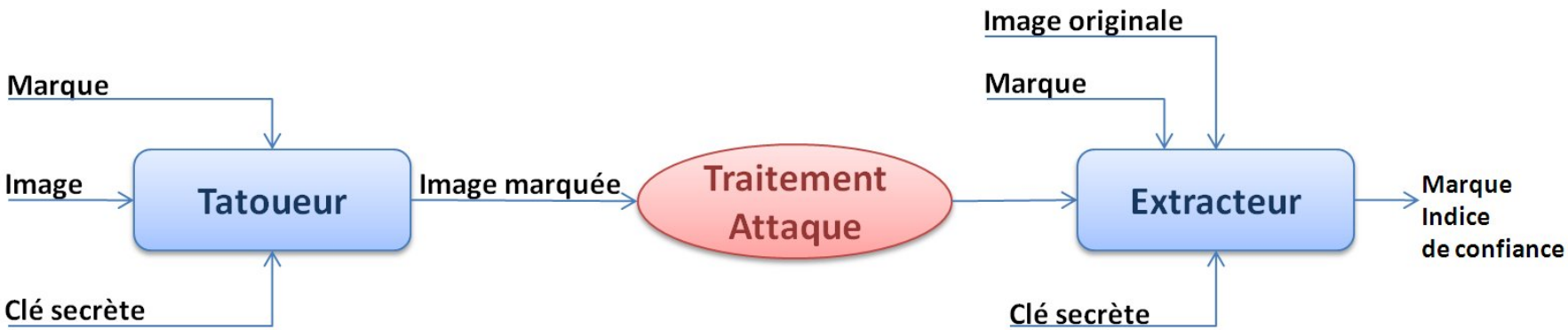
Problématique de tatouage d'image

Tatouage d'image (watermarking)

Rajouter dans une image une marque qui doit être :

1. Imperceptible:
   - invisible.
   - ne pas détériorer l'image.
2. Spécificité:
   - suffisamment spécifique pour être clairement identifiable lors de son extraction.
3. Robuste:
   - décelable même après un traitement du contenu de l'image (résultant d'une attaque, etc.).
   - robuste contre les dégradations (e.g. compression, ré-échantillonnage, déformation non-linéaires)

Chaîne de marquage



- Tatoueur: algorithme qui incruste la marque dans l'image.
- Extracteur: algorithme d'extraction qui retrouve la marque.
  1. Type I: détermine lui-même la marque.
  2. Type II: nécessite de savoir à l'avance la marque.
     Réponse : oui/non ou indice de confiance (entre 0 et 1).

Plan

Attaques de tatouage d'image

Attaque

Modification intentionnelle du contenu tatoué.

Deux types d'attaques:

1. Attaques liées au signal/image
   - s'attaquent au signal lui-même pour retirer la marque ou en ajoutant une sur-marque qui va masquer la première.
   - combinent des déformations géométriques imperceptiblement (e.g. rotations, filtrage, compression avec perte).
2. Attaques de nature cryptologique
   - déterminent des informations sur les clés à partir des images marquées.
   - cherchent à utiliser ces informations pour attaquer d'autres images.

Attaques de filtrage

Attaques de filtrage

Chercher à effacer le tatouage (marque).

Exemples

- compression (JPEG, JEPG 2000, etc.)
- ajout de bruit.
- restauration/débuitage.
- amplification.
- effets des logiciels de retouche.
- etc.

Attaque de débruitage

Restauration par transformée multi-échelle (ondelettes).

Plan

Séries d'ondelettes

- Ondelettes correspond à la projection successive d'un signal $\mathbf{Y}$ sur une base orthonormale, $L^2(\mathbb{R})$, de fonctions formées par dilatation et translation à partir d'une fonction d'échelle (ondelette père), Φ, et d'une ondelette mère Ψ.

$$\Psi_{j,k}(t) = \frac{1}{\sqrt{2^j}}\Psi(2^{-j}t - k)$$

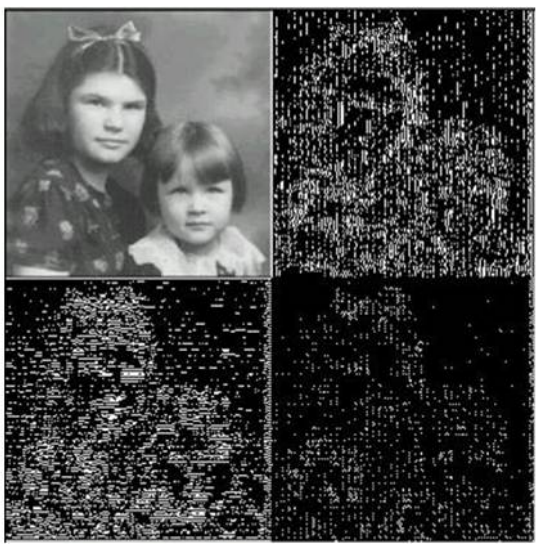
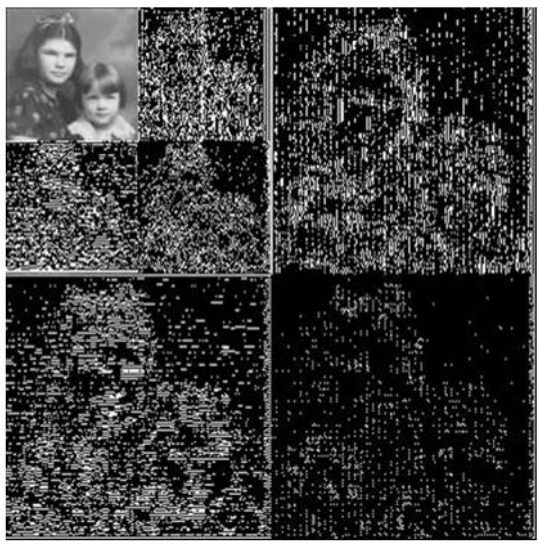
$$\Phi_{j,k}(t) = \frac{1}{\sqrt{2^j}}\Phi(2^{-j}t - k),\ t \in [0,1]$$

- Décomposition en séries d'ondelettes
  la décomposition du signal $\mathbf{Y}$ au niveau de résolution j est réalisé par une combinaison linéaire de fonction d'échelle Φ et d'ondelette Ψ

$$\mathbf{y(t)} = \sum_{k=0}^{2^{j_0}-1} c_{j_0,k}\Phi_{j_0,k}(t) + \sum_{j=j_0}^{\infty}\sum_{k=0}^{2^j-1} d_{j,k}\Psi_{j,k}(t) \text{ où } j_0 \geqslant 0, t \in [0,1]$$

$$c_{j_0,k} =< \mathbf{y}, \Phi_{j_0,k} > \text{ et } d_{j,k} =< \mathbf{y}, \Psi_{j,k} >$$

Transformée en ondelette séparable 2D

Relation entre les coefficients d'ondelettes

$$\sum_k c_{j+1,k}\Phi_{j+1,k}(t) = \sum_k c_{j,k}\Phi_{j,k}(t) + \sum_k d_{j,k}\Psi_{j,k}(t)$$

Plan

Problème de la régression nonparamétrique

Régression nonparamétrique=débruitage

recouvrer une fonction inconnue $\mathbf{X}$ sans spécification d'un modèle explicite.

$\mathbf{Y} = \mathbf{X} + \epsilon$ où $\epsilon \sim \mathcal{N}(0, \sigma^2)$

- $\mathbf{Y}$ l'image bruitée.
- $\mathbf{X}$ l'image originale.
- ϵ bruit.

Débruitage multi-échelle des images

- Modèle d'observation: $\mathbf{Y} = \mathbf{X} + \epsilon$ où $\epsilon \sim \mathcal{N}(0, \sigma^2)$
- Application de TOD

$$\begin{cases} c_{mn} = a_{mn} + \epsilon_{mn} \\ d^{oj}_{mn} = s^{oj}_{mn} + \epsilon_{mn} \end{cases} \quad j = J_c, ..., J-1 \; ; m, n = 0, 1, ..., 2^j - 1$$

- Argument: **la parcimonie**
  - Les coefficients significatifs contribuent au signal/image à recouvrer $\mathbf{X}$.
  - Les coefficients de faibles valeurs sont essentiellement dus au bruit.

Comment distinguer les coefficients significatifs de ceux dus au bruit?

- conserver les coefficients c_{mn}, relatifs aux composantes basses fréquences $\Longrightarrow$ caractéristiques du signal original.
- une sélection judicieuse des coefficients $d^{oj}_{mn} \Longrightarrow$ opérateurs de contraction ou de seuillage

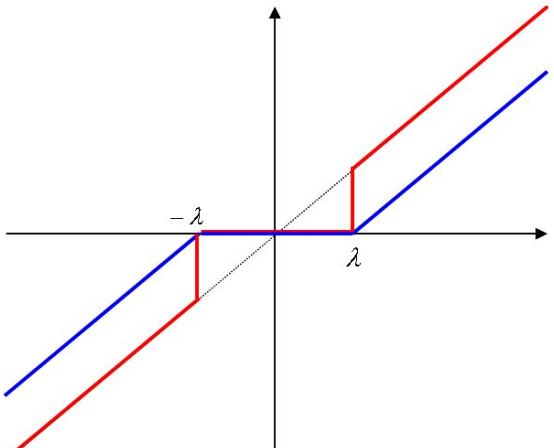
Processus de débruitage

$$Y \xrightarrow{\Phi^T} \{c_{mn}, d^{oj}_{mn}\} \xrightarrow{\text{estimateur non-linéaire } \delta} \{c_{mn}, \delta(d^{oj}_{mn})\} \xrightarrow{\mathcal{R}} \hat{X}$$

Approches de débruitage multi-échelle

1- Débruitage classique

- Travaux fondateurs de Donoho & Johnstone dans le domaine des ondelettes.
- **Idée de base:** appliquer un estimateur non-linéaire (e.g. opérateur de seuillage) sur les coefficients.
- Extraction des coefficients de détail significatifs par comparaison avec un paramètre de seuillage $\lambda > 0$.
- **Exemples d'opérateur d'estimation:** seuillage doux et dur.

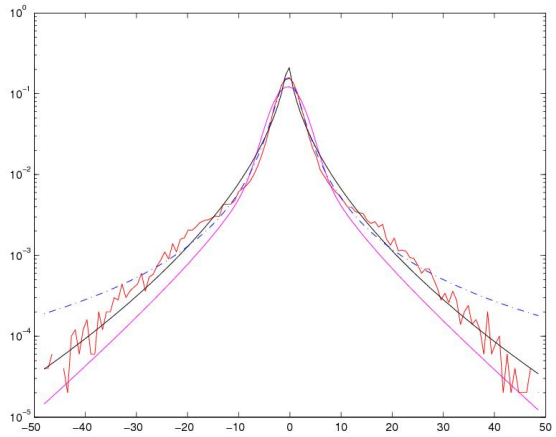


2- Débruitage bayésienne

- Modèle statistique *a priori*: gaussienne, DGG, α-stable, etc.
- Meilleurs performances.

Difficultés:

- Problème de modélisation des queues de distribution lourdes.
- Aucune forme analytique pour le débruiteur.
- Problème d'estimation des hyperparamètres.



Plan

Estimation bayésienne

- Approche bayésienne: l'image est vue comme les réalisations d'une variable aléatoire ou d'un champ aléatoire.
- Modèle de dégradation: $Y = X + \epsilon$
  - La loi a priori $p(x|\theta_1, H)$
  - La loi conditionnelle $p(y|x, \theta_2; H)$
  - La loi marginale $p(y|\theta; H) = \int p(y|x)p(x)dx$
    où $\theta = (\theta_1, \theta_2)$ est l'ensemble des hyperparamètres
  - Règle de Bayes:

    la loi *a posteriori* $p(x|y, \theta; H) = \frac{p(y|x,\theta_2;H)p(x|\theta_1;H)}{p(y|\theta;H)}$

Le problème de l'estimation bayésienne revient à trouver l'opérateur $\mathcal{D}$ tels que:

$$\hat{x} = \arg\inf_{\mathcal{D}\in\mathcal{O}_n} R(x, \hat{x} = \mathcal{D}y) = E_{Y,X}[L(x, \mathcal{D}y)]$$

Coût $L(x, \hat{x})$	Estimateur	Implémentation
$0-1$	MAP	optimisation
L_2	ECP	intégration

Plan

Modèle a priori: Formes K de Bessel

Définition

La PDF d'une distribution BKF est donnée par [Grenander (2001)]

$$f(x;p,c) = \frac{1}{\sqrt{\pi}\Gamma(p)} \left(\frac{c}{2}\right)^{-\frac{p}{2}-\frac{1}{4}} \left|\frac{x}{2}\right|^{p-\frac{1}{2}} K_{p-\frac{1}{2}}\left(\sqrt{\frac{2}{c}}|x|\right)$$

où K_ν est la fonction de Bessel modifiée.

Propriétés

- $p > 0$ et $c > 0$ représentent les paramètres de forme et d'échelle.
- Loi est unimodale, symétrique autour du mode, leptokurtique, et à queues lourdes.

Débruiteur bayésien MAP

- Modèle d'observation: $\mathbf{Y} = \mathbf{X} + \epsilon$
- Les coefficients de détail à chaque échelle et à chaque orientation du signal à estimer $\mathbf{X}$ suivent une distribution BKF: $\mathbf{x} \sim \text{BKF}(p, c)$
- Cas du bruit additif gaussien blanc: $\mathbf{Y}|\mathbf{X} \sim \mathcal{N}(\mathbf{x}, \sigma_\epsilon^2)$

Débruiteur MAP BKF

La forme analytique de l'estimateur MAP est donnée par l'expression suivante:

$$\hat{\mathbf{x}}_{MAP}(\mathbf{y}) = \begin{cases} 0 & |\mathbf{y}| \leq \nu \\ \frac{\text{sgn}(\mathbf{y})}{2}\left(\left(|\mathbf{y}| - \sqrt{\frac{2}{c}}\sigma_\epsilon^2\right) + \sqrt{\left(|\mathbf{y}| - \sqrt{\frac{2}{c}}\sigma_\epsilon^2\right)^2 + 4(p-1)\sigma_\epsilon^2}\right) & |\mathbf{y}| > \nu \end{cases}$$

où $\nu = \sqrt{2}\sigma_\epsilon\left(\sqrt{2(1-p)} + \frac{\sigma_\epsilon}{\sqrt{c}}\right)$.

L. Boubchir and J. Fadili
"Bayesian Denoising Based on The MAP Estimation in Wavelet-domain Using Bessel K Form Prior", *ICIP 2005.*

Estimation des hyperparamètres

L'ensemble des hyperparamètres: $\theta = \{p, c, \sigma_\epsilon\}$

- Méthode des cumulants

$$\hat{p} = \frac{3\kappa_2^2}{\kappa_4}, \quad \hat{c} = \frac{\kappa_2}{\hat{p}}$$

où κ_i est le cumulant d'ordre i.

- Estimateur MAD [Donoho & Johnston (1994)]

$$\hat{\sigma}_\epsilon = \frac{\text{MAD}(\mathbf{y}_{mn}^{HH_1})}{0.6745}$$

Plan

Modèle d'observation

Tatouage d'image

$\mathbf{Y}_{mn} = \mathbf{X}_{mn} + \mathbf{W}_{mn}$ où $m, n = 0, \ldots, N-1$ avec $N = 2^J$

- $\mathbf{Y}_{mn}$ l'image tatouée.
- $\mathbf{X}_{mn}$ l'image originale.
- $\mathbf{W}_{mn}$ la marque.

Régression nonparamétrique=débruitage

$\mathbf{Y}^{\mathbf{N}}_{mn} = \mathbf{X}_{mn} + \epsilon_{mn}$ où $\epsilon_{mn} \sim \mathcal{N}(0, \sigma^2)$

- $\mathbf{Y}^{\mathbf{N}}_{mn}$ l'image tatouée bruitée.
- $\mathbf{X}_{mn}$ l'image originale.
- ϵ_{mn} bruit blanc gaussien.

Plan

Étape 1

Algorithm 1

initialiser: $\sigma = 1$
répéter
  $\sigma = \sigma + 1$
  $\mathbf{Y}^{\mathbf{N}}_{mn} = \mathbf{X}_{mn} + \epsilon_{mn}$ où $\epsilon_{mn} \sim \mathcal{N}(0, \sigma^2)$
  Applique TOD à l'image tatoué bruitée $\mathbf{Y}^{\mathbf{N}}_{mn}$
  A chaque échelle et à chaque orientation appliquer le débruiteur bayésien MAP BKF aux coefficients de détails.
  Appliquer la TODI pour reconstruire l'image débruitée $\hat{\mathbf{X}}$.
jusqu'á la marque n'est pas présente

Plan

Étape 2

Algorithm 2

initialiser: $\lambda = 1$
répéter
  pour $i = 1$ to N **faire**
    pour $j = 1$ to N **faire**
      si $(|\mathbf{D}_{ij}| > \lambda)$ **alors**
        $\mathbf{X}_{ij}^{new} = \mathbf{X}_{ij}$
      sinon
        $\mathbf{X}_{ij}^{new} = \texttt{Average}(\mathbf{Y}_{ij})$ //Pixel à la position (i, j) est remplacé par la moyenne de ses 8 voisins.
      finsi
    fin pour
  fin pour
  $\lambda = \lambda + 1$
jusqu'á la marque n'est pas présente

Plan

Contexte

Challenge "Break Our Watermarking System" 2ème édition

- Organiser par "Watermarking Virtual Laboratory" (WAVILA) du Réseau d'Excellence Européen ECRYPT, et supporter par le projet ANR Nebbiano.
- Objectif
  - Enlever la marque ou rendre la marque indétectable.
  - $PSNR(I_m, I_a) = 10 \log_{10} \left(\frac{D^2}{EQM} \right) > 20dB$
    avec D la dynamique du signal (255 pour des pixels codés sur un octet) et $EQM(I_m, I_a) = \frac{1}{M\,N} \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} (I_m(j,i) - I_a(j,i))^2$


Fall

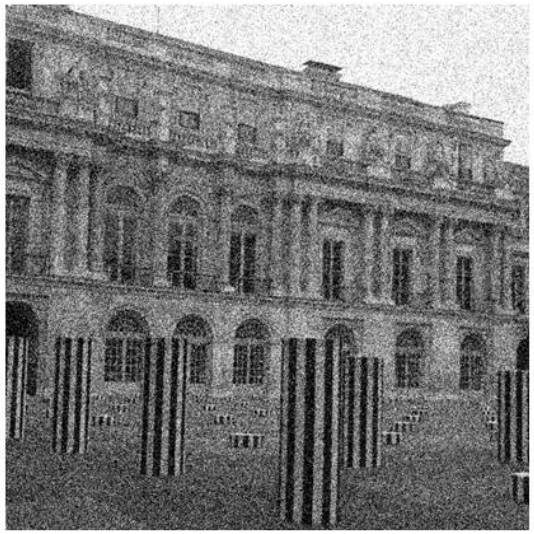

Palais Royal

Casimir

Plan

Résultats

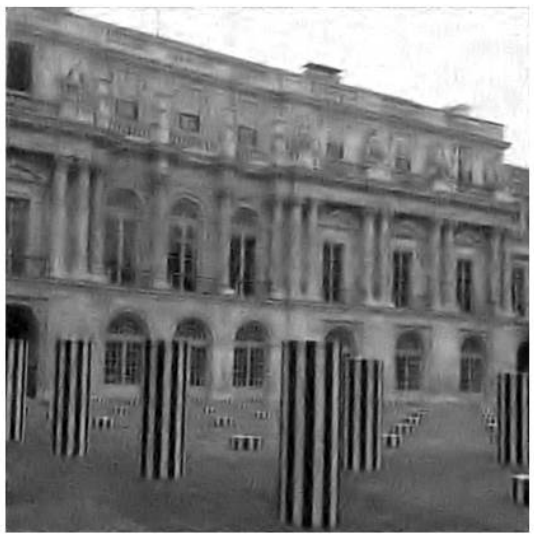
sigma= 56



sigma= 42

sigma= 73

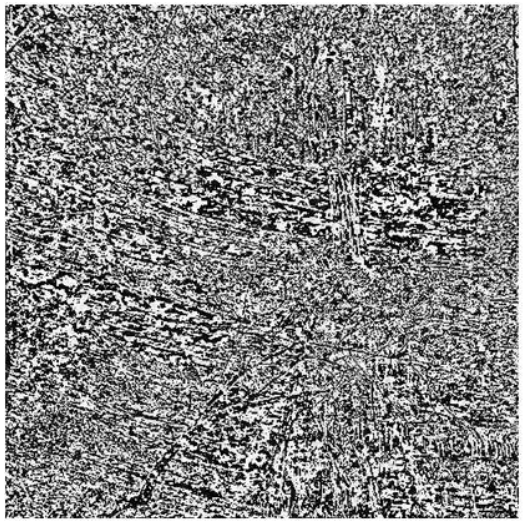
PSNR= 20.16dB



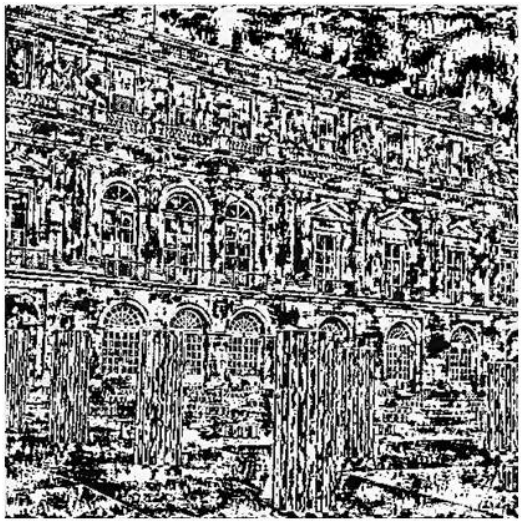
PSNR= 25.27dB

PSNR= 27.94dB

Résultats



Seuil=130



Seuil=27

Seuil=31

PSNR= 21.43dB

+1.27dB

PSNR= 27.79dB

+2.52dB

PSNR= 30.93dB

+2.99dB

Conclusion & perspectives

- Adapter le débruiteur bayésien BKF comme une attaque de filtrage dans le domaine de tatouage d'images.
- Efficacité de notre stratégie d'attaque dans le cadre du challenge BOWS-2.

Perspectives

- Mettre en place un algorithme de tatouage d'image en exploitant la propriété de la parcimonie dans le domaine des transformées multi-échelles non-orientées (e.g. ondelettes) et orientées (e.g. curvelets).

Merci de votre attention

Vos questions?