

SYSTÈME POUR DETERMINER L'IDENTIFICATION D'UN APPAREIL PHOTOGRAPHIQUE À PARTIR D'UNE PHOTOGRAPHIE ET PROCÉDÉ MIS EN OEUVRE DANS UN TEL SYSTÈME

5 DOMAINE TECHNIQUE DE L'INVENTION

L'invention se rapporte à l'identification d'un appareil photographique plus particulièrement, l'invention concerne un système pour déterminer l'identification d'un appareil photographique à partir d'une photographie numérique. L'invention concerne aussi un procédé mis en œuvre dans un tel système. Ces systèmes trouvent des applications importantes pour déterminer la provenance d'une photographie.

[001] La criminalistique numérique ou la recherche de preuves dans un média numérique a connu un développement important au cours de la dernière décennie. Dans ce domaine les méthodes proposées se distinguent en deux catégories selon que l'on souhaite identifier le modèle d'appareil photographique ou l'appareil lui-même (une instance d'un certain modèle).

[002] De manière générale les méthodes d'identification sont passives ou actives. Dans le cas des méthodes actives, les données numériques représentant le contenu de l'image sont modifiées afin d'insérer un identifiant (méthode dites de tatouage ou de watermarking). Lorsque l'image inspectée ne contient pas de tatouage, l'identification de l'appareil d'acquisition doit se faire à partir des données de l'image.

ÉTAT DE LA TECHNIQUE ANTÉRIEURE

[003] De nombreuses méthodes d'identification passives d'un appareil photographique ont été proposées. La plupart de ces méthodes sont basées sur

la présence d'un bruit de non-uniformité de la photosensibilité (NUPS). Ces bruits de NUPS sont utilisés pour identifier de l'appareil d'acquisition d'une image au travers d'une analyse comparative par corrélation dans [Lukas2006, Sutcu2007, LI2010] ou encore avec une approche de reconnaissance de formes dans 5[Chen2008, Filler2008].

[004] De même, les méthodes d'apprentissage supervisées ont également été appliquées à la problématique dans [Kharrazi2004, Choi2006]. Cependant, ces méthodes possèdent des inconvénients majeurs tels que: 1) leur performances ne sont pas établies mais uniquement mesurées empiriquement en utilisant de 10grandes base de données images et, 2) le bruit (NUPS) résulte de la présence de pixels dits « chauds » ou « morts » dans la matrice du capteur et la présence de ces derniers peut être facilement corrigée, ou au moins compensée par la calibration de l'appareil (photographie d'une mire par exemple).

[005] La méthode proposée dans l'invention a pour but d'identifier de façon 15passive un modèle d'appareil photographique. Cette méthode est basée sur l'hétéroscédasticité du bruit présent dans une image. On sait que le bruit qui perturbe les pixels d'une photographie numérique dépend linéairement de l'intensité lumineuse incidente sur le pixel comme cela a été modélisé dans [Healey 1994]. En outre une méthode été proposé dans [Foi 08] pour estimer les 20paramètres caractérisant cette dépendance entre la valeur moyenne des pixels et la variance du bruit d'acquisition. Les paramètres d'hétéroscédasticité du bruit, dépendent du modèle d'appareil photographique utilisé et peuvent donc servir à l'identification de ce dernier. En utilisant ce modèle paramétrique de l'hétéroscédasticité du bruit des images, la méthode d'identification proposée 25consiste en un rapport de vraisemblance afin de tester statistiquement si une image donnée provient d'un appareil photographique A ou d'un appareil

photographique B. Les propriétés statistiques du rapport de vraisemblance ont été théoriquement établies ce qui permet de fixer à priori le seuil de décision permettant de respecter une probabilité d'erreur prescrite (décider A quant B est vraie ou inversement).

5[006] Il faut noter que cette propriété d'hétéroscédasticité du bruit des images a déjà été utilisée de façon ad-hoc, sans exploitation statistique du modèle paramétrique. A titre d'exemple dans [Maillard 2008], cette propriété est utilisée afin d'identifier dans une vidéo (succession d'images) les objets qui auraient éventuellement été ajoutés après l'acquisition.

10[007] L'article « Detection of digital processing of Images through a realistic model of CCD noise », [Maillard2008], décrit un processus pour détecter si une séquence vidéo a été prise par une ou au moins deux caméras.

[008] Cependant ce qui est décrit dans cet article ne correspond pas exactement à la détermination de l'identification d'un appareil photographique. Le procédé qui y
15est appliqué est par essence distinct de ce que propose l'invention.

[009] L'article « Digital caméra identification from sensor pattern noise », [Lukas 2006], propose une méthode pour identifier des appareils numériques à partir d'images fondées sur des bruits de modèle de capteur. Pour chaque appareil (caméra), on détermine d'abord le bruit de motif de référence, qui sert comme une
20identification des empreintes digitales uniques. Ce résultat est obtenu en faisant la moyenne du bruit obtenu à partir de plusieurs images en utilisant un filtre de débruitage. Pour identifier la caméra à partir d'une image donnée, on considère le bruit de modèle de référence en tant que filigrane à étalement de spectre ; la présence de l'image est établie en utilisant un détecteur de corrélation.

[0010] L'article « Determining Image Origin and Integrity Using Sensor Noise », [Chen2008], décrit un cadre pour l'identification d'un appareil photo numérique à partir d'images numériques et pour révéler des images retouchées à l'aide du bruit non-uniformité photo-réponse (NUPS). Le principe de cette méthode est d'utiliser ce bruit de NUPS comme une empreinte stochastique unique pour chaque capteur d'image. Le NUPS est obtenu en utilisant un estimateur du maximum de vraisemblance dérivé d'un modèle simplifié de la sortie du capteur d'image.

[0011] L'invention a pour but de résoudre le problème de l'identification passive d'un modèle d'appareil photographique d'acquisition à partir d'une image donnée. Par identification passive, on entend une prise de décision dans le cas où l'image n'est pas supposée contenir d'information identifiant la source. Plus précisément, l'invention envisage de résoudre les problèmes suivants: 1) assuré qu'une photographie n'a pas été prise par un appareil donné lorsque cette dernière est compromettant ou, 2) à l'inverse garantir qu'une photographie inspectée a bien été prise par un modèle d'appareil photographique et non par un autre. Ces procédés sont caractérisés par le fait que les probabilités d'erreurs du systèmes proposées sont garantie et peuvent être prédéfinie par un utilisateur, ce qui permet de certifier les résultats. Les exemples possibles d'application pratiques sont nombreux.

20 EXPOSÉ DE L'INVENTION

[0012] Le but de l'invention est de fournir un système pour déterminer l'identification d'un premier appareil photographique à partir d'une photographie, comportant un dispositif d'analyse de photographies pour traiter ladite photographie et un dispositif de traitement pour établir des statistiques à partir de ladite analyse, le système opérant ladite identification sur la base de paramètres

« a » et « b » définissant un appareil photographique et répondant à la relation ci-dessous :

$$z_{m,n} : N(y_{m,n}, \alpha y_{m,n} + b)$$

où $z_{m,n}$ est représentatif de la valeur d'un pixel de coordonnées (m,n) d'une photographie, $N(\cdot)$ représente une distribution gaussienne et $y_{m,n}$ est l'espérance mathématique du pixel situé à la position (m,n). Le système est caractérisé en ce qu'il comprend en outre un organe de sortie qui opère par comparaison statistique avec un deuxième appareil photographique pour déterminer si la photographie a été prise par ledit premier appareil ou ledit deuxième appareil.

[0013] L'invention permet plus précisément, l'identification d'un appareil photographique en certifiant l'exactitude de l'identification, avec une précision préalablement définie.

15[0014] L'invention concerne encore un procédé, mis en œuvre dans un tel système, caractérisé en ce qu'il comporte les étapes suivantes :

- lecture de la photographie en vue de déterminer les matrices descriptives (K12),
- estimation du contenu et du bruit de l'image (K14),
- 20 - estimation de la variance des pixels par niveaux d'intensité estimée du contenu (K22),
- estimation des paramètres caractérisant la relation moyenne variance du bruit (K24),
- exécution de tests statistiques en vue de donner l'identification.
- 25 (K26).

[0015] Avantageusement, à l'étape (K26) les tests statistiques sont exécutés en fonction de la probabilité d'erreur exigée, et une décision est prise.

[0016] Selon un mode de réalisation de l'invention, la photographie est en un format non compressé ou en un format compressé sans pertes.

5[0017] L'invention concerne en outre l'utilisation du procédé ci-dessus pour la détection, de façon supervisée, de la falsification d'une zone dans une image.

[0018] L'invention concerne encore l'utilisation du procédé ci-dessus dans la recherche de preuves à partir d'une image compromettante.

[0019] L'invention concerne l'application du procédé ci-dessus dans des logiciels
10spécialisés, dans la recherche de preuves à partir de média numériques.

BRÈVE DESCRIPTION DES FIGURES

[0020] D'autres caractéristiques, détails et avantages de l'invention ressortiront à la lecture de la description qui suit, en référence aux figures annexées, qui illustrent:

- 15
- la figure 1 montre un système pour déterminer l'identification d'un appareil photographique conforme à l'invention ;
 - la figure 2 montre un organigramme explicitant le procédé mis en œuvre dans le système conforme à l'invention ;
 - les figures 3 et 4 présente une série des paramètres caractérisant les
- 20
- propriétés d'hétéroscédasticité du bruit de quelques modèles d'appareils ;
 - la figure 5 montre la comparaison des propriétés du bruit des trois appareils photographiques distincts ;

- la figure 6 montre les performances du système d'identification lorsque les modèles d'appareils sont connus. Ces performances sont présentées sous la forme de courbes COR ("Caractéristique de Opérationnelle du Récepteur") et offre une comparaison entre la puissance théoriquement calculée et la puissance obtenue sur des images réelles ; et

- la figure 7 montre les valeurs du rapport de vraisemblance (statistique sur laquelle est fondé le système d'identification) calculée sur les colonnes d'une image de mire (bande grise allant du noir au blanc) lorsque l'appareil photographique est inconnu.

10[0021] Pour plus de clarté, les éléments identiques ou similaires sont repérés par des signes de référence identiques sur l'ensemble des figures.

DESCRIPTION DETAILLEE D'UN MODE DE RÉALISATION

[0022] La figure 1, représente un système pour déterminer l'identification d'un appareil photographique. La référence 1 indique le système et la référence 2 15l'appareil photographique qui a pris une photographie 3.

[0023] C'est à partir de cette photographie que le système 1 va déterminer l'appareil photographique 2 qui a pris cette photographie. Ce système se compose d'un analyseur de photo qui va examiner cette photographie 3. De préférence la photographie se présente sous forme de fichier non compressé (ou compressé 20sans pertes) apte au traitement qui va suivre. Le format du type RAW (*format* de fichier pour les images numériques) est un format apte à ce traitement.

[0024] Le système 1 peut être mis en œuvre sur un ordinateur du type PC. Ce système 1 est muni d'un organe d'entrée 10 pour pouvoir accueillir les données de la photographie 3. Ces données sont traitées par un organe de traitement 12 qui

met en œuvre un traitement qui sera explicité ci-dessous. Un organe de sortie 14 fournira une indication sur l'identification de l'appareil photographique responsable de ladite photographie.

[0025] La figure 2 montre un organigramme explicitant le procédé conforme à 5 l'invention. A la première étape, montrée à la case K10, la photographie numérique 2 est vue comme une ou plusieurs matrices dont les éléments représentent la valeur de chacun des pixels. Dans le cas d'une image en niveau de gris (parfois appelée « noir et blanc ») la photographie peut être représentée par une unique matrice : $Z = z_{m,n}$ avec $1 \leq m \leq M$ et $1 \leq n \leq N$

10 Pour les images en couleurs, trois couleurs distinctes sont usuellement utilisées: le rouge, le vert et le bleu. Dans ce cas, une image est assimilable à 3 matrices distinctes :

$$Z = z_{m,n}^k \text{ avec } 1 \leq k \leq 3 \text{ une matrice par canal de couleurs.}$$

[0026] La seconde étape, case K12, consiste à séparer les différents canaux de 15 couleurs, lorsque l'image analysée est en couleurs. La suite des opérations étant réalisées de manière identique avec chacune des matrices représentant les canaux de couleurs, nous considérons que l'image est représentée par une unique matrice (l'indice k est omis).

[0027] Lors de l'acquisition d'une photographie numérique, des erreurs de mesure 20 d'origine distincte corrompent légèrement l'intensité enregistrée pour chacun des pixels. La valeur de chacun des pixels peut être considérée comme la somme d'un bruit aléatoire (représentant l'ensemble des erreurs de mesures) et d'un contenu déterministe : $z_{m,n} = y_{m,n} + \xi_{m,n}$

[0028] Le bruit présent dans les photographies numériques, présente la propriété d'être hétéroscédastique : Les propriétés stochastiques (aléatoires) de bruit ne sont pas constantes sur l'ensemble des pixels de l'image. Plus précisément, la valeur de chaque pixel dépend linéairement du nombre de photons incidents noté $Np_{m,n}$. La conversion photo-électrique peut être assimilée à un processus de comptage de Poisson. Aussi, le nombre d'électrons collectés au sein de chacun des pixels est donné par :

$$Ne_{m,n} : P(r_{m,n} Np_{m,n} + Nt_{m,n}) \quad (1)$$

avec $r_{m,n}$ qui est le facteur de conversion représentant, entre autre, la transmittance des filtres optiques et la sensibilité quantique (nombre moyen d'électrons générés par photon incident) et avec $Nt_{m,n}$ le nombre moyen d'électrons d'origine thermique collectés.

[0029] On peut considérer, en première approximation, que la photo-sensibilité $r_{m,n}$ ainsi que le bruit thermique $Nt_{m,n}$ sont constants pour l'ensemble des pixels d'un appareil photographique donné (voir [Healey1998]). Par conséquent on omettra les indices m et n de ces quantités. Les électrons collectés dans chacun des pixels sont ensuite transférés à une unité de lecture de charge. Durant ces phases de transfert et de lecture, plusieurs sources de phénomènes parasites interviennent ; ces dernières peuvent être modélisées comme une variable aléatoire Gaussienne (voir [Healey1998]). Ainsi, la valeur enregistrée d'un pixel $z_{m,n}$ est donnée par :

$$z_{m,n} : aN(Ne_{m,n}, \sigma^2) \quad (2)$$

où a est un paramètre de sensibilité linéairement dépendant du facteur ISO de l'appareil photographique et défini par l'utilisateur.

[0030] En raison du grand nombre de photons incidents sur les capteurs, il est possible d'approximer avec une grande précision le processus de comptage par une variable aléatoire Gaussienne.

En notant dans la suite $y_{m;n} = a(r Np_{m;n} + Nt)$ l'espérance mathématique (moyenne) du pixel à la position (m; n), les équations (1) et (2) permettent d'écrire la valeur $z_{m;n}$ comme suit, voir [Healey94,Foi2008]:

$$z_{m,n} \sim N(a(rNp_{m,n} + Nt), a^2(rNp_{m,n} + Nt + \sigma^2))$$

$$\sim N(y_{m,n}, ay_{m,n} + b) \quad (3)$$

avec $b = a^2\sigma^2$

[0031] Il est à noter que les paramètres a et b varient pour différents modèles d'appareils photographiques et peuvent donc être utilisés pour permettre une identification de ces derniers.

[0032] La troisième étape, case K14, consiste en une séparation du contenu, case K16, et du bruit de l'image, case K18. Il faut préciser que ce problème fait l'objet de travaux de recherche et que plusieurs solutions ont été proposées. Dans le cas des travaux présentés, une méthode publiée dans la littérature spécialisée a été implémentée (voir [Foi2008]). Brièvement, le principe de cette méthode est d'utiliser une décomposition de l'image sur une base d'ondelettes en utilisant les coefficients de basses fréquences comme estimateurs de contenu, c'est-à-dire de l'espérance mathématique de chacun des pixels notée ici : $z_{m,n}^{wapp}$ et, à l'inverse, en utilisant les coefficients de hautes fréquences, ou de détails, comme estimateurs du bruit notés ici $z_{m,n}^{wdet}$. Dans le document [Foi2008], on suggère de réaliser une étape de segmentation préalable afin de n'appliquer cette opération que dans les zones d'intensité approximativement constante de l'image. Cette opération

préalable permet d'estimer l'espérance mathématique et la variance des pixels par zone.

[0033] La troisième étape montrée à la case K20 concerne l'estimation de la variance des pixels par niveaux d'intensité estimée du contenu.

5[0034] À partir de la « sous-image » du contenu $z_{m,n}^{wapp}$, on construit L vecteurs z_l^{wapp} , $1 \leq l \leq L$ contenant des pixels dont les espérances estimées sont proches.

Chacun des vecteurs z_l^{wapp} contient v_l pixels:

$z_l^{wapp} = z_{l,i}^{wapp}, 1 \leq i \leq v_l$. Il est montré dans le document [Foi2008] que chacun des 10vecteurs z_l^{wapp} suit la loi de distribution suivante :

$$z_l^{wapp} \sim N(y_l \mathbf{1} \mid \phi|_2^2 \sigma_l^2 \mathbf{I}_{v_l})$$

où $\mathbf{1}$ est un vecteur constitué de 1 (de la dimension $v_l + 1$) \mathbf{I}_{v_l} est la matrice identité de taille $v_l * v_l$ et $|\phi|_2^2$ est la norme de la fonction d'échelle de l'ondelette utilisée.

Par ailleurs, compte tenu de la relation (3), la variance σ_l^2 est donnée par :

$$15 \sigma_l^2 = a y_l + b$$

L'estimateur du maximum de l'espérance est donné par :

$$\hat{y}_l = \frac{1}{v_l} \sum z_{l,i}$$

et est statistiquement distribué suivant la loi Gaussienne suivante :

$$\hat{y}_l : N\left(\hat{y}_l, \frac{|\phi|_2^2 \sigma_l^2}{v_l}\right) \quad (4)$$

20Enfin, en utilisant la relation (3), l'estimation de la variance des pixels du vecteur

z_l^{wapp} est donnée par :

$$\hat{\sigma}_l^2 = a y_l + b \quad (5) ;$$

lorsque les paramètres a et b des appareils photographiques sont connus.

[0035] Lorsque les paramètres a et b ne sont pas connus et obtenus par le maximum de vraisemblance, la variance est caractérisée par l'équation (4) :

$$\hat{\sigma}_l^2 = \frac{1}{|\phi|_2^2 (v_l - 1)} \sum_{i=1}^{v_l} (z_{l,i}^{wapp} - \hat{y}_l)^2 \quad (6)$$

[0036] Suivant la méthode proposée dans l'article [Foi2008], l'estimation de la variance est réalisée en tenant compte du phénomène de quantification des pixels permettant que la valeur de ces derniers soit un entier positif ou nul. Ce phénomène n'étant pas nécessairement pris en compte, les détails techniques sont laissés dans [Foi2008].

[0037] A la cinquième étape, montrée à la case K24, on élabore l'estimation des paramètres a et b caractérisant la relation affine entre la moyenne des pixels et la variance du bruit.

[0038] À partir de l'estimation de la variance des pixels de même niveau d'intensité et à partir de la relation entre espérance et variance des pixels (3), il est possible de fournir des estimations statistiques \hat{a} et \hat{b} des paramètres a et b . Cette estimation statistique est réalisée par maximum de vraisemblance en tenant compte du phénomène de quantification. Il est important de préciser que cette étape n'est pas nécessaire à l'exécution du test statistique visant à décider si une photographie numérique donné provient d'un premier appareil photographique ou d'un deuxième appareil photographique.

[0039] Enfin, le test statistique est exécuté en fonction de la probabilité d'erreur exigée et une décision est prise (voir case K26). À ce niveau, plusieurs cas peuvent être envisagés. Ils sont listés ci-dessous et accompagnés d'une description de leur mise en œuvre :

[0040] Dans le premier cas, l'utilisateur souhaite décider si la photographie numérique inspectée provient du premier appareil photographique encore appelé « appareil 1 » dans la suite de ce texte ou du deuxième appareil photographique (encore appelé « appareil 2 » dans la suite de ce texte) dont les propriétés du bruit sont connues pour chacun des appareils (les coefficients notés a_1 , b_1 , a_2 et b_2). Dans ce cas, on procède à un test du rapport de vraisemblance dont la fonction de décision est donnée par :

$$\begin{aligned} \text{Appareil n}^\circ 1 \text{ si } \Lambda_1(Z) &= \prod_{l=1}^L \Lambda_l(z_l^{wapp}) < \tau_{\alpha 0} \\ \text{Appareil n}^\circ 2 \text{ si } \Lambda_1(Z) &= \prod_{l=1}^L \Lambda_l(z_l^{wapp}) \geq \tau_{\alpha 0} \end{aligned}$$

où le seuil de décision est défini a priori selon l'équation (10) de sorte que la probabilité d'erreur exigée soit garantie asymptotiquement (lorsque le nombre de pixels est très élevé). Après quelques calculs, le logarithme du rapport de vraisemblance $\Lambda_l(z_i)$ est donné par

$$\Lambda_l(z_i) = v_l \log \frac{\frac{\hat{\sigma}_{l,1}^2}{\hat{\sigma}_{l,2}^2} \frac{\hat{\sigma}_{l,2}^2 - \hat{\sigma}_{l,1}^2}{2\hat{\sigma}_{l,2}^2 \hat{\sigma}_{l,1}^2} v_l (z_{l,i}^{wapp} - \hat{y}_l)^2}{\hat{\sigma}_{l,1}^2} \quad (8)$$

où $\hat{\sigma}_{l,1}^2$ est l'estimation de la variance donnée par l'équation (5) en utilisant les paramètres a_1 et b_1 de l'appareil photographique 1 (respectivement, $\hat{\sigma}_{l,2}^2$ pour l'appareil photographique 2). Quelques calculs permettent de montrer que les deux premiers moments (espérance mathématique et variance) du rapport de vraisemblance $\Lambda_l(z_l^{wapp})$ sont donnés, dans le cas où l'appareil photographique n°1 est utilisé, par:

$$E_1 \log \Lambda_l(z_l^{wapp}) = \frac{v_l}{2} g(y_l) + v_l h(y_l)$$

$$\begin{aligned} \text{Var}_1 \left[\log \Lambda_l(z_l^{wapp}) \right] &= \frac{v_l^2}{4} \text{Var}_0 \left[g(\hat{y}_l) \right] + v_l^2 \text{Var}_0 \left[h(\hat{y}_l) \right] \\ &+ 2v_l \left(\text{Var}_0 \left[h(\hat{y}_l) \right] + (h(y_l))^2 \right) \end{aligned}$$

où les fonctions $g(x)$, $h(x)$ et $q(x)$ sont données par:

$$g(x) = \log \frac{a_1x + b_1}{a_2x + b_2} ; \quad h(x) = \frac{1 - \frac{a_1x + b_2}{a_1x + b_2}}{2} ; \quad q(x) = \frac{\frac{a_2x + b_2}{a_1x + b_1} - 1}{2}$$

[0041] Dans le cas où l'appareil n°2 est utilisé, les mêmes calculs permettent de

5montrer que les deux premiers moments du rapport de vraisemblance $\Lambda_l(z_l^{wapp})$

sont donnés, par :

$$\begin{aligned} E_2 \left[\log \Lambda_l(z_l^{wapp}) \right] &= \frac{v_l}{2} g(y_l) + v_l q(y_l) \\ \text{Var}_2 \left[\log \Lambda_l(z_l^{wapp}) \right] &= \frac{v_l^2}{4} \text{Var}_1 \left[g(y_l) \right] + v_l^2 \text{Var}_1 \left[q(y_l) \right] + 2v_l \left(\text{Var}_1 \left[q(y_l) \right] + (q(y_l))^2 \right) \end{aligned}$$

[0042] En définissant les deux équations ci-dessus, les deux premiers moments

10du rapport de vraisemblance $\Lambda_l(Z)$ sont donnés, dans le cas où l'appareil

photographique $j = \{1,2\}$ est utilisé, par :

$$\begin{aligned} m_j &= E_j \left[\log \Lambda_l(z_l^{wapp}) \right] = \frac{1}{L} \sum_{l=1}^L E_j \left[\log \Lambda_l(z_l^{wapp}) \right] \\ s_j^2 &= \text{Var}_j \left[\log \Lambda_l(z_l^{wapp}) \right] = \frac{1}{L} \sum_{l=1}^L \text{Var}_j \left[\log \Lambda_l(z_l^{wapp}) \right] \end{aligned} \quad (9)$$

[0043] En prenant le théorème de la limite centrale de Lindeberg [Lehmann2005,

15Théorème 11.2.5], les propriétés des variables aléatoires Gaussiennes permettent

d'obtenir la valeur du seuil de décision suivante :

$$\tau_{\alpha 0} = s_l \Phi^{-1}(1 - \alpha/2) + m_0 \quad (10)$$

où Φ^{-1} est la réciproque de la fonction de répartition Gaussienne centrée réduite

qui permet de garantir que asymptotiquement (lorsque le nombre total de pixels

MN $\rightarrow \infty$) la probabilité d'erreur de première espèce du test (7) vérifie $\alpha_0(\delta_1(Z)) \leq \alpha_0$.

L'erreur de première espèce est caractérisée par la prise de décision « Appareil n°2 » lorsque l'image inspectée provient de l'appareil n°1. Néanmoins en s'intervertissant le rôle des appareils photographiques, il est possible de garantir la probabilité de l'erreur de seconde espèce (réciproquement décider « Appareil n°1 » lorsque l'image inspectée provient de l'appareil n°2).

[0044] La puissance du test est définie comme la probabilité que le test prenne la décision « Appareil n°2 » lorsque l'image inspectée provient effectivement de l'appareil n°2. En utilisant le théorème de la limite centrale de Lindeberg [Lehmann2005, Théorème 11.2.5], les propriétés des variables aléatoires Gaussiennes permettent d'obtenir la puissance du test proposé :

$$\beta = 1 - \Phi\left(\frac{\sigma_0}{\sigma_1} \Phi^{-1}(1 - \alpha_0) - \frac{m_1 - m_0}{s_1}\right) \quad (11)$$

[0045] Dans ce second cas, l'utilisateur souhaite décider si la photographie numérique inspectée peut provenir d'un appareil photographique n°1 dont les propriétés du bruit sont connues (les paramètres notés a_1 et b_1). Les paramètres a^* et b^* de l'appareil photographique potentiellement à l'origine de la photographie ne sont pas connus. Cependant, on suppose dans la suite que ces derniers vérifient $a_1 < a^*$ et $b_1 < b^*$ ou bien $a_1 > a^*$ et $b_1 > b^*$.

[0046] On procède alors à un test, basé sur le rapport de vraisemblance précédent, dont la fonction de décision est donnée par

$$\begin{aligned} & \text{Appareil 1 si } \Lambda_2(Z) = \Lambda_2(z_l^{wap}) \geq \tau_{\alpha_0}^{\min}, \tau_{\alpha_0}^{\max} \\ & \text{Sinon} \quad \text{Appareil différent} \end{aligned} \quad (12)$$

où le seuil de décision est défini a priori par l'équation (14) de sorte que la probabilité d'erreur exigée soit garantie. Après quelques calculs, le logarithme du rapport de vraisemblance $\Lambda_2(z_i)$ est donné par:

$$\Lambda_2(z_l^{wapp}) = \sum_{i=1}^{v_l} \frac{(z_{l,i}^{wapp} - \hat{y}_l)^2}{\hat{\sigma}_l^2} \quad (13)$$

où $\hat{\sigma}_l^2$ est l'estimation de la variance donnée par l'équation (5) en utilisant les paramètres a_1 et b_1 de l'appareil photographique n°1 (seul appareil considéré comme étant connu dans ce second cas).

[0047] En utilisant les propriétés de la loi Gaussienne centrée réduite, la quantité

$\Lambda_2(z_l^{wapp})$ est distribuée, lorsque l'appareil n°1 est à l'origine de la photographie

inspectée, suivant la loi de distribution suivante :

$\Lambda_2(z_l^{wapp}) : \chi_{v_l}^2$ où $\chi_{v_l}^2$ représente la loi de distribution du χ^2 (« chi2 ») à v_l

degrés de liberté. Ainsi, il en découle que la quantité $\Lambda_2(Z)$ est distribuée suivant

la loi de distribution suivante : $\Lambda_2(Z) : \chi_{MN}^2$

[0048] Finalement, quelque soit le nombre MN de pixels, les propriétés de la loi de

distribution du χ^2 permettent d'établir que les seuils de décision donnés par:

$$\tau_{\alpha 0}^{\min} = \left[\chi_{MN}^2 \right]^{-1} \left(\frac{\alpha 0}{2} \right) \quad \text{et} \quad \tau_{\alpha 0}^{\max} = \left[\chi_{MN}^2 \right]^{-1} \left(1 - \frac{\alpha 0}{2} \right) \quad (14)$$

où $\left[\chi_{MN}^2 \right]^{-1}$ est la réciproque de la fonction de répartition du χ^2 , permettant de

garantir que la probabilité d'erreur de première espèce du test (12) vérifie :

$$\alpha 0(\delta_1(Z)) \leq \alpha 0$$

[0049] Enfin, la puissance du test, c'est-à-dire la probabilité de rejeter l'hypothèse que la photographie inspectée provient de l'appareil n°1 lorsque c'est

effectivement le cas, dépend des paramètres a^* et b^* de l'appareil photographique (inconnu) à la source de lumière inspectée.

[0050] Dans ce troisième cas, le test vise à vérifier si une image bruitée a été prise par l'appareil 0 lorsque les paramètres de l'image sont inconnus. Le problème est rendu plus compliqué parce que l'hypothèse alternative devient complexe lorsque l'on tient compte de la présence de paramètres de nuisance. Pour surmonter cette sorte de problème le test RVG (Rapport de Vraisemblance Généralisé) est utilisé.

On réécrit les deux hypothèses alternatives pour une observation.

$$10 \begin{cases} \mathcal{H}_0 : z_i \sim P_{y_i, a_0, b_0} = \mathcal{N}(y_i, \sigma_{i,0}^2) \\ \mathcal{H}_1 : z_i \sim P_{y_i, a, b} = \mathcal{N}(y_i, \sigma_i^2) \quad (a, b) \neq (a_0, b_0) \end{cases}$$

Dans cette équation : $\sigma_{i,0}^2 = a_0 y_i + b_0 \quad \sigma_i^2 = a y_i + b$

Les paramètres (a, b) et y_i sont inconnus. En utilisant leur EMV (Estimateur du Maximum de Vraisemblance), le test RVG devient :

$$15 \delta_3 = \begin{cases} \mathcal{H}_0 & \text{if } \hat{\Lambda}_3(Z) = \prod_{i=1}^M \hat{\Lambda}_3(z_i) < \tau_{a_0} \\ \mathcal{H}_1 & \text{if } \hat{\Lambda}_3(Z) = \prod_{i=1}^M \hat{\Lambda}_3(z_i) \geq \tau_{a_0} \end{cases}$$

où RVG $\hat{\Lambda}_3(z_i)$ est donné par :

$$\hat{\Lambda}_3(z_i) = \frac{\sup_{a,b} \sup_{y_i,a,b} P_{y_i,a,b}}{\sup_{y_i} P_{y_i,a_0,b_0}} = \frac{P_{y_i,\hat{a},\hat{b}}}{P_{y_i,a_0,b_0}}$$

[0051] Il est particulièrement important de calculer d'une manière analytique les performances statistiques du RVG afin de garantir une probabilité d'erreur prescrite.

Estimation des paramètres de l'image.

[0052] L'algorithme d'estimation est clairement indiqué dans le document [Foi2008]. Cette étape, montrée à la case K20, a été brièvement décrite au paragraphe [0032] et est ici présentée plus en détails. Tout d'abord l'image bruitée est traitée dans le domaine des « ondelettes » en vue de faciliter l'analyse du 5bruit. Elle est ensuite segmentée en K ensembles de niveaux sans chevauchements ; les données sont donc « adoucies ». Il est considéré d'une manière raisonnable que chaque ensemble de même niveau d'intensité forme une région uniforme. Par conséquent, les pixels de chaque ensemble de même niveau partagent la même moyenne et la même variance, les pixels sont statistiquement 10indépendants et identiquement distribués.

Soit :

$$z_i^{wapp} = \left\{ z_{i,i}^{wapp} \right\}_{i=1}^{n_k}$$

le vecteur des coefficients d'approximations de l'ondelette représentant l'ensemble de niveau k qui contient n_k pixels. Ces coefficients sont distribués comme suit:

$$15 \left\{ \begin{array}{l} z_k^{wapp} \sim \mathcal{N}\left(y_k, \|\varphi\|_2^2 \sigma_{k,0}^2\right) \text{ dans le cas } \mathcal{H}_0 \\ z_k^{wapp} \sim \mathcal{N}\left(y_k, \|\varphi\|_2^2 \sigma_k^2\right) \text{ dans le cas } \mathcal{H}_1 \end{array} \right.$$

où φ est la fonction d'échelle d'ondelette normalisé en 2D.

Il en découle que le EMV de l'espérance y_k dans chaque ensemble de niveau k, peut être défini par :

$$\hat{y}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} z_{k,i}^{wapp}$$

20La distribution de \hat{y}_k peut donc être définie par :

$$\left\{ \begin{array}{l} \hat{y}_k \sim \mathcal{N}\left(y_k, c_k \sigma_{k,0}^2\right) \text{ dans le cas } \mathcal{H}_0 \\ \hat{y}_k \sim \mathcal{N}\left(y_k, c_k \sigma_k^2\right) \text{ dans le cas } \mathcal{H}_1 \end{array} \right.$$

$$\text{où } c_k = \frac{\|\varphi\|_2^2}{n_k}$$

Estimation des paramètres d'un appareil photographique.

[0053] On va présenter comment estimer les paramètres d'un appareil photographique au moyen d'une image donnée bruitée.

5 Soit : $z_k^{w \det} = \left\{ z_{k,i}^{w \det} \right\}_{i=1}^{n_k}$ le vecteur des coefficients de détail de l'ondelette représentant l'ensemble de niveau « k ». En considérant l'hypothèse que sa distribution peut être déterminée par :

$$z_k^{w \det} : N \left(0, \sigma_k^2 \right)$$

Une estimation non biaisée de σ_k^2 est alors donnée par :

$$10 \sigma_k^2 = \frac{1}{n_k - 1} \sum_{i=1}^{n_k} \left(z_{k,i}^{w \det} - \bar{z}_k^{w \det} \right)^2 \approx \sigma_k^2 \rightarrow \hat{\sigma}_k^2$$

$$\text{où } \bar{z}_k^{w \det} = \frac{1}{n_k} \sum_{i=1}^{n_k} z_{k,i}^{w \det} .$$

[0054] Les estimées $\hat{\sigma}_k^2$ sont distribuées, à un facteur d'échelle près, suivant une loi de distribution du χ^2 avec $n_k - 1$ degrés de liberté. Cette distribution peut asymptotiquement, pour n_k suffisamment grand, être approchée, avec une bonne
15 précision, par la loi de distribution normale suivante :

$$\hat{\sigma}_k^2 : N \left(\sigma_k^2, e_k \sigma_k^4 \right) \text{ où } e_k = \frac{2}{n_k - 1}$$

Alors la variance $\hat{\sigma}_k^2$ peut être traitée comme un modèle hétéroscédastique de l'espérance $\hat{\mu}_k$ qui peut s'écrire :

$$\hat{\sigma}_k^2 = b + a \hat{y}_k + s_k \varepsilon_k \quad (k=1, \dots, K)$$

$$20 \text{ où : } s_k^2 = \varepsilon_k (a \hat{y}_k + b)^2 - a^2 c_k (a \hat{y}_k + b)$$

les résidus ε_k sont indépendants et identiquement distribués comme une distribution normale standard.

[0055] L'hétéroscédasticité dans le modèle est régit par les différents s_k^2 . Il est possible d'exploiter l'approche des moindres carrés pondérés dans le but de minimiser les résidus pondérés et de fournir un modèle convenable. Une stratégie bien connue pour estimer (a, b) est d'obtenir, dans un premier temps les estimées \hat{s}_k^2 et dans un deuxième temps d'appliquer les moindres carrés pondérés en

utilisant une pondération telle que : $\hat{w}_k = \frac{1}{\hat{s}_k^2}$.

[0056] Les moindres carrés donnant une estimation (\hat{a}_L, \hat{b}_L) utilisée comme estimation préliminaire sont définis par :

$$\hat{\beta}_L = (X^T X)^{-1} X^T V \quad \text{où:}$$

$$X = \begin{pmatrix} 1 & \hat{y}_1 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & \hat{y}_k \end{pmatrix} \quad V = \begin{pmatrix} \hat{\sigma}_1^2 \\ \cdot \\ \cdot \\ \cdot \\ \hat{\sigma}_k^2 \end{pmatrix}$$

10

Il en découle que les estimations de \hat{s}_k^2 peuvent être directement calculées par :

$$\hat{s}_k^2 = e_k (\hat{a}_L \hat{y}_k + \hat{b}_L)^2 - \hat{\sigma}_L^2 c_k (\hat{a}_L \hat{y}_k + \hat{b}_L)$$

En utilisant les pondérations estimées : $\hat{w}_k = \frac{1}{\hat{s}_k^2}$, les estimateurs WLS peuvent

15 être calculés par :

$$\hat{\beta}_W = (X^T W X)^{-1} X^T W V$$

$$\text{où : } W = \text{diag}(\hat{w}_1, \dots, \hat{w}_k)$$

[0057] Il est montré dans le document [Shao1989] que compte tenu de la normalité et de la forme connue de s_k , le WLS donne une estimation en utilisant \hat{w}_k qui devient d'une manière asymptotique équivalente aux estimations EMV

(maximum de vraisemblance) pour un grand échantillonnage. En outre les estimations WLS et EMV partagent la même distribution normale asymptotique.

Variance des estimées WLS.

5[0058] Comme les estimées des moindres carrés ont la propriété d'être non

biaisées, il s'en suit que: $E\left[\frac{\hat{s}_k^2}{s_k^2}\right] = 1$ et que :

$$\begin{aligned} \text{Var}\left[\frac{\hat{s}_k^2}{s_k^2}\right] &= e_k^2 \text{Var}\left[\left(\hat{a}_L \hat{y}_k + \hat{b}_L\right)^2\right] + c_k^2 \text{Var}\left[\hat{a}_L^2 \left(\hat{a}_L \hat{y}_k + \hat{b}_L\right)\right] \\ &= \sigma(n_k^2) \end{aligned}$$

En conséquence, on peut calculer d'une manière approximative \hat{s}_k^2 et \hat{w}_k par :

$$\begin{aligned} \hat{s}_k^2 &\approx s_k^2 \\ \hat{w}_k &\approx w_k \end{aligned}$$

10 On peut réécrire les estimées WLS :

$$\begin{aligned} \hat{a}_w &= \frac{A_1}{A_2} \\ \hat{b}_w &= \bar{\sigma}^2 - \hat{a}_w \bar{y} \end{aligned}$$

où :

$$A_1 = \sum_{k=1}^K w_k (\hat{y}_k - \bar{y}) \hat{\sigma}_k^2 \quad \text{et} \quad A_2 = \sum_{k=1}^K w_k (\hat{y}_k - \bar{y})^2$$

$$\bar{y} = \frac{\sum_{k=1}^K w_k \hat{y}_k}{\sum_{k=1}^K w_k} \quad \bar{\sigma}^2 = \frac{\sum_{k=1}^K w_k \hat{\sigma}_k^2}{\sum_{k=1}^K w_k}$$

15[0059] On peut observer que \bar{y} et $\bar{\sigma}^2$ se présentent sous forme de distribution gaussienne ayant respectivement les espérances et les variances :

$$\mathbb{E}(\bar{y}) = y = \frac{\sum_{k=1}^K w_k y_k}{\sum_{k=1}^K w_k}$$

$$\text{Var}(\bar{y}) = \frac{\sum_{k=1}^K w_k^2 c_k \sigma_k^2}{\left(\sum_{k=1}^K w_k\right)^2}$$

$$\mathbb{E}(\bar{\sigma}^2) = \sigma^2 = \frac{\sum_{k=1}^K w_k \sigma_k^2}{\sum_{k=1}^K w_k} = ay + b$$

$$\text{Var}(\bar{\sigma}^2) = \frac{\sum_{k=1}^K w_k^2 e_k \sigma_k^4}{\left(\sum_{k=1}^K w_k\right)^2}$$

[0060] En application de la méthode du Delta [Lehmann2005, Théorème 11.2.14], il découle que pour deux variables aléatoires X et Y indépendantes, si on définit respectivement, m_x, s_x^2 et m_y, s_y^2 , l'espérance et la variance des variables aléatoires

X et Y , alors on a une approximation du rapport $R = \frac{X}{Y}$ selon les relations ci-dessous :

$$E[R] = \frac{m_x}{m_y} \qquad \text{Var}[R] = \frac{m_x^2}{m_y^2} \frac{s_x^2}{m_x^2} + \frac{s_y^2}{m_y^2}$$

10[0061] Il découle également de la méthode du Delta [Lehmann2005, Théorème

11.2.14], que l'espérance et la variance de la variable aléatoire $S = \log \frac{X}{Y}$ peuvent

être calculées d'une manière approximative par :

$$E[S] = \log \frac{m_x}{m_y}$$

$$\text{Var}[S] = \frac{s_x^2}{m_x^2} + \frac{s_y^2}{m_y^2}$$

15En considérant l'équation (4) il s'en suit :

$\hat{y}_k - y \sim \mathcal{N}\left(y_k - y, c_k \sigma_k^2 + \text{Var}(\bar{y})\right)$ dont il découle finalement :

$$\mathbb{E}\left[w_k (\hat{y}_k - \bar{y}) \hat{\sigma}_k^2\right] = w_k (y_k - y) \sigma_k^2$$

$$\text{Var}\left[w_k (\hat{y}_k - \bar{y}) \hat{\sigma}_k^2\right] = w_k^2 \sigma_k^4 \left[e_k (y_k - y)^2 + c_k \sigma_k^2 + \text{Var}(\bar{y}) + o\left(n_k^{-2}\right) \right]$$

En tenant compte du théorème de la limite centrale de Linderberg [Lehmann2005, Théorème 11.2.5] : A_1 est une valeur de type gaussien présentant une espérance et une variance :

$$\mathbb{E}(A_1) = \sum_{k=1}^K w_k (y_k - y) \sigma_k^2$$

$$\text{Var}(A_1) = \sum_{k=1}^K w_k^2 \sigma_k^4 \left[e_k (y_k - y)^2 + c_k \sigma_k^2 + \text{Var}(\bar{y}) \right]$$

On peut alors écrire:

$$\begin{aligned} 10 \quad \mathbb{E}\left(w_k (\hat{y}_k - \bar{y})^2\right) &= w_k (y_k - y)^2 \\ \text{Var}\left(w_k (\hat{y}_k - \bar{y})^2\right) &= 4w_k^2 (y_k - y)^2 \left(c_k \sigma_k^2 + \text{Var}(\bar{y}) + o\left(n_k^{-2}\right) \right) \end{aligned}$$

En conséquence, et compte tenu du théorème de la limite centrale de Lindeberg [Lehmann2005, Théorème 11.2.5], A_2 se présente sous la forme d'une distribution gaussienne ayant une espérance et une variance telles que :

$$\begin{aligned} 15 \quad \mathbb{E}(A_2) &= \sum_{k=1}^K w_k (y_k - y)^2 \\ \text{Var}(A_2) &= 4 \sum_{k=1}^K w_k^2 (y_k - y)^2 \left(c_k \sigma_k^2 + \text{Var}(\bar{y}) \right) \end{aligned}$$

De la méthode du Delta, il en découle :

$$\text{Var}[\hat{a}_w] = \frac{\text{Var}[A_1] + a^2 \text{Var}[A_2]}{\mathbb{E}^2[A_2]}$$

$$\text{Var}[\hat{\delta}_w] = \text{Var}(\sigma^2) + a^2 \text{Var}(\bar{y}) + y^2 \text{Var}[\hat{a}_w]$$

En outre la covariance entre les deux estimées peut se définir par :

$$\text{Cov}[\hat{a}_w, \hat{b}_w] = \text{Cov}[\hat{a}_w, \sigma^2 - \hat{a}_w \bar{y}] = -y \text{Var}[\hat{a}_w]$$

Performance statistique du test.

[0062] Le test RVG (Rapport de Vraisemblance Généralisé) peut être réécrit sous

la forme :

$$\delta_3 = \begin{cases} H_0 & \text{if } \hat{\Lambda}_3(Z) = \prod_{k=1}^K \prod_{i=1}^{n_k} \hat{\Lambda}_3(z_{k,i}^{wapp}) < \tau_{a0} \\ H_1 & \text{if } \hat{\Lambda}_3(Z) = \prod_{k=1}^K \prod_{i=1}^{n_k} \hat{\Lambda}_3(z_{k,i}^{wapp}) \geq \tau_{a0} \end{cases}$$

où τ_{a0} est la solution de l'équation :

$$P_0(\hat{\Lambda}_3(Z) \geq \tau_{a0}) = a_0. \quad \text{Le test RVG } \hat{\Lambda}_3(z_{k,i}^{wapp}) \text{ est donné par :}$$

$$\hat{\Lambda}_3(z_{k,i}^{wapp}) = \frac{\bar{\sigma}_{k,0}}{\bar{\sigma}_k} \exp \left[\frac{\bar{\sigma}_k^2 - \bar{\sigma}_{k,0}^2}{2 \|\varphi\|_2^2 \bar{\sigma}_k^2 \bar{\sigma}_{k,0}^2} (z_{k,i}^{wapp} - \hat{y}_k)^2 \right]$$

où

$$\bar{\sigma}_{k,0}^2 = a_0 \hat{y}_k + b_0 \quad \text{et} \quad \bar{\sigma}_k^2 = \hat{a}_w \hat{y}_k + \hat{b}_w$$

La variance de $\bar{\sigma}_{k,0}^2$ peut être donnée par :

$$\begin{cases} \text{Var}_0[\bar{\sigma}_{k,0}^2] = a^2 c_k \bar{\sigma}_{k,0}^2 + y_k^2 \text{Var}[\hat{a}_w] + 2y_k \text{Cov}[\hat{a}_w, \hat{b}_w] + \text{Var}[\hat{b}_w] \\ \text{Var}_1[\bar{\sigma}_k^2] = a^2 c_k \bar{\sigma}_k^2 + y_k^2 \text{Var}[\hat{a}_w] + 2y_k \text{Cov}[\hat{a}_w, \hat{b}_w] + \text{Var}[\hat{b}_w] \end{cases}$$

Sur la base des logarithmes, la fonction de vraisemblance en logarithme se

calculer directement :

$$\begin{aligned} \log \hat{\Lambda}_3(z_{k,i}^{wapp}) &= \log \frac{\bar{\sigma}_{k,0}}{\bar{\sigma}_k} + \frac{\bar{\sigma}_k^2 - \bar{\sigma}_{k,0}^2}{2 \|\varphi\|_2^2 \bar{\sigma}_k^2 \bar{\sigma}_{k,0}^2} (z_{k,i}^{wapp} - \hat{y}_k)^2 \\ &= \frac{T}{2} + \frac{1-Z}{2} \frac{(z_{k,i}^{wapp} - \hat{y}_k)^2}{\|\varphi\|_2^2 \bar{\sigma}_{k,0}^2} \end{aligned}$$

$$\text{où } Z = \frac{\partial \log \bar{\sigma}_{k,0}}{\partial \rho} \quad \text{et} \quad T = \log \frac{\partial \bar{\sigma}_{k,0}}{\partial \rho}$$

En conséquence, les deux premiers moments de $\log \hat{\Lambda}_3(z_{k,i}^{wapp})$ sous l'hypothèse H_0 , selon laquelle l'image provient de l'appareil n°1, sont donnés par :

$$E_0 \log \hat{\Lambda}_3(z_{k,i}^{wapp}) = \frac{E_0[T]}{2} + \frac{1 - E_0[Z]}{2}$$

$$Var_0 \log \hat{\Lambda}_3(z_{k,i}^{wapp}) = \frac{Var_0[T] + 3 Var_0[Z]}{4} + \frac{(1 - E_0[Z])^2}{2}$$

On peut aussi écrire la vraisemblance en logarithme de $\hat{\Lambda}_3(z_{k,i}^{wapp})$

$$\log \hat{\Lambda}_3(z_{k,i}^{wapp}) = \log \frac{\bar{\sigma}_{k,0}}{\bar{\sigma}_{k,1}} + \frac{\bar{\sigma}_{k,1}^2 - \bar{\sigma}_{k,0}^2}{2 \|\varphi\|_2^2 \sigma_{k,1}^2 \sigma_{k,0}^2} (z_{k,i}^{wapp} - \hat{y}_k)^2$$

$$= \frac{T}{2} + \frac{Q-1}{2} \frac{(z_{k,i}^{wapp} - \hat{y}_k)^2}{\|\varphi\|_2^2 \sigma_k^2}$$

où $Q = \frac{\partial^2 \log \hat{\Lambda}_3}{\partial \theta^2}$.

Les deux premiers moments de $\log \hat{\Lambda}_3(z_{k,i}^{wapp})$ sous l'hypothèse H_1 , selon laquelle l'image provient de l'appareil n°2, sont donnés par les formules suivantes :

$$E_1 \log \hat{\Lambda}_3(z_{k,i}^{wapp}) = \frac{E_1[T]}{2} + \frac{E_1[Q] - 1}{2}$$

$$Var_1 \log \hat{\Lambda}_3(z_{k,i}^{wapp}) = \frac{Var_1[T] + 3 Var_1[Q]}{4} + \frac{(E_1[Q] - 1)^2}{2}$$

[0063] Enfin, il est possible d'envisager un quatrième cas d'utilisation du système d'identification d'un appareil photographique en vue de déterminer, de façon supervisée, si une zone de l'image n'a pas été falsifiée (par copier/coller depuis une autre photographie ou par suppression d'un élément). On entend ici par "supervisé" le fait que l'utilisateur souhaite s'assurer de l'intégrité d'une zone préalablement définie. Le principe est alors d'appliquer la méthode d'identification aux deux "sous-images" issues respectivement de la zone ciblée par l'utilisateur et de son complémentaire (le reste de l'image). Si l'élément inspecté provient d'une autre photographie et a été ajouté par

copier/coller, les propriétés de bruit seront différentes ce que le système proposé sera capable d'identifier (en supposant que les photographies n'ont pas été prise avec dans les mêmes conditions d'acquisition et avec le même modèle d'appareil photographique, ce qui semble raisonnable).

5

[0064] Les figures 3 et 4 présentent une série des paramètres caractérisant les propriétés d'hétéroscédasticité du bruit de quelques modèles d'appareils ; ces résultats sont obtenus avec la méthode proposée dans [Foi2008] et illustrent la séparabilité des modèles d'appareils photographiques.

10[0065] La figure 5 illustre la comparaison des propriétés du bruit entre les trois appareils photographiques suivants : Canon EOS7D, Nikon D60 et Nikon D3000. La figure 5 présente la variance du bruit d'acquisition en fonction de l'espérance mathématique de la valeur des pixels, voir relations (3) et (5). On constate que les trois modèles d'appareils photographiques distincts pour lesquelles les paramètres 15a et b, caractérisant cette relation, ont été estimés présentent des propriétés d'hétéroscédasticité fort différentes. Cette figure permet de visualiser la différence qui existe entre les propriétés des bruits présents dans des images capturées par des modèles d'appareils photographiques distincts.

[0066] La figure 6, illustre les résultats obtenus pour le cas où les appareils 20photographique 1 et 2 sont connus. Cette figure montre la comparaison entre la puissance de détection théorique (traits pointillés) et celle obtenue lors de simulations numériques (traits pleins). Les résultats présentés dans la figure 6 ont été obtenus à partir d'une base d'images naturelles provenant de deux modèles d'appareil photographique distincts. Les paramètres de bruit, a_1 , b_1 et a_2 , b_2 de 25chaque appareil ont été calculés sur l'ensemble des images de cette base de données (les appareils utilisées sont les NikonD60 et NikonD3000 présentés sur

la figure 5). On constate enfin dans la figure 6 que, outre la puissance de détection très importante qu'il est possible d'obtenir pour une probabilité de fausse-alarme très faible, que les résultats théoriques et empiriques sont en très bonne adéquation.

5[0067] La figure 7 montre les résultats obtenus dans le cas où l'appareil photographique 2 est inconnu. Le résultat est simulé avec l'intensité « y » croissante de la gauche vers la droite, image d'une mire constituée de bandes grises allant du noir au blanc, photographiée avec deux modèles d'appareil photographique distincts. Il a été considéré que les propriétés de l'un des deux
10appareils photographique sont connues (ce modèle d'appareil photographique ayant été utilisée pour obtenir les résultats présentés dans la figure 6. La figure 7 présente en ordonné le numéro de ligne inspecté et en abscisse la valeur du rapport de vraisemblance calculé sans connaissance sur les propriétés de l'appareil photographique n°2. La figure 7 montre que 1) à un facteur d'échelle
15près, le rapport de vraisemblance est normalisé de sorte que sa distribution reste identique pour l'appareil photographique n°1 et, 2) que pour l'appareil photographique n°2, le rapport de vraisemblance, sur lequel est fondé le système d'identification proposé, prend des valeurs largement supérieur (ce qui permet une prise de décision).

20[0068] La méthode d'identification du modèle d'appareil photographique proposée répond aux deux inconvénients ou faiblesses des méthodes brièvement présentées dans l'état de l'art. Premièrement, dans les méthodes de l'état de l'art, les performances ne sont pas établies et de plus ces méthodes peuvent être mises en échec par la calibration de l'appareil photographique. La méthode
25proposée repose sur la propriété d'hétéroscédasticité du bruit, caractéristique inhérente à l'acquisition des photographies. Cette dernière est applicable de façon

générale quels que soient les traitements post-acquisition appliqués par un utilisateur (notamment en vue d'améliorer la qualité visuelle). En outre la modélisation paramétrique de la relation entre moyenne et variance de la valeur des pixels permet de fournir de façon analytique les performances du test proposé. Cet avantage permet notamment d'assurer le respect d'une contrainte prescrite sur la probabilité d'erreur.

[0069] Les principaux domaines d'applications de l'invention sont d'une part, la recherche de preuve à partir d'une image « compromettante » et d'autre part la garantie qu'une photographie a été acquise par un appareil donné.

10[0070] La méthode proposée peut être étendue au contrôle de l'intégrité d'une photographie. Le but est alors de garantir qu'une photographie n'a pas été modifiée/falsifiée depuis son acquisition. Cela permet par exemple de détecter les photographies comportant des éléments provenant d'un appareil photographique différent, i.e. importés après l'acquisition, ou encore d'assurer l'intégrité d'un
15document scanné ou photographié (un document juridique par exemple).

[0071] Le procédé de l'invention pourra être développé dans des logiciels spécialisés des fabricants logiciels, dans la recherche de preuve à partir de média numériques. Le procédé selon l'invention peut être utilisé auprès des tribunaux en vu de fournir un outil d'aide à la décision.

Bibliographie:

- [Shao1989]** "Asymptotic distribution of the weighted least squares estimator", Shao, J., *Annals of the Institute of Statistical Mathematics*, vol.41, no.2, pp 365-382, 1989.
- 5 **[Healey1994]** "Radiometric CCD camera calibration and noise estimation", Healey, G.E.; Kondepudy, R, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol.16 , no.3, pp 267-276, 1994
- [Lehmann2005]** "Testing Statistical Hypotheses", Lehmann, E.L. ; Romano, J.P. , 3ième édition, ISBN 0-387-98864-5, Springer, 2005.
- 10 **[Lukas2006]** "Digital camera identification from sensor pattern noise", Lukas, J.; Fridrich, J.; Goljan, M.; *Information Forensics and Security, IEEE Transactions on* , vol.1, no.2, pp. 205- 214, June 2006
- [Chen2008]** "Determining Image Origin and Integrity Using Sensor Noise", Chen, M.; Fridrich, J.; Goljan, M.; Lukas, J.; *Information Forensics and Security, IEEE Transactions on*, vol.3, no.1, pp.74-90, March 2008
- 15 **[Foi2008]** "Practical Poissonian-Gaussian Noise Modeling and Fitting for Single-Image Raw-Data", Foi, A.; Trimeche, M.; Katkovnik, V.; Egiazarian, K.; *IEEE Transactions on Image Processing*, vol.17, no.10, pp.1737-1754, Oct. 2008
- 20 **[Maillard2008]** "Detection of digital processing of images through a realistic model of CCD noise", Maillard, J.-B.; Levesque, D.; Deschenes, F. *International Conference on Pattern Recognition, ICPR*, 8-11 Dec. 2008

REVENDECATIONS

1. Système pour déterminer l'identification d'un premier appareil photographique (2) à partir d'une photographie (3), système comportant un dispositif d'analyse de photographies (10) pour traiter ladite photographie et
 5 un dispositif de traitement (12) pour établir des statistiques à partir de ladite analyse, le système opérant ladite identification sur la base de paramètres « a » et « b » définissant un appareil photographique et répondant à la relation ci-dessous :

$$z_{m,n} : N(y_{m,n}, \alpha y_{m,n} + b)$$

10 où $z_{m,n}$ est représentatif de la valeur d'un pixel de coordonnées m et n d'une photographie,,
 N (..) représente une distribution gaussienne

$y_{m,n}$ est l'espérance mathématique du pixel situé à la position m,n.

caractérisé en ce qu'il comprend en outre est prévu un organe de sortie (14)

15 qui opère par comparaison statistique avec un deuxième appareil photographique pour déterminer si la photographie a été prise par ledit premier appareil ou ledit deuxième appareil.

2. Système selon la revendication 1, caractérisé en ce que l'organe de sortie (14) fournit une indication sur l'identification d'un appareil photographique en
 20 certifiant l'exactitude de l'identification suivant une précision préalablement définie.

3. Procédé mis en œuvre dans le système de la revendication 1 caractérisé en ce qu'il comporte les étapes suivantes :

25 - lecture de la photographie en vue de déterminer les matrices descriptives (K12),

- estimation du contenu et du bruit de l'image (K14),
- estimation de la variance des pixels par niveaux d'intensité estimée du contenu (K22),
- estimation des paramètres caractérisant la relation moyenne variance du bruit (K24),
- exécution de tests statistiques en vue de donner l'identification. (K26).

4. Procédé selon la revendication 3, caractérisé en ce que à l'étape(K26) les tests statistiques sont exécutés en fonction de la probabilité d'erreur exigée et une décision est prise.

5. Procédé selon la revendication 3 caractérisé en ce que la photographie est en un format non compressé ou en un format compressé sans pertes.

6. Utilisation du procédé selon l'une des revendications 3 à 5 pour la détection, de façon supervisée, de la falsification d'une zone dans une image.

7. Utilisation du procédé selon l'une des revendications 3 à 5 dans la recherche de preuves à partir d'une image compromettante.

8. Application du procédé selon l'une des revendications 3 à 5 dans des logiciels spécialisés, dans la recherche de preuves à partir de média numériques.

Abrégé

L'invention concerne un système pour déterminer l'identification d'un premier appareil photographique à partir d'une photographie ainsi que le procédé mis en œuvre dans un tel système. Ce système comporte un dispositif (10) d'analyse de photographies pour traiter ladite photographie et un dispositif de traitement (12) pour établir des statistiques à partir de ladite analyse, le système opérant ladite identification sur la base de paramètres « a » et « b » définissant un appareil photographique selon l'invention le système comporte un organe de sortie (14) qui opère par comparaison statistique avec un deuxième appareil photographique pour déterminer si la photographie a été prise par ledit premier appareil ou ledit deuxième appareil.

Figure pour l'abrégé : figure 2