



# Détection statistique d'informations cachées dans une image naturelle à partir d'un modèle physique

Rémi Cogranne

Sous la direction de Lionel Fillatre et Florent Retrait, ICD - LM2S - UTT - UMR STMR CNRS.

Soutenance de thèse publique du 02 décembre 2011.



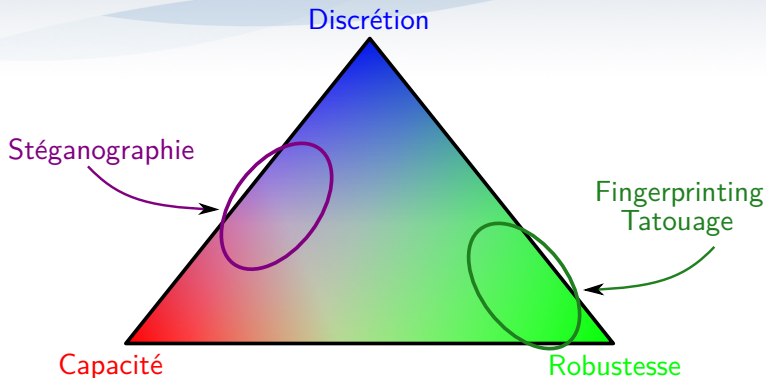
- L'ère numérique a fait émerger un besoin de communications "sécurisées".
- La **cryptographie** a pour but de **cacher le sens** d'un message mais elle est facilement décelable.
- La **stéganographie** permet de **cacher l'existence** d'un message, qui est alors difficilement décelable.



Image "saine"

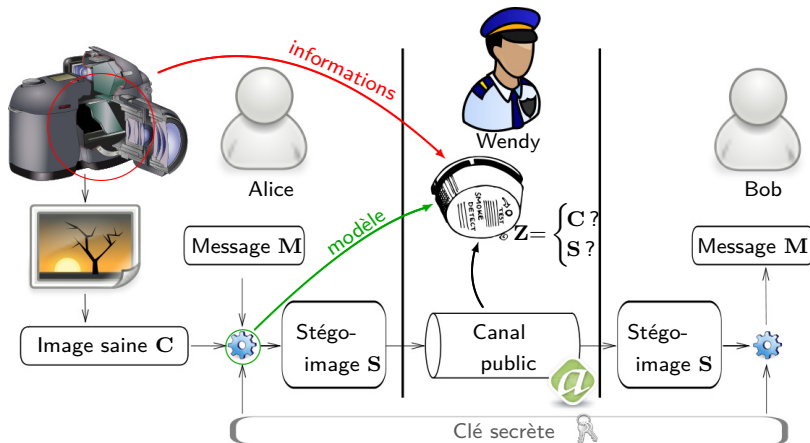


Image stéganographiée



- **Tatouage** et **fingerprinting** ont pour but l'insertion d'un identifiant.
- La **stéganographie** a pour objectif la dissimulation d'un message.
- C'est donc un compromis entre **capacité** et **discrétion**.

- Stéganographie : Insertion d'informations dans un média numérique.
- Stéganalyse : Détection de la présence d'informations cachées dans un média.





- L'insertion d'un message repose sur la limite des sens humains ...
- et doit se conformer à la structure du média de couverture.



image originale

$$R=254 = 11111110_2$$

$$G=183 = 10110111_2$$

$$B=156 = 10011100_2$$



modification des LSB

$$R=255 = 11111111_2$$

$$G=182 = 10110110_2$$

$$B=157 = 10011101_2$$



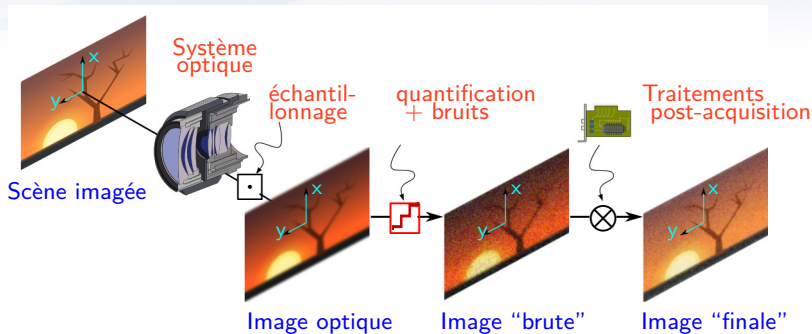
modification du MSB vert

$$R=254 = 11111110_2$$

$$G=55 = 00110111_2$$

$$B=156 = 10011100_2$$

Exemple de dissimulation d'informations par substitution des bits de poids faibles (LSB).



- La structure d'une image naturelle résulte de phénomènes connus.
- Ces éléments peuvent être utilisés par le stéganalyste.

- 1 Quelques détecteurs exploitent la **théorie de la décision statistique** (test du rapport de vraisemblance de Dabeer,  $\chi^2$ ) :  
↪ *reposent sur un modèle d'image naïf ; performances médiocres.*
- 2 De nombreux détecteurs (RS, SPA/LSM, etc. . . ) appartiennent à la classe des **détecteur structuraux** :  
↪ *exploitent de façon ad-hoc la structure entre pixels voisins ; difficilement caractérisables statistiquement.*
- 3 Les détecteurs du **WS (Weighted Stego)** sont une classe à part :  
↪ *modèle auto-regressif local de l'image peu précis ; propriétés statistiques inconnues.*
- 4 Les **détecteurs universels** reposent sur l'apprentissage supervisé :  
↪ *permet de s'affranchir d'un modèle d'image ; caractéristiques difficile à choisir ; probabilités d'erreurs peu maîtrisables.*

## Limites de ces approches

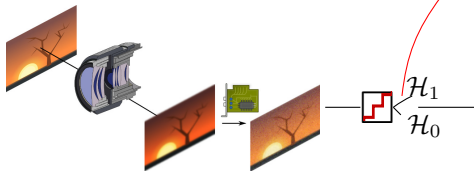
- Pas d'étude statistique réaliste sur la détection d'informations cachées.
- L'expression des probabilités d'erreur est obtenue empiriquement.
- Borne sur la puissance d'un test inconnue.
- L'utilisation d'un modèle d'image précis n'a pas été proposée.

## Principaux enjeux des travaux

- Les performances du test doivent être analytiquement connue.
- La prise en compte de la quantification des données est cruciale.
- Une image naturelle est un objet complexe à représenter.
- Nécessité d'un modèle adapté aux images simple et précis.

## Espace expérimental

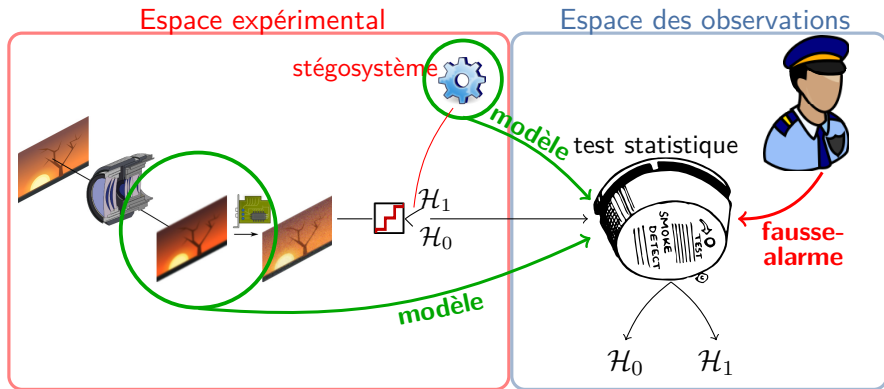
stégosystème 



## Espace des observations



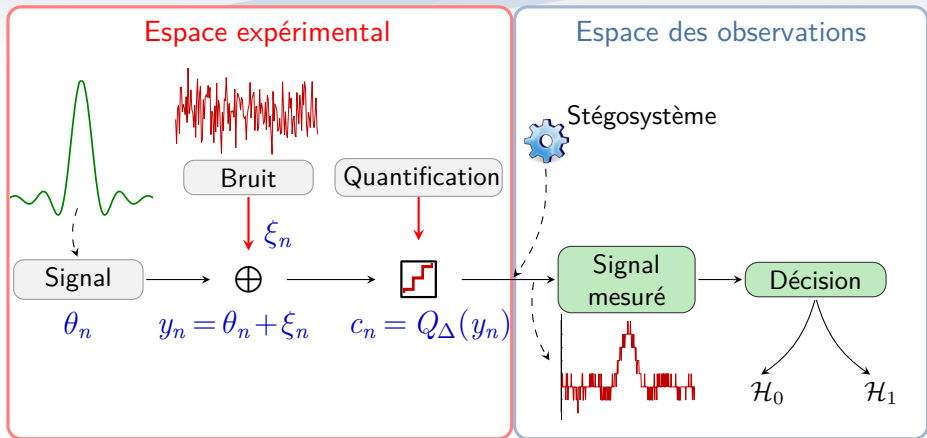
- Garantir une probabilité de fausse-alarme maximale.
- Pouvoir décrire statistiquement les performances du test proposé.
- Prendre une décision indépendante du contenu des images.



- 1 Détection statistique optimale (paramètres de l'image connus)
  - Taux d'insertion  $R$  connu, hypothèses simples
  - Taux d'insertion  $R$  inconnu, hypothèses composites
- 2 Modélisation d'une image naturelle
  - Processus d'acquisition d'une image naturelle
  - Modèle de la scène imagée
- 3 Test statistique exploitant le modèle d'image (image inconnue)
  - Construction du test adapté aux images naturelles
  - Comparaison de puissance avec la borne d'optimalité
- 4 Résultats numériques
- 5 Synthèse, conclusions et perspectives

- 1 Détection statistique optimale (paramètres de l'image connus)
  - Taux d'insertion  $R$  connu, hypothèses simples
  - Taux d'insertion  $R$  inconnu, hypothèses composites
- 2 Modélisation d'une image naturelle
  - Processus d'acquisition d'une image naturelle
  - Modèle de la scène imagée
- 3 Test statistique exploitant le modèle d'image (image inconnue)
  - Construction du test adapté aux images naturelles
  - Comparaison de puissance avec la borne d'optimalité
- 4 Résultats numériques
- 5 Synthèse, conclusions et perspectives



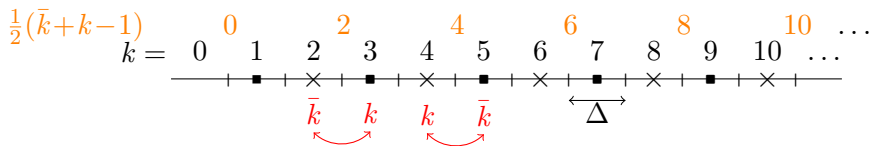


- La valeur  $c_n$  est la réalisation d'une variable aléatoire discrète dont la loi de distribution est notée  $P_{\theta_n} = \{P_{\theta_n}[k]\}_{k \in \mathcal{I}}$ .

- Deux suppositions permettent de modéliser la stéganographie :
  - 1 Le message est compressé et/ou crypté :  $\forall l, m_l \sim \mathcal{B}(1, 1/2)$ .
  - 2 La probabilité de d'insertion est identique pour chacun des pixels.
- Après insertion d'information avec un taux  $R = \ell/N$  la loi de probabilité de  $s_n$  est donnée  $\forall k \in \mathcal{Z}$  par :

$$Q_{\theta_n}^R[k] = \mathbb{P}[s_n = k] = \left(1 - \frac{R}{2}\right) P_{\theta_n}[k] + \frac{R}{2} P_{\theta_n}[\bar{k}]$$

où l'entier  $\bar{k} = k + (-1)^k$  correspond à  $k$  avec LSB inversé :



- Supposons que  $\forall n = \{1, \dots, N\}$  les distributions  $P_{\theta_n}$  et  $Q_{\theta_n}^R$  sont connues lors de l'analyse du média  $\mathbf{Z}$ .
- Pour décider entre  $\mathcal{H}_0 = \{z_n \sim P_{\theta_n}\}$  et  $\mathcal{H}_1 = \{z_n \sim Q_{\theta_n}^R\}$ , le test le plus puissant (PP) vérifiant  $\mathbb{P}_{\mathcal{H}_0}[\delta(\mathbf{Z}) = \mathcal{H}_1] \leq \alpha_0$  est le test  $\delta_R$  du rapport de vraisemblance (RV) :

$$\delta_R : \Lambda_R(\mathbf{Z}) = \prod_{n=1}^N \Lambda_R(z_n) \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\gtrless}} \tau_{\alpha_0}$$

- où  $\tau_{\alpha_0}$  est la solution de l'équation  $\mathbb{P}_{\mathcal{H}_0}[\Lambda_R(\mathbf{Z}) = \tau_{\alpha_0}] \leq \alpha_0$ .

$$\Lambda_R(z_n) = \frac{Q_{\theta_n}^R[z_n]}{P_{\theta_n}[z_n]} = (1 - R) + R \underbrace{\frac{P_{\theta_n}[\frac{1}{2}(\bar{k} + k - 1) + m_l]}{P_{\theta_n}[z_n]}}_{\Lambda_1(z_n)}.$$

- En supposant que les pixels sont indépendants  $y_n \sim \mathcal{N}(\theta_n, \sigma_n^2)$ , la loi de probabilité  $P_{\theta_n} = \{P_{\theta_n}[k]\}_{k \in \mathcal{Z}}$  est donnée par  $\forall k \in \mathcal{Z}$  :

$$P_{\theta_n}[k] = \frac{1}{\sigma_n} \int_{\Delta(k-\frac{1}{2})}^{\Delta(k+\frac{1}{2})} \phi\left(\frac{x - \theta_n}{\sigma_n}\right) dx = \frac{\Delta}{\sigma_n} \phi\left(\frac{\Delta k - \theta_n}{\sigma_n}\right) + o\left(\frac{\Delta^2}{\sigma_n^2}\right).$$

- La quantité  $\log \Lambda_1(z_n)$  est donnée par :

$$\log \Lambda_1(z_n) = \sum_{i=1}^N \frac{\Delta}{2\sigma_n^2} (\Delta z_n - \theta_n)(z_n - \bar{z}_n) + C + o_p\left(\frac{\Delta^2}{\sigma_n^2}\right).$$

- En supposant que les pixels sont indépendants  $y_n \sim \mathcal{N}(\theta_n, \sigma_n^2)$ , la loi de probabilité  $P_{\theta_n} = \{P_{\theta_n}[k]\}_{k \in \mathcal{Z}}$  est donnée par  $\forall k \in \mathcal{Z}$  :

$$P_{\theta_n}[k] = \frac{1}{\sigma_n} \int_{\Delta(k-\frac{1}{2})}^{\Delta(k+\frac{1}{2})} \phi\left(\frac{x - \theta_n}{\sigma_n}\right) dx = \frac{\Delta}{\sigma_n} \phi\left(\frac{\Delta k - \theta_n}{\sigma_n}\right) + o\left(\frac{\Delta^2}{\sigma_n^2}\right).$$

- La quantité  $\log \Lambda_1(z_n)$  est donnée par :

$$\log \Lambda_1(z_n) = \sum_{i=1}^N \frac{\Delta}{2\sigma_n^2} (\Delta z_n - \theta_n)(z_n - \bar{z}_n) + C + o_p\left(\frac{\Delta^2}{\sigma_n^2}\right).$$

- En comparaison, les détecteurs du WS sont basés sur la statistique :

$$\Lambda_{ws} = \sum_{n=1}^N \underbrace{w_n}_{\text{poids}} \underbrace{(z_n - \hat{c}_n)}_{\text{résidus}} \underbrace{(z_n - \bar{z}_n)}_{2 \cdot \text{lsb}(z_{k,l}) - 1 = \pm 1},$$

- Le RV  $\log \Lambda_1(\mathbf{Z})$  est facile à caractériser asymptotiquement, lorsque  $N \rightarrow \infty$  en vertu du théorème de la limite centrale :

$$\frac{1}{\sqrt{N}} \log \Lambda_1(\mathbf{Z}) = \frac{1}{\sqrt{N}} \sum_{n=1}^N \log \Lambda_1(z_n) \rightsquigarrow \begin{cases} \mathcal{N}(\mu_0 \sqrt{N}; \sigma_0^2) & \text{sous } \mathcal{H}_0 \\ \mathcal{N}(\mu_1 \sqrt{N}; \sigma_1^2) & \text{sous } \mathcal{H}_1 \end{cases}$$

- Un calcul direct montre l'impact de la quantification sur ces moments, au travers des quantités  $\zeta_i(\theta_n, \Delta)$  :

$$\mu_0 = \frac{-\Delta^2}{8\sigma_n^2} + \frac{\zeta_1(\theta_n, \Delta)}{\sigma^2}, \quad \sigma_0^2 = \frac{\Delta^4}{4\sigma_n^4} + \frac{\zeta_2(\theta_n, \Delta)}{\sigma_n^4}$$

$$\mu_1 = \frac{\Delta^2}{8\sigma_n^2}, \quad \sigma_1^2 = \frac{\Delta^4}{4\sigma_n^4} + \frac{\zeta_3(\theta_n, \Delta)}{\sigma_n^4}$$

- Pour tout  $R \in ]0; 1]$ , la probabilité de fausses-alarmes du test  $\delta_R$  est majorée par  $\alpha_0$  en fixant le seuil :

$$\tau_{\alpha_0} = \left( \sigma_0 \sqrt{N} \right) \Phi^{-1}(1 - \alpha_0) + (\mu_0 N)$$

## Théorème 1 :

La puissance du test PP  $\beta_{\delta_R}$  basé sur le RV approché  $\log \Lambda_1(z_n)$  est :

$$\beta_{\delta_R} \simeq 1 - \Phi \left( \frac{\sigma_0 \Phi^{-1}(1 - \alpha_0) - (\mu_1 - \mu_0) R \sqrt{N}}{\sigma_R} \right)$$

- où la variance du  $\sigma_R^2$  du log RV  $\log \Lambda_1(z_n)$  est :

$$\sigma_R^2 = (1 - R)(\sigma_0^2 + \mu_0^2) + R(\sigma_1^2 + \mu_1^2) - ((1 - R)\mu_1 + R\mu_0)^2$$





## Problèmes

- Le seuil de décision (et les performances du test) dépendent des paramètres de l'images :  $R$ ,  $\theta_n$  et  $\sigma_n$ .
- Le taux d'insertion  $R$  est inconnu.
- Les hypothèses n'admettent pas un rapport de vraisemblance monotone ; l'existence d'un test UPP est compromise.

## Approche proposée

- Des **hypothèses supplémentaires** sont nécessaires pour résoudre **de façon simple** le problème lorsque  $R$  est inconnu.

- En préambule définissons le “rapport insertion sur bruit” :

$$\varrho_N = \frac{R\Delta}{2\bar{\sigma}} \sqrt{N} \quad \text{où} \quad \frac{1}{\bar{\sigma}^2} = \sum_{n=1}^N \frac{1}{\sigma_n^2}.$$

Considérons que  $\Delta$ ,  $\sigma$  et  $R$  peuvent évoluer de sorte que :

$$\varrho_N \rightarrow \varrho \in \mathbb{R}_*^+ \text{ lorsque } N \rightarrow +\infty.$$

- Sous ces hypothèses, un développement de  $\log \Lambda_1(\mathbf{Z})$  autour de  $\frac{\Delta}{\sigma} = 0$  montre que, lorsque  $N \rightarrow +\infty$  :

$$\log \Lambda_1(\mathbf{Z}) \xrightarrow{p} \log \bar{\Lambda}(\mathbf{Z}) = \frac{\Delta \bar{\sigma}}{\sqrt{N}} \sum_{i=1}^N \frac{(\Delta z_n - \theta_n)(z_n - \bar{z}_n)}{\sigma_n^2} \rightsquigarrow \mathcal{N}(\varrho; 1)$$

- Pour tout  $R \in ]0; 1]$ , la probabilité de fausses-alarmes du test  $\bar{\delta}$ , reposant sur  $\log \bar{\Lambda}(\mathbf{Z})$ , est majorée par  $\alpha_0$  en fixant le seuil :

$$\tau_{\alpha_0} = \Phi^{-1}(1 - \alpha_0)$$

### Théorème 2 :

La puissance  $\beta_{\bar{\delta}}$  du test  $\bar{\delta}$  basé sur le RV approché  $\log \bar{\Lambda}(\mathbf{Z})$  est :

$$\beta_{\bar{\delta}} \simeq 1 - \Phi \left( \Phi^{-1}(1 - \alpha_0) - \varrho \right)$$

### Théorème 3 :

Si les paramètres  $\theta_n$  et  $\sigma_n$  sont connus pour tout  $n \in \mathcal{I}$  et si lorsque  $N \rightarrow +\infty$  on vérifie  $\varrho_N \rightarrow \varrho \in \mathbb{R}_*^+$  alors le test  $\bar{\delta}$  est asymptotiquement uniformément le plus puissant (AUPP).



## Problèmes

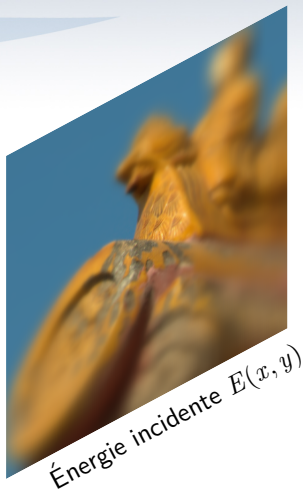
- Les paramètres  $\theta_n$  et  $\sigma_n$  ne sont pas connus.
- Difficile d'estimer  $\theta_n$  de façon simple car une image naturelle est un objet difficile à modéliser.
- Cette procédure d'estimation doit garantir le respect d'une probabilité de fausse alarme du test.

## Méthodologie proposée

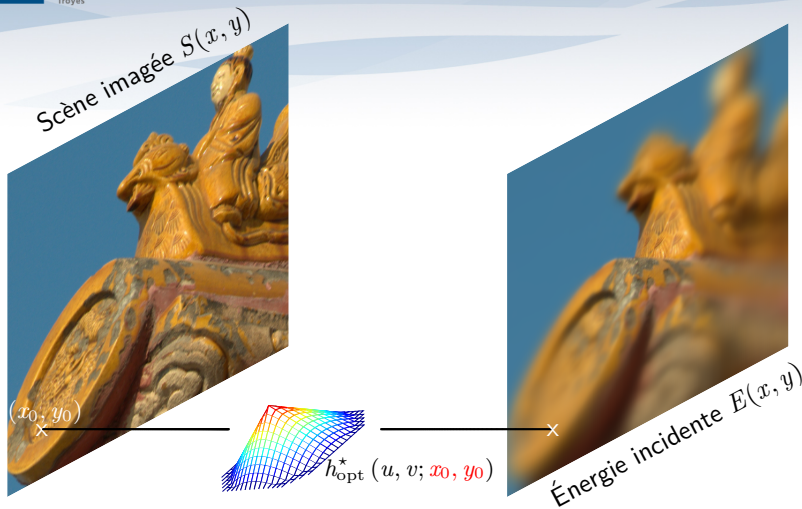
- La modélisation des images doit être précise compte tenu des faibles modifications qu'il faut détecter.
- Le respect d'un taux de fausse-alarme exige un modèle d'image dont les paramètres sont maîtrisables.
- Pour obtenir de telles informations sur l'image il est proposé de modéliser la scène imagée et le processus d'acquisition d'une image.

- 1 Détection statistique optimale (paramètres de l'image connus)
  - Taux d'insertion  $R$  connu, hypothèses simples
  - Taux d'insertion  $R$  inconnu, hypothèses composites
- 2 Modélisation d'une image naturelle
  - Processus d'acquisition d'une image naturelle
  - Modèle de la scène imagée
- 3 Test statistique exploitant le modèle d'image (image inconnue)
  - Construction du test adapté aux images naturelles
  - Comparaison de puissance avec la borne d'optimalité
- 4 Résultats numériques
- 5 Synthèse, conclusions et perspectives

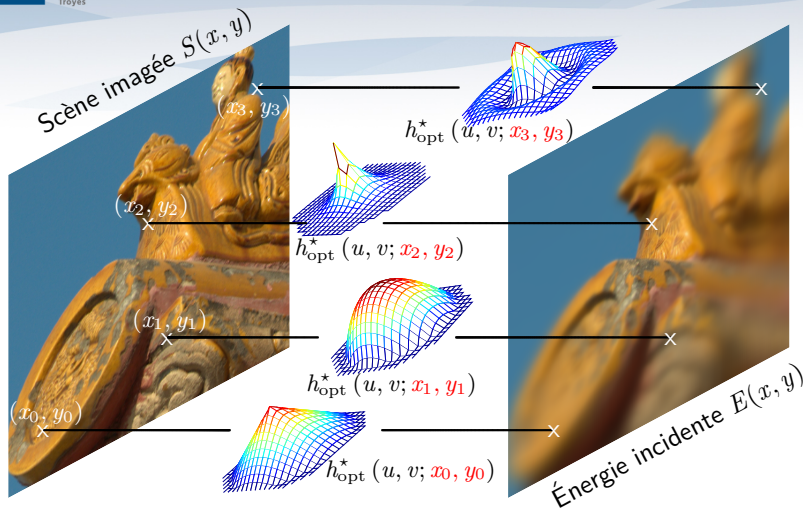




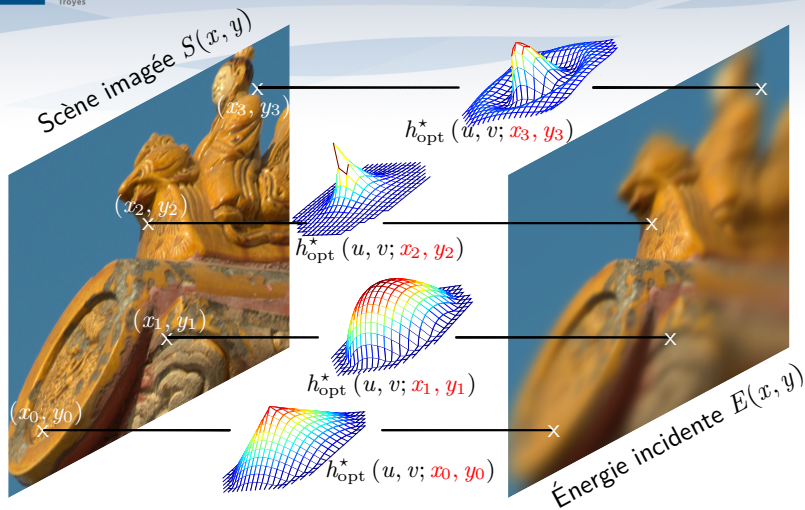




$$E(x_0, y_0) = \iint_{\mathcal{X}} S(x_0 - u, y_0 - v) h_{\text{opt}}^*(u, v; x_0, y_0) du dv$$



$$E(x, y) = \iint_{\mathcal{X}} S(x - u, y - v) h_{\text{opt}}^*(u, v; x, y) du dv$$



$$E(x, y) = \iint_{\mathcal{X}} S(x - u, y - v) h_{\text{opt}}^*(u, v; x, y) du dv$$

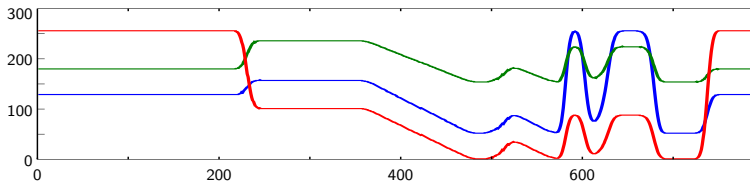
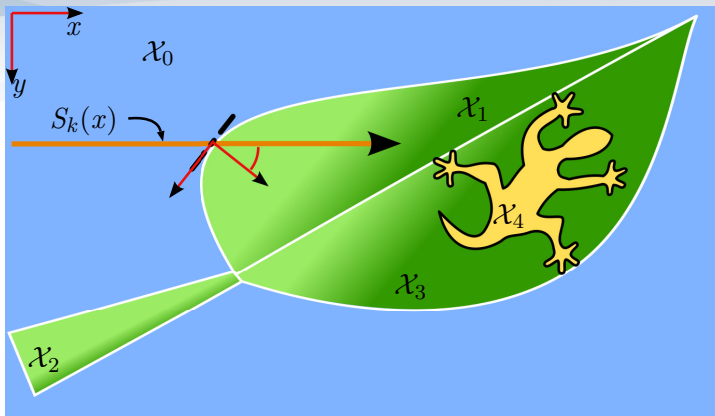
$$h \text{ localement stationnaire} \Rightarrow E(x, y) = S ** h_{\text{opt}}^* \parallel_{(x, y)}$$

- Afin de simplifier l'application de la méthodologie proposée, il est admis que :
  - ①  $h_{\text{opt}}^*$  est non-stationnaire par changement d'échelle ;
  - ②  $h_{\text{opt}}^*$  est correctement modélisée par un filtre Gaussien.

$$h_{\text{opt}}(u, v; \varsigma_0(x, y)) \approx \frac{1}{\varsigma_0^2} \phi\left(\frac{\rho}{\varsigma_0}\right)$$

$$\text{avec } \rho = \sqrt{u^2 + v^2} \quad \text{et} \quad \phi(x) = \frac{1}{2\pi} \exp\left(-\frac{x^2}{2}\right).$$

- L'énergie incidente sur le photo-détecteur  $\theta$  est donnée par :
$$\theta(x, y) \propto E(x, y) = S(x, y) ** h_{\text{opt}}(x, y; \varsigma_0(x, y)).$$



- La puissance émise par la  $k$ -ième ligne,  $y = y_k$ , peut s'écrire :

$$S(x, y_k) = S_k(x) = S_{k,c}(x) + S_{k,s}(x).$$

- avec  $S_{k,c} \in C^0$  (partie continue) et  $S_{k,s}, C^0$  (discontinuités) :

$$S_{k,s}(x) = \sum_{i=1}^{r_k} u_i \delta(x - \nu_{k,i}) \quad \text{avec} \quad \delta(x - \nu_{k,i}) = \begin{cases} 0 & \text{si } \nu_{k,i} < x \\ 1 & \text{si } \nu_{k,i} \geq x \end{cases}$$

et  $\nu_{k,i}$  la localisation de la  $i$ -ième discontinuité du  $k$ -ième segment.

- L'énergie incidente sur le photo-détecteur est :

$$\theta_k(x) = \theta_{k,c}(x) + \theta_{k,s}(x),$$

- où  $\theta_{k,c}$  correspond à la partie continue  $S_{k,c}$  et admet la décomposition :

$$\theta_{k,c} = \sum_{j=1}^p c_{k,j} f_j.$$

- La fonction  $\theta_{k,s}$  correspondant aux discontinuités  $S_{k,s}$  est donnée par :

$$\theta_{k,s}(x) \propto \sum_{i=1}^{r_k} u_i H_{\text{opt}} \left( \frac{x - \nu_{k,i}}{S_{k,i}} \right) = \sum_{i=1}^{r_k} u_i \Phi \left( \frac{x - \nu_{k,i}}{S_{k,i}} \right) .$$

- Notons  $\boldsymbol{\eta}_k = (\nu_{k,1}, S_{k,1}, \dots, \nu_{k,d}, S_{k,d})$  le paramètre de discontinuité.
- Après intégration sur la surface des pixels, le signal  $\mathbf{z}_k$  issu de  $\theta_k$  vérifie  $\forall k$  :

$$\mathbb{E}[\mathbf{z}_k] = \boldsymbol{\theta}_k = \mathbf{F} \mathbf{c}_k + \mathbf{H}(\boldsymbol{\eta}_k) \mathbf{u}_k$$

- où  $\mathbf{F}$  représente la partie linéaire continue.
  - et  $\mathbf{H}(\boldsymbol{\eta}_k)$  représente les discontinuités, partie non-linéaire.
- L'exploitation d'un modèle avec des paramètre de nuisances non-linéaire est complexe.

- 1 Détection statistique optimale (paramètres de l'image connus)
  - Taux d'insertion  $R$  connu, hypothèses simples
  - Taux d'insertion  $R$  inconnu, hypothèses composites
- 2 Modélisation d'une image naturelle
  - Processus d'acquisition d'une image naturelle
  - Modèle de la scène imagée
- 3 Test statistique exploitant le modèle d'image (image inconnue)
  - Construction du test adapté aux images naturelles
  - Comparaison de puissance avec la borne d'optimalité
- 4 Résultats numériques
- 5 Synthèse, conclusions et perspectives



- En pratique l'image est "découpée" en  $K$  segments de  $L$  pixels.
- Des méthodes existent afin d'estimer le paramètre  $\boldsymbol{\eta}_k$ .
- Notons  $\hat{\boldsymbol{\eta}}_k$  une estimation de  $\boldsymbol{\eta}_k$  telle que  $\|\hat{\boldsymbol{\eta}}_k - \boldsymbol{\eta}_k\|_1 \leq \varepsilon$ .
- Autour de  $\hat{\boldsymbol{\eta}}_k$  le modèle local peut être "linearisé" par :

$$\begin{aligned}\boldsymbol{\theta}_k &= \mathbf{F} \mathbf{s}_k + \mathbf{H}(\hat{\boldsymbol{\eta}}_k) \mathbf{u}_k + \dot{\mathbf{H}}(\hat{\boldsymbol{\eta}}_k) \mathbf{u}_k (\hat{\boldsymbol{\eta}}_k - \boldsymbol{\eta}_k) + o(\varepsilon^2) \\ &= \mathbf{G}(\hat{\boldsymbol{\eta}}_k) \mathbf{v}_k + o(\varepsilon^2)\end{aligned}$$

avec  $\mathbf{G}(\hat{\boldsymbol{\eta}}_k) = (\mathbf{F} | \mathbf{H}(\hat{\boldsymbol{\eta}}_k) | \dot{\mathbf{H}}(\hat{\boldsymbol{\eta}}_k))$  et  $\mathbf{v}_k = (\mathbf{s}_k, \mathbf{u}_k, \mathbf{u}_k (\hat{\boldsymbol{\eta}}_k - \boldsymbol{\eta}_k))^T$

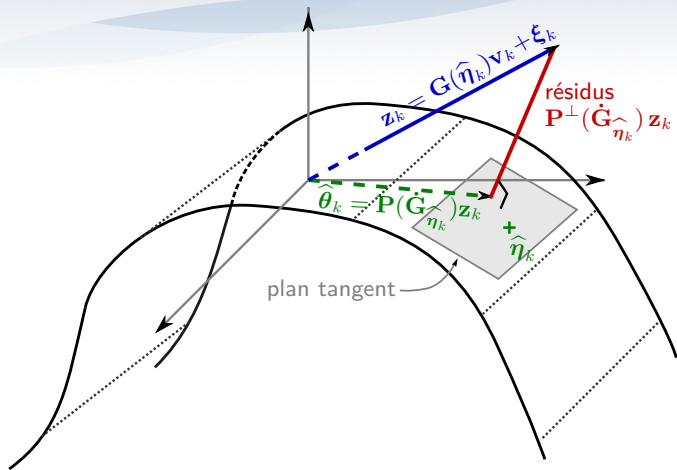


Illustration d'une estimation dans un contexte non-linéaire.

- Notons  $\Sigma_k$  la covariance du  $k$ -ième segment, l'estimation par maximum de vraisemblance de  $\theta_k$  est :

$$\hat{\theta}_k = \mathbf{P}_{\dot{\mathbf{G}}(\hat{\eta}_k)} \mathbf{z}_k = \mathbf{G}(\hat{\eta}_k) \left( \mathbf{G}(\hat{\eta}_k)^T \Sigma_k^{-1} \mathbf{G}(\hat{\eta}_k) \right)^{-1} \mathbf{G}(\hat{\eta}_k)^T \Sigma_k^{-1} \mathbf{z}_k \quad (1)$$

- En utilisant l'estimation  $\hat{\theta}_k$  dans le RV  $\log \bar{\Lambda}(\mathbf{Z})$  on a :

$$\log \hat{\Lambda}(\mathbf{Z}) = \frac{\bar{\sigma}}{\sqrt{\kappa}} \sum_{k=1}^K \sum_{l=1}^L \frac{(z_{k,n} - \hat{\theta}_{k,n})(z_{k,n} - \bar{z}_{k,n})}{\sigma_{k,l}^2} \rightsquigarrow \mathcal{N} \left( \varrho \sqrt{\frac{\kappa}{N}}; 1+b \right) \quad (2)$$

avec  $\kappa = K(L-p) - 3r_K$  et  $0 < b < b_{max} = \frac{o(\varepsilon^2)}{\bar{\sigma}^2(L-p-3)}$ .

- Quel est l'impact de (1) sur la performance du test reposant sur (2) ?
- La quantité  $b$  est bornée, la puissance l'est alors également.

- Pour tout  $\alpha_0 \in ]0; 0,5]$ , la probabilité de fausses-alarmes du test  $\hat{\delta}$  est majorée par  $\alpha_0$  en fixant le seuil :

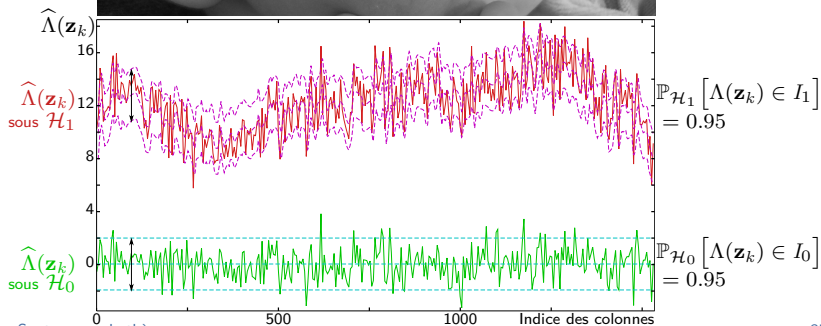
$$\hat{\tau}_{\alpha_0} = \Phi^{-1}(1 - \alpha_0)\sqrt{1 + b_{\max}} \geq \Phi^{-1}(1 - \alpha_0)\sqrt{1 + \bar{b}},$$

#### Théorème 4 :

La puissance  $\beta_{\hat{\delta}}$  du test  $\hat{\delta}$  basé sur  $\log \hat{\Lambda}(\mathbf{Z})$  est :

$$1 - \Phi \left( \frac{\hat{\tau}_{\alpha_0} - \varrho \sqrt{\frac{\kappa}{N}}}{1 + b_{\max}} \right) \leq \beta_{\hat{\delta}} \leq 1 - \Phi \left( \hat{\tau}_{\alpha_0} - \varrho \sqrt{\frac{\kappa}{N}} \right)$$

- La perte de puissance (en comparaison à  $\bar{\delta}$ ) est due :
  - 1 au terme  $\sqrt{\frac{\kappa}{N}}$  représentant la réduction de “paramètres libres”.
  - 2 au biais  $b$  dû à la linéarisation du modèle autour de  $\hat{\eta}_k \neq \eta_k$ .



- 1 Détection statistique optimale (paramètres de l'image connus)
  - Taux d'insertion  $R$  connu, hypothèses simples
  - Taux d'insertion  $R$  inconnu, hypothèses composites
- 2 Modélisation d'une image naturelle
  - Processus d'acquisition d'une image naturelle
  - Modèle de la scène imagée
- 3 Test statistique exploitant le modèle d'image (image inconnue)
  - Construction du test adapté aux images naturelles
  - Comparaison de puissance avec la borne d'optimalité
- 4 Résultats numériques
- 5 Synthèse, conclusions et perspectives

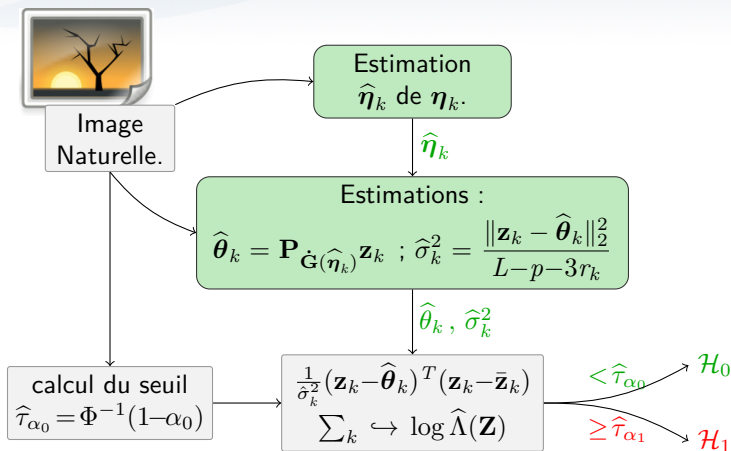
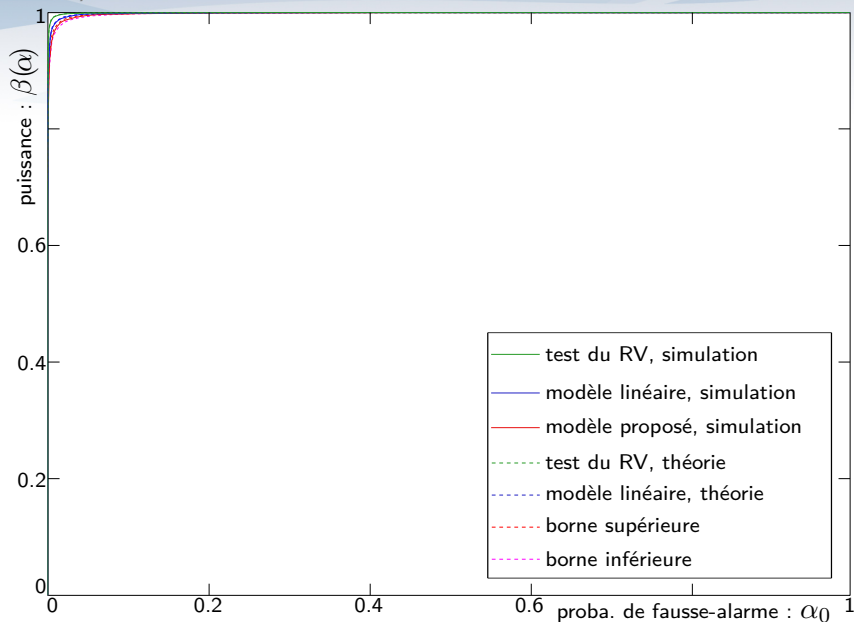
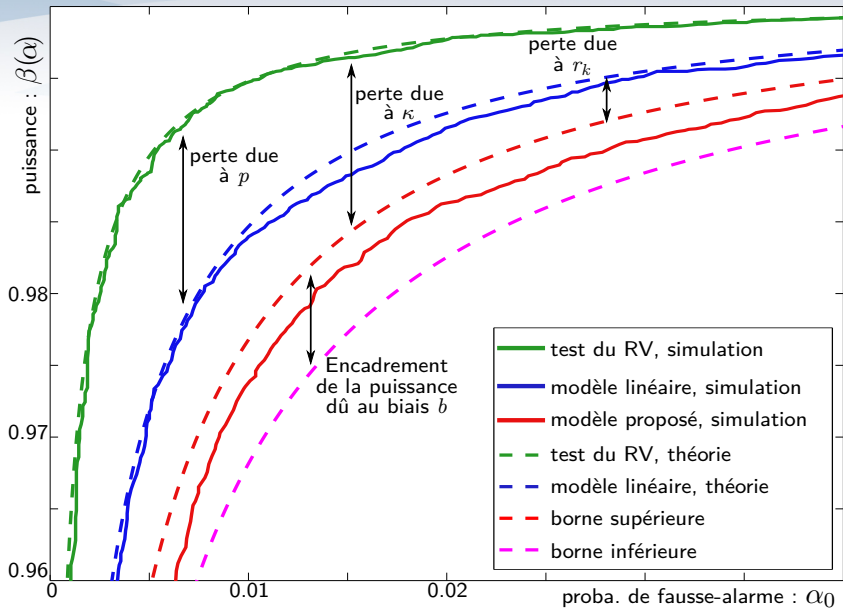
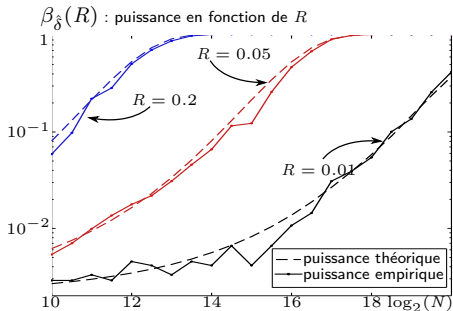
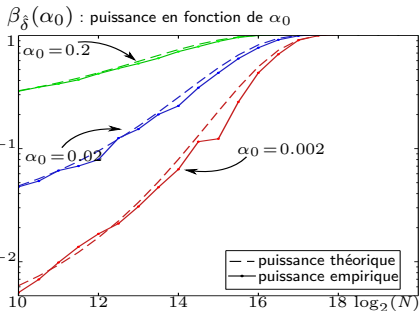
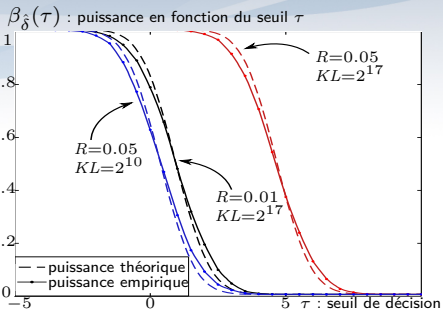


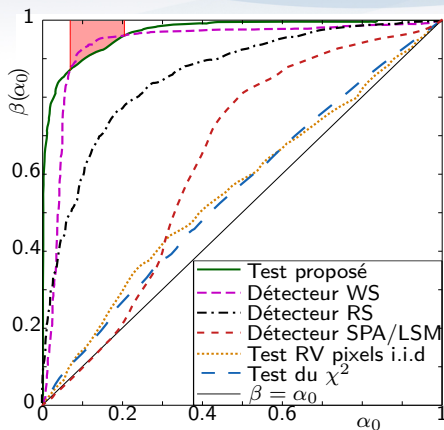
Schéma d'implémentation de la méthodologie proposée.



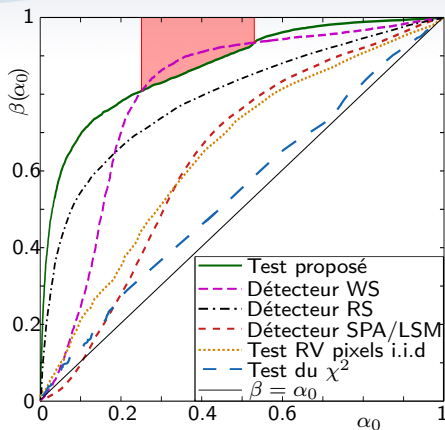




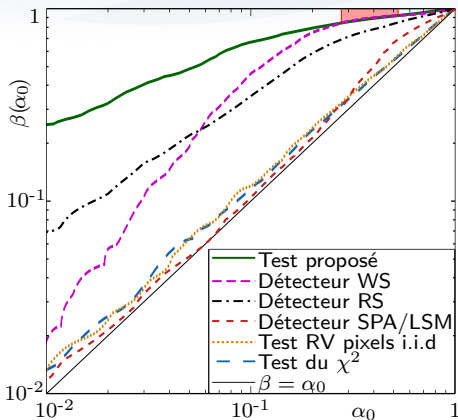




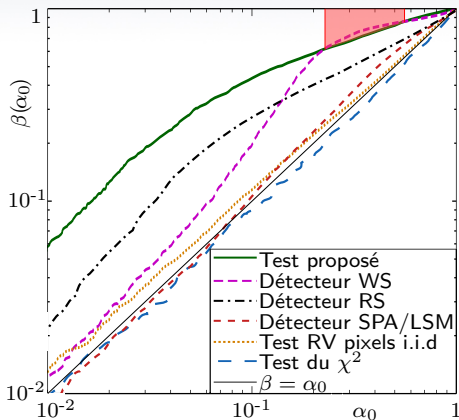
Base UCID,  $R = 0.1$



Base BOSS,  $R = 0.1$



Base UCID,  $R = 0.05$



Base BOSS,  $R = 0.05$

- 1 Détection statistique optimale (paramètres de l'image connus)
  - Taux d'insertion  $R$  connu, hypothèses simples
  - Taux d'insertion  $R$  inconnu, hypothèses composites
- 2 Modélisation d'une image naturelle
  - Processus d'acquisition d'une image naturelle
  - Modèle de la scène imagée
- 3 Test statistique exploitant le modèle d'image (image inconnue)
  - Construction du test adapté aux images naturelles
  - Comparaison de puissance avec la borne d'optimalité
- 4 Résultats numériques
- 5 Synthèse, conclusions et perspectives

- La théorie de la décision statistique permet la conception de tests optimaux selon le cas considéré et la connaissance supposée.
- S'appuyer sur la modélisation du système d'acquisition permet d'appliquer ces tests en pratique.
- La méthodologie de linéarisation permet d'expliciter les performances du test (et de réduire le temps de calcul).
- Les résultats numériques présentés attestent de la pertinence de la méthodologie.

Ces travaux ouvrent de vastes perspectives dans différents domaines :

- Pour la théorie de la décision et de l'estimation statistique : impact de la quantification sur la performance d'un test et d'un estimateur.
- Dans la modélisation des images : utilisation du modèle à d'autres fins ; applications à la détection d'"anomalies" différentes.
- À des fins de criminalistique : certification et authentification des images, détection d'information cachées par "correspondance de LSB" (*LSB matching* ,  $\pm 1$ ).

