CAMERA MODEL IDENTIFICATION BASED ON HYPOTHESIS TESTING THEORY

Thanh Hai Thai, Rémi Cogranne, Florent Retraint

ICD - LM2S - Université de Technologie de Troyes - UMR STMR CNRS
12, rue Marie Curie - B.P. 2060 - 10010 Troyes cedex - France
E-mail : {thanh_hai.thai, remi.cogranne, florent.retraint}@utt.fr

ABSTRACT

This paper aims to study the problem of imaging device identification using the heteroscedastic property of uncompressed image noise. Noise variance depends on pixels intensity through two parameters which uniquely represent a camera model and hence, enable to identify imaging device. The decision problem is cast in the framework of hypothesis testing theory. First, the theoretical context in which both the inspected image parameters and imaging device properties are known is considered. The most powerful Likelihood Ratio Test (LRT) is presented and its detection performance is analytically calculated. Then, the practical situation when inspected image parameters are unknown, but imaging device properties remain known, is studied. Based on a simple yet efficient image model, the inspected image parameters are estimated. This leads to the designed Generalized Likelihood Ratio Test (GLRT) whose statistical performances are analytically given. Numerical simulations and experimentations on natural images show the relevance of the proposed approach.

Index Terms— Hypthesis testing, Information forensics, Noise and system model, Camera model identification.

1. INTRODUCTION

It is nowadays within the reach of anyone to easily manipulate a digital image by using image editing software. Therefore, the use of such medium as a reliable witness has been questioned, in a court for instance. In this context, it becomes a crucial and useful challenge for security forces to be able to warrant origin and integrity of digital images. This paper address the problem of imaging device identification, which belongs to digital forensics.

Several approaches have been proposed in the literature to solve the problem of imaging device identification [1]. The state-of-the-art methods investigate the image processing pipeline in order to identify characteristics which are unique for each camera and could thus identify the source camera. On the one hand, the majority of latest algorithms exploits specific fingerprint in the acquisition stage such as lens aberration [2], sensor imperfections [3]. On the other hand, the post-acquisition processes inherent to image acquisition have also been explored. Particularly, it is proposed in [4] to use CFA interpolation assuming that the interpolation algorithm and the design of the CFA filter pattern of each camera model are somewhat different from others. Additionally, an approach based on supervised learning has also been investigated in [5]. As in all applications of machine learning the two main difficulties are the choice of accurate features set and the establishment of detection performance which remain an open problem.

In an operational context, the design of an accurate detector might not be enough. The most important and challenging problem is to provide a test with analytically predictable results. Methods already proposed in the literature for the problem of camera identification from a single image are very interesting and efficient. However, they have been designed with a very limited exploitation of hypothesis testing theory and a lack of a statistical image model. Therefore their detection performance can only be approximated using simulation on a large database. Similarly exploiting imaging device imperfection might lead to a lack of robustness especially if one intent to correct such defect using a calibration process.

In this paper, it is proposed to exploit hypothesis testing theory based on a statistical noise model of natural raw images [6]. The main contribution is twofold. First, when image and camera parameters are known, the most powerful test is designed and its statistical performances are theoretically calculated. Second, the practical context in which the image parameters are unknown is considered. To overcome the difficulty of a presence of nuisance parameters, a GLRT is designed using their Maximum Likelihood Estimation (MLE) and the detection probability is analytically established. This enables to warrant a prescribed constraint on false-alarm probability.

The paper is organized as follows. The statistical noise model of natural raw images is presented and the problem of camera model identification is stated in section 2. The theoretical context in which inspected image parameters and camera noise properties are known is addressed in section 3. Based on the MLE proposed in [7], section 4 is devoted to design the GLRT and then establish statistical performances of the proposed test. Numerical results are presented in section 5 and section 6 concludes the paper.

2. STATISTICAL NOISE MODEL

Let $\mathbf{Z} = \{z_i\}_{i=1}^{M}$ be a vector representing a natural image of $M = M_x \times M_y$ grayscale level pixels. This paper mainly focuses on raw images which can be simply modeled by considering the image acquisition process [6]. Indeed, the photo-electron conversion essentially consists in a counting process

With the financial support from the Prevention of and Fight against Crime Programme of the European Union European Commission - Directorate-General Home Affairs. Research partially funded by Troyes University of Technology (UTT) strategic program COLUMBO.

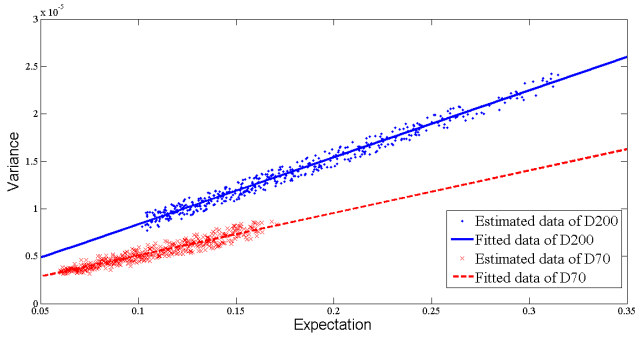

Fig. 1: Heteroscedastic relation between expectation and variance of a natural image.

which can be modeled as a Poisson process. Hence, the number of electrons collected on i-th pixel, denoted Ne_i is given by:

$$Ne_i \sim \mathscr{P}(\eta_i Np_i + Nt_i) \tag{1}$$

where Np_i is the number of incident photons, η_i is a conversion factor representing filter transmittance and quantum sensitivity, and Nt_m is the number of dark electrons generated by thermal noise. It is assumed that the photo-sensitivity and the thermal noise are constant for every pixel, the index i is therefore omitted from η and Nt.

The number of collected electrons is then transferred to a read-out unit. During the read-out process, recorded signal is corrupted by different sources of electronic noise which all can be modeled as a zero-mean Gaussian random variable with variance σ^2. Therefore, the value of raw pixel z_i is finally given by [6, 7]:

$$z_i \sim a\mathscr{N}(Ne_i, \sigma^2). \tag{2}$$

where a is a sensitivity factor related to ISO speed number. For the sake of simplification, the Normal approximation of the Poisson distribution may be exploited due to the large number of counted electrons. It finally follows that :

$$z_i \sim \mathscr{N}(\mu_i,\ a\mu_i + b) \tag{3}$$

where $\mu_i = a(\eta Np_i + Nt)$ and $b = a^2\sigma^2$.

Equation (3) shows that the expectation and variance of pixel follow a heteroscedastic relation in which the variance linearly depends on the expectation. Moreover, this relation is entirely defined by the two parameters a and b which change for different camera models. An example is given in figure 1 which illustrates both the heteroscedastic relation and the camera model dependence of parameters a and b for two *Nikon* reflex camera, which are expected to exhibit similar characteristics. In the present paper, it is proposed to exploit this unique characteristic to identify the imaging device which captured a given image.

3. LIKELIHOOD RATIO TEST FOR TWO SIMPLE HYPOTHESES

3.1. Problem statement

The test aims to identify either camera 0 or camera 1 captured the given image. Let us assume that the characteristics of two cameras are known (e.g (a_0, b_0) and (a_1, b_1) respectively). In such theoretical situation, it is desirable to decide between the two following hypothesis :

$$\begin{cases} \mathscr{H}_0 : z_i \sim P_{\mu_i,a_0,b_0} = \mathscr{N}(\mu_i, \sigma_{i,0}^2), \forall i \in \{1,\dots,M\} \\ \mathscr{H}_1 : z_i \sim P_{\mu_i,a_1,b_1} = \mathscr{N}(\mu_i, \sigma_{i,1}^2), \forall i \in \{1,\dots,M\} \end{cases} \tag{4}$$

where $\sigma_{i,j} = \sqrt{a_j\mu_i + b_j}$, $j \in \{0,1\}$. However, it can be observed that μ_i acts as a nuisance parameter when it has no interest for decision problem (4) while appearing in the model of raw pixel. Hence, let us firstly suppose that the parameter μ_i is exactly known. From Neyman-Pearson lemma [8, theorem 3.2.1], the most powerful test δ_1 over the class

$$\mathscr{K}_{\alpha_0} = \{\delta : \mathbb{P}_0[\delta_1(\mathbf{Z}) = \mathscr{H}_1] \leq \alpha_0\}$$

where $\mathbb{P}_j[\dots]$ represents the probability under hypothesis $\mathscr{H}_j, j \in \{0,1\}$, is given by the following decision rule :

$$\delta_1(\mathbf{Z}) = \begin{cases} \mathscr{H}_0 & \text{if} \quad \Lambda_1(\mathbf{Z}) = \prod_{i=1}^{M} \Lambda_1(z_i) < \tau_{\alpha_0} \\ \mathscr{H}_1 & \text{if} \quad \Lambda_1(\mathbf{Z}) = \prod_{i=1}^{M} \Lambda_1(z_i) \geq \tau_{\alpha_0} \end{cases} \tag{5}$$

where the decision threshold τ_{α_0} is the solution of the equation $\mathbb{P}_0[\Lambda_1(\mathbf{Z}) \geq \tau_{\alpha_0}] = \alpha_0$ to insure that $\delta_1 \in \mathscr{K}_{\alpha_0}$. This most powerful test maximizes the power

$$\beta_{\delta_1} = 1 - \mathbb{P}_1[\Lambda_1(\mathbf{Z}) < \tau_{\alpha_0}] \tag{6}$$

over the class $\mathscr{K}_{\alpha_0}$.

Under assumption that μ_i is known, the distributions P_{μ_i,a_j,b_j}, $j \in \{0,1\}$ are well defined. A short algebra shows that the Likelihood Ratio (LR) for one observation z_i is given by :

$$\Lambda_1(z_i) = \frac{P_{\mu_i,a_1,b_1}}{P_{\mu_i,a_0,b_0}} = \frac{\sigma_{i,0}}{\sigma_{i,1}} \exp\left[\frac{(z_i-\mu_i)^2}{2\sigma_{i,0}^2} - \frac{(z_i-\mu_i)^2}{2\sigma_{i,1}^2}\right] \tag{7}$$

Subsequently, the log-likelihood ratio can be computed as :

$$\log\Lambda_1(z_i) = \log\frac{\sigma_{i,0}}{\sigma_{i,1}} + \frac{\sigma_{i,1}^2 - \sigma_{i,0}^2}{2\sigma_{i,1}^2\sigma_{i,0}^2}\,(z_i-\mu_i)^2 \tag{8}$$

It can be noted that the first term of (8) remains unchanged under either $\mathscr{H}_0$ or $\mathscr{H}_1$. Hence, let us define the equivalent log LR for the sake of clarity :

$$\log\Lambda_1^\star(z_i) = \frac{\sigma_{i,1}^2 - \sigma_{i,0}^2}{\sigma_{i,1}^2\sigma_{i,0}^2}\,(z_i-\mu_i)^2 \tag{9}$$

Therefore, the equivalent test is defined by :

$$\delta_1^\star(\mathbf{Z}) = \begin{cases} \mathscr{H}_0 & \text{if} \quad \Lambda_1^\star(\mathbf{Z}) = \prod_{i=1}^{M} \Lambda_1^\star(z_i) < \tau_{\alpha_0}^\star \\ \mathscr{H}_1 & \text{if} \quad \Lambda_1^\star(\mathbf{Z}) = \prod_{i=1}^{M} \Lambda_1^\star(z_i) \geq \tau_{\alpha_0}^\star \end{cases} \tag{10}$$

where $\tau_{\alpha_0}^\star$ is the solution of the equation $\mathbb{P}_0[\Lambda_1^\star(\mathbf{Z}) \geq \tau_{\alpha_0}^\star] = \alpha_0$.

3.2. Statistical performance of the test

Characterizing the distribution of $\log\Lambda_1^\star(\mathbf{Z})$ is of crucial importance to analytically compute the test performance. However, the exact distribution of $\log\Lambda_1^\star(\mathbf{Z})$ can not be explicitly given. Alternatively, an asymptotic approach can be exploited because the number of pixels M is sufficiently large in any natural image. It follows from the Lindeberg central limit theorem (CLT) [8, theorem 11.2.5] that :

$$\frac{\sum_{i=1}^{M}\log\Lambda_1^\star(z_i)-\sum_{i=1}^{M}\mathbb{E}_j\Big[\log\Lambda_1^\star(z_i)\Big]}{\sqrt{\sum_{i=1}^{M}\mathrm{Var}_j\Big[\log\Lambda_1^\star(z_i)\Big]}}\xrightarrow{D}\mathscr{N}(0,1) \quad (11)$$

where $\mathbb{E}_j$ and Var_j denote respectively the mathematical expectation and variance under $\mathscr{H}_j$, $j\in\{0,1\}$, and $\xrightarrow{D}$ represents the convergence in distribution. As a result, $\log\Lambda_1^\star(\mathbf{Z})$ can be asymptotically distributed as :

$$\log\Lambda_1^\star(\mathbf{Z})\sim\mathscr{N}\Big(m_j^{(1)},v_j^{(1)}\Big) \quad (12)$$

where

$$m_j^{(1)}=\sum_{i=1}^{M}\mathbb{E}_j\Big[\log\Lambda_1^\star(z_i)\Big] \quad v_j^{(1)}=\sum_{i=1}^{M}\mathrm{Var}_j\Big[\log\Lambda_1^\star(z_i)\Big] \quad (13)$$

Because $z_i\sim\mathscr{N}(\mu_i,\sigma_{i,j}^2)$, it is immediate to verify that

$$\left(\frac{z_i-\mu_i}{\sigma_{i,j}}\right)^2\sim\chi_1^2 \quad \text{under} \quad \mathscr{H}_j \quad (14)$$

Hence, for one observation z_i, the expectation and variance of $\log\Lambda_1^\star(z_i)$ can be computed as :

$$\begin{cases}\mathbb{E}_0\Big[\log\Lambda_1^\star(z_i)\Big] & =1-\dfrac{\sigma_{i,0}^2}{\sigma_{i,1}^2}\\ \mathrm{Var}_0\Big[\log\Lambda_1^\star(z_i)\Big] & =2\Big(1-\dfrac{\sigma_{i,0}^2}{\sigma_{i,1}^2}\Big)^2\\ \mathbb{E}_1\Big[\log\Lambda_1^\star(z_i)\Big] & =\dfrac{\sigma_{i,1}^2}{\sigma_{i,0}^2}-1\\ \mathrm{Var}_1\Big[\log\Lambda_1^\star(z_i)\Big] & =2\Big(\dfrac{\sigma_{i,1}^2}{\sigma_{i,0}^2}-1\Big)^2\end{cases} \quad (15)$$

For the sake of clarity, let us define the normalized log LR

$$\log\tilde{\Lambda}_1(\mathbf{Z})=\frac{\log\Lambda_1^\star(\mathbf{Z})-m_0^{(1)}}{\sqrt{v_0^{(1)}}} \quad (16)$$

Theorem 1. *Assuming that parameters μ_i, a_j, b_j are exactly known, the decision threshold of the test $\delta_1^\star$ (10) is given by :*

$$\tilde{\tau}_{\alpha_0}=\Phi^{-1}(1-\alpha_0) \quad (17)$$

Theorem 2. *Consequently, the power of the test $\delta_1^\star$ based on $\log\tilde{\Lambda}_1(\mathbf{Z})$ is given by :*

$$\beta_{\delta_1^\star}=1-\Phi\left(\frac{m_0^{(1)}-m_1^{(1)}+\tilde{\tau}_{\alpha_0}\sqrt{v_0^{(1)}}}{\sqrt{v_1^{(1)}}}\right) \quad (18)$$

where Φ and Φ^{-1} are respectively the standard normal cumulative distribution function (cdf) and its inverse, $m_j^{(1)}$, $v_j^{(1)}$, $j=0,1$ are computed from (13) *and* (15)*.*

4. UNKNOWN IMAGE PARAMETERS

4.1. Problem statement

In a practical context, the image parameters are unfortunately unknown. In such case of nuisance parameters, the LR can not be directly computed because the distributions P_{μ_i,a_j,b_j}, $j\in\{0,1\}$ are not defined. Hence, the Generalized Likelihood Ratio Test based on the ML estimate $\hat{\mu}_i$ of μ_i can be used :

$$\delta_2(\mathbf{Z})=\begin{cases}\mathscr{H}_0 & \text{if} \quad \hat{\Lambda}_2(\mathbf{Z})=\prod_{i=1}^{M}\hat{\Lambda}_2(z_i)<\tau_{\alpha_0}\\ \mathscr{H}_1 & \text{if} \quad \hat{\Lambda}_2(\mathbf{Z})=\prod_{i=1}^{M}\hat{\Lambda}_2(z_i)\geq\tau_{\alpha_0}\end{cases} \quad (19)$$

where

$$\hat{\Lambda}_2(z_i)=\frac{\sup_{\mu_i}P_{\mu_i,a_1,b_1}}{\sup_{\mu_i}P_{\mu_i,a_0,b_0}}=\frac{P_{\hat{\mu}_i,a_1,b_1}}{P_{\hat{\mu}_i,a_0,b_0}} \quad (20)$$

The estimation quality of μ_i will affect on the statistical performance of the test. Therefore, an algorithm which can provide a reliable estimate $\hat{\mu}_i$ in a given noisy image is of crucial importance.

4.2. Expectation estimation

The estimation algorithm is clearly presented in [7]. First, the noisy image is transformed to the wavelet domain to facilitate the noise analysis and then segmented into K non-overlapping level sets where the data is smooth. It can be reasonably assumed that each level set tend to be an uniform region. Therefore, the pixels in each level set may be independent and identically distributed. Let $\mathbf{z}_k^{\text{wapp}}=\{z_{k,i}^{\text{wapp}}\}_{i=1}^{n_k}$ be the vector of wavelet approximation coefficients representing the level set k which contains n_k pixels. These coefficients are hence distributed as :

$$\mathbf{z}_k^{\text{wapp}}\sim\mathscr{N}(\mu_k,\|\varphi\|_2^2\,\sigma_{k,j}^2) \quad \text{under} \quad \mathscr{H}_j,\ j=0,1 \quad (21)$$

where $\sigma_{k,j}=\sqrt{a_j\mu_k+b_j}$ and φ is the 2-D normalized wavelet scaling function. Hence, the MLE of the expectation μ_k in each level set k can be defined as :

$$\hat{\mu}_k=\frac{1}{n_k}\sum_{i=1}^{n_k}z_{k,i}^{\text{wapp}} \quad (22)$$

The distribution of $\hat{\mu}_k$ can be subsequently defined as :

$$\hat{\mu}_k\sim\mathscr{N}(\mu_k,c_k\sigma_{k,j}^2) \quad \text{under} \quad \mathscr{H}_j,\ j\in\{0,1\} \quad (23)$$

where $c_k = \frac{\|\varphi\|_2^2}{n_k}$.

As a result, the GLRT can be written as follows :

$$\delta_2 = \begin{cases} \mathscr{H}_0 \quad \text{if} \quad \hat{\Lambda}_2(\mathbf{Z}) = \prod_{k=1}^{K}\prod_{i=1}^{n_k} \hat{\Lambda}_2(z_{k,i}^{\text{wapp}}) < \tau_{\alpha_0} \\ \mathscr{H}_1 \quad \text{if} \quad \hat{\Lambda}_2(\mathbf{Z}) = \prod_{k=1}^{K}\prod_{i=1}^{n_k} \hat{\Lambda}_2(z_{k,i}^{\text{wapp}}) \geq \tau_{\alpha_0} \end{cases} \tag{24}$$

where τ_{α_0} is the solution of the equation $\mathbb{P}_0[\hat{\Lambda}_2(\mathbf{Z}) \geq \tau_{\alpha_0}] = \alpha_0$. The GLR $\hat{\Lambda}_2(z_{k,i}^{\text{wapp}})$ is given by :

$$\hat{\Lambda}_2(z_{k,i}^{\text{wapp}}) = \frac{\hat{\sigma}_{k,0}}{\hat{\sigma}_{k,1}} \cdot \exp\left[\frac{\hat{\sigma}_{k,1}^2 - \hat{\sigma}_{k,0}^2}{2\|\varphi\|_2^2 \hat{\sigma}_{k,1}^2 \hat{\sigma}_{k,0}^2}(z_{k,i}^{\text{wapp}} - \hat{\mu}_k)^2\right] \tag{25}$$

where $\hat{\sigma}_{k,j} = \sqrt{a_j \hat{\mu}_k + b_j}$, $j \in \{0,1\}$.

4.3. Statistical performance of the test

Let consider two functions

$$g(x) = \log\frac{a_0 x + b_0}{a_1 x + b_1} \quad h(x) = \frac{1 - \frac{a_0 x + b_0}{a_1 x + b_1}}{2}$$

which are continuous and differentiable on $\mathbb{R}^+$. Their derivative can be expressed as :

$$g'(x) = \frac{a_0 b_1 - a_1 b_0}{(a_0 x + b_0)(a_1 x + b_1)} \quad h'(x) = \frac{a_1 b_0 - a_0 b_1}{2(a_1 x + b_1)^2}$$

It can be noted that their derivative is equal zero if and only if two cameras 0 and 1 are identical. For obvious reasons, this case is not considered in the present paper.

Taking the logarithm of $\hat{\Lambda}_2(z_{k,i}^{\text{wapp}})$, a straightforward calculation from (25) permits to have :

$$\log\hat{\Lambda}_2(z_{k,i}^{\text{wapp}}) = \frac{1}{2} g(\hat{\mu}_k) + h(\hat{\mu}_k)\frac{(z_{k,i}^{\text{wapp}} - \hat{\mu}_k)^2}{\|\varphi\|_2^2 \hat{\sigma}_{k,0}^2} \tag{26}$$

Under hypothesis $\mathscr{H}_0$, it follows from the Delta method [8, theorem 11.2.14] that :

$$\begin{cases} g(\hat{\mu}_k) \xrightarrow{D} \mathscr{N}\left(g(\mu_k),\ (g'(\mu_k))^2 c_k \sigma_{k,0}^2\right) \\ h(\hat{\mu}_k) \xrightarrow{D} \mathscr{N}\left(h(\mu_k),\ (h'(\mu_k))^2 c_k \sigma_{k,0}^2\right) \end{cases} \tag{27}$$

Moreover, it can be asymptotically shown that :

$$\frac{(z_{k,i}^{\text{wapp}} - \hat{\mu}_k)^2}{\|\varphi\|_2^2 \hat{\sigma}_{k,0}^2} \sim \chi_1^2 \quad \text{under} \quad \mathscr{H}_0 \tag{28}$$

Consequently, the two first moments of $\log\hat{\Lambda}_2(z_{k,i}^{\text{wapp}})$ under $\mathscr{H}_0$ can be computed as :

$$\begin{aligned} \mathbb{E}_0\left[\log\hat{\Lambda}_2(z_{k,i}^{\text{wapp}})\right] &= \frac{1}{2} g(\mu_k) + h(\mu_k) \\ \text{Var}_0\left[\log\hat{\Lambda}_2(z_{k,i}^{\text{wapp}})\right] &= \frac{1}{4}\text{Var}_0[g(\hat{\mu}_k)] + \text{Var}_0[h(\hat{\mu}_k)] \\ &\quad + 2\Big(\text{Var}_0[h(\hat{\mu}_k)] + h^2(\mu_k)\Big) \end{aligned} \tag{29}$$

where $\text{Var}_0[g(\hat{\mu}_k)]$ and $\text{Var}_0[h(\hat{\mu}_k)]$ can be directly derived from (27).

Similarly, a little calculation permits to obtain the two first moments of $\log\hat{\Lambda}_2(z_{k,i}^{\text{wapp}})$ under $\mathscr{H}_1$:

$$\begin{aligned} \mathbb{E}_1\left[\log\hat{\Lambda}_2(z_{k,i}^{\text{wapp}})\right] &= \frac{1}{2} g(\mu_k) + q(\mu_k) \\ \text{Var}_1\left[\log\hat{\Lambda}_2(z_{k,i}^{\text{wapp}})\right] &= \frac{1}{4}\text{Var}_1[g(\hat{\mu}_k)] + \text{Var}_1[q(\hat{\mu}_k)] \\ &\quad + 2\Big(\text{Var}_1[q(\hat{\mu}_k)] + q^2(\mu_k)\Big) \end{aligned} \tag{30}$$

where

$$q(x) = \frac{\frac{a_1 x + b_1}{a_0 x + b_0} - 1}{2}$$

and

$$\begin{cases} g(\hat{\mu}_k) \xrightarrow{D} \mathscr{N}\left(g(\mu_k),\ (g'(\mu_k))^2 c_k \sigma_{k,1}^2\right) \\ q(\hat{\mu}_k) \xrightarrow{D} \mathscr{N}\left(q(\mu_k),\ (q'(\mu_k))^2 c_k \sigma_{k,1}^2\right) \end{cases} \quad \text{under} \quad \mathscr{H}_1 \tag{31}$$

On the other hand, an application of the Lindeberg CLT provides :

$$\log\hat{\Lambda}_2(\mathbf{Z}) \xrightarrow{D} \mathscr{N}\left(m_j^{(2)}, v_j^{(2)}\right) \quad \text{under} \quad \mathscr{H}_j \tag{32}$$

where

$$\begin{cases} m_j^{(2)} &= \sum_{k=1}^{K}\sum_{i=1}^{n_k} \mathbb{E}_j\left[\log\hat{\Lambda}_2(z_{k,i}^{\text{wapp}})\right] \\ v_j^{(2)} &= \sum_{k=1}^{K}\sum_{i=1}^{n_k} \text{Var}_j\left[\log\hat{\Lambda}_2(z_{k,i}^{\text{wapp}})\right] \end{cases} \tag{33}$$

Because μ_k is unknown in the practical context, the expectation and variance of $\log\hat{\Lambda}_2(\mathbf{Z})$ are not defined. Hence, μ_k can be replaced by $\hat{\mu}_k$ in (29) and (30) to finally obtain the estimation of $m_j^{(2)}$ and $v_j^{(2)}$, denoted $\hat{m}_j^{(2)}$, $\hat{v}_j^{(2)}$ respectively.

For the sake of clarity, let us define the normalized log LR

$$\log\hat{\Lambda}_2^{\star}(\mathbf{Z}) = \frac{\log\hat{\Lambda}_2(\mathbf{Z}) - \hat{m}_0^{(2)}}{\sqrt{\hat{v}_0^{(2)}}} \tag{34}$$

Theorem 3. *For any natural image whose parameters are unknown, the decision threshold $\tau_{\alpha_0}^{\star}$ is given by :*

$$\tau_{\alpha_0}^{\star} = \Phi^{-1}(1 - \alpha_0) \tag{35}$$

Theorem 4. *The power of the test δ_2 based on $\log\hat{\Lambda}_2^{\star}(\mathbf{Z})$ is given by :*

$$\beta_{\delta_2} = 1 - \Phi\left(\frac{\hat{m}_0^{(2)} - \hat{m}_1^{(2)} + \tau_{\alpha_0}^{\star}\sqrt{\hat{v}_0^{(2)}}}{\sqrt{\hat{v}_1^{(2)}}}\right) \tag{36}$$

It can be noted that the main advantage of using $\log\hat{\Lambda}_2^{\star}(\mathbf{Z})$ defined in (34) is that the decision threshold given in (35) is independent from image parameters and thus remains the same for nay inspected image.

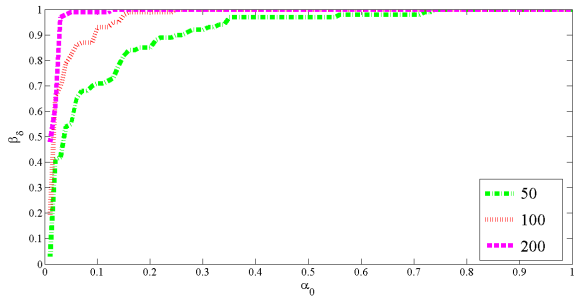

Fig. 2: ROC for different numbers of pixels .

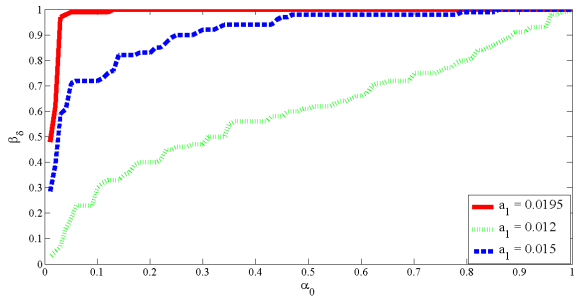

Fig. 3: ROC for different camera parameters .

5. NUMERICAL RESULTS

To emphasis the efficiency of the proposed test, a Monte Carlo simulation was performed on a synthetic image with 1000 repetitions while considering different numbers of pixels and different imaging device parameters. Figure 2 shows the detection performance as a Receiver Operating Characteristic (ROC) respectively for $M = \{50, 100, 200\}$ pixels in which the camera 0 and 1 are characterized by $a_0 = 0.0115$, $b_0 = 2.56\ 10^{-4}$ and $a_1 = 0.0195$, $b_1 = 2.56\ 10^{-4}$. These values correspond to *Nikon D70* and *Nikon D200* respectively. It turns out that the test power significantly drops when the number of pixels decreases.

Similarly, figure 3 shows the detection performance obtained by keeping the camera 0 parameters and setting camera 1 ones to $a_1 = \{0.0195, 0.015, 0.012\}$. A total number of $M = 200$ pixels is considered. As expected, figure 3 shows that when the parameter a_1 tends to a_0, the proposed test power declines and the ROC curves tends to $\beta_\delta = \alpha_0$.

Finally, the experiments were conducted on the Dresden image database [9] captured by the *Nikon D70* and *Nikon D200* camera. Similarly to figure 2, the results presented in figure 4 show the empirical ROC for different numbers of pixels. These results are near to the ones obtained with a synthetic image presented in figure 2 and the small loss of optimality can be explained by the following reasons. First, the clipping or censoring phenomenon [7] have not been taken into account in proposed method and the segmentation method used to obtain the level sets is probably not perfect. These might lead to a slight error in the estimation of variance in each level set. Second, the number of images in Dresden database is rather small, thus the estimation of parameters (a, b) might be slightly biased. However, the test reveals ability of identifying the camera models with a high performance even when the image parameters are unknown and the camera model exhibit similar characteristics. Indeed, for only 500 pixels, the camera model identification is mostly perfect.

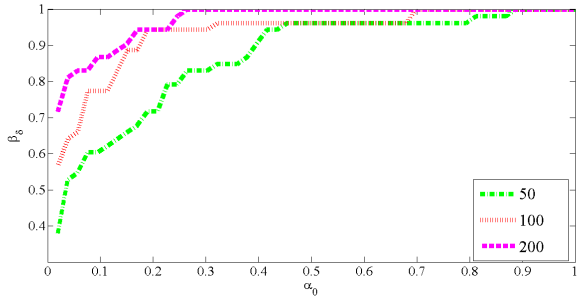

Fig. 4: ROC for different numbers of pixels on Dresden image database .

6. CONCLUSION

The paper provides an approach based on the statistical decision theory for the problem of camera model identification. A statistical noise model is exploited and is characterized by two parameters which can identify the imaging device . The main contribution of the paper is to design an efficient test in presence of nuisance parameters when putting into a practical context. The two first moments of the GLRT are obtained and the statistical performances of the test are analytically computed as well. The paper encourages a deeper research when considering that the imaging device properties are unknown.

7. REFERENCES

[1] V.L Tran et al., "A survey on digital camera image forensics methods," *IEEE ICME*, pp. 16–19, July 2007.

[2] K.S. Choi et al., "Source camera identification using footprints from lens aberration," *Proc. SPIE*, vol. 6069, pp. 155–162, 2006.

[3] J. Lukas et al., "Digital camera identification from sensor pattern noise," *IEEE Trans. Inf. Forensics Security*, vol. 1, no. 2, pp. 205–214, June 2006.

[4] S. Bayram et al., "Source camera identification based on cfa interploation," *IEEE ICIP*, pp. 69–72, September 2005.

[5] M. Kharrazi et al., "Blind source camera identification," *IEEE ICIP*, pp. 24–27, October 2004.

[6] G. E. Healey and R. Kondepudy, "Radiometric ccd camera calibration and noise estimation," *IEEE Trans. Pattern Anal. Mach. Intell*, vol. 16, Mars.

[7] A. Foi et al., "Practical poissonian-gaussian noise modeling and fitting for single-image raw-data," *IEEE Trans. Image Process.*, vol. 17, pp. 1737 –1754, October 2008.

[8] E. L. Lehman and J. P. Romano, *Testing statistical hypotheses*, Springer, 3rd edition, 2005.

[9] T. Gloe and R. Bohme, "The 'dresden image database' for benchmarking digital image forensics," *Proc. ACM SAC*, vol. 2, pp. 1585–1591, 2010.