Camera identification based on sensor noise pattern: a practical procedure for open scenarios

Hermeson Barbosa da Costa, Ronaldo de Freitas Zampolo, Eurípedes Pinheiro dos Santos,
Diego Marques do Carmo and Adalbery Rodrigues Castro

***Abstract*— In this paper, the problem of device identification based on sensor noise pattern in open scenarios is addressed. This context is common in practice and quite challenging because of the lack of reference to evaluate statistical similarity measures between image and suspect camera noise patterns. A device identification procedure based on an artificial neural network classifier is proposed, whose parameters are optimised by extreme learning machine algorithm and repeated double cross validation techniques. In addition, a strategy to select training patterns, aiming at open scenario situations, is presented. Experimental results are shown for the sake of performance assessment.**

***Keywords*— Device identification, sensor noise pattern, extreme learning machine, repeated double cross validation, artificial neural networks.**

I. Introduction

This work addresses the problem of device identification based on the photo-response non-uniformity (PRNU) of camera sensors in open scenarios. In forensics, device identification techniques verify the connection between a signal under analysis and an acquisition equipment. We are particularly interested in digital camera identification. In this context, both digital image/video to be analysed and the suspect camera must be available. In order to find out whether such a camera produced the photo/video in hand, the following approaches can be adopted alone or combined: metadata extraction and evaluation, watermark checking, and sensor noise pattern analysis [1].

The first approach is the simplest and the least robust among the three. It consists in extracting the metadata contained in an image or video file, which is written automatically by the acquisition equipment itself. For camera identification, the information of interest is the model and serial number of the camera. Nevertheless, metadata is considered a weak evidence provider, because such an information generally remains unprotected and can be easily modified or erased by simple file manipulation.

The next approach, analysis of watermark consistency, requires a camera capable to insert a digital watermark in files it produces, containing identification data. In addition, to serve as a reliable feature for forensic purposes, such a watermark should be relatively robust to re-quantisation, resizing, and filtering operations. In this case, the origin of picture/video would be evaluated by comparing the watermark data extracted from the test signal with that of the suspect camera. The drawback here is exactly the necessity of having an acquisition equipment with watermarking capabilities, which is not common if one considers the typical context the most forensic institutes deal with.

The third approach aforementioned is the analysis of sensor noise pattern, also known as PRNU, which is due to differences among photo-sensor elements in converting light to electrical signal [2], [3]. Such differences provide a sort of inherent watermark, which uniquely characterises each photo-sensor and constitutes the so-called *sensor fingerprint*. For such an approach, the identification relies upon the comparison between the estimated camera PRNU and the PRNU obtained from the picture under analysis.

Lately, device identification techniques based on the analysis of sensor noise pattern have gained much attention over other approaches because of the following PRNU properties: a) the noise pattern is unique to each sensor; b) every sensor exhibits it and every picture as well (with the exception of completely dark images); c) the PRNU is relatively robust to a wide range of image processing operations, such as lossy compression, filtering, and gamma adjustment; and d) the PRNU characteristics are stable in time and under a wide range of physical conditions. Nevertheless, PRNU-based identification is still an issue because correlation values between picture and camera PRNUs are quite low, even for images/videos taken from the suspect camera. Such an aspect is mainly due to a typically long acquisition chain, often comprising demosaicing, gamma correction, colour space conversion and lossy compression.

In device identification research papers, in general, techniques are presented and evaluated under the so-called closed scenario, where we have a finite set of cameras and the test picture is known to be taken by one of them. However, in most practical scenarios, just one camera is available (open scenario). In this latter case, interpretation of PRNU statistical similarity values becomes more difficult, due to the absence of reliable reference, which would allow a clear discrimination whether a picture is strongly or weakly linked to a particular camera.

The contributions of this paper are twofold: (a) we present a strategy for training a device identification method to be applied in open scenarios, and (b) we also propose an approach to classify test images, concerning their connection to the suspect camera, based on neural networks and extreme learning machine (ELM) [4].

The authors are with the Computer and Telecommunications Engineering Department, Institute of Technology, Federal University of Pará (FCT/ITEC/UFPA), Belém-PA, Brazil. Emails: hermeson.costa@itec.ufpa.br, zampolo@ufpa.br, epsantos@ufpa.br, diego.carmo@itec.ufpa.br, adalbery@ufpa.br.

The remaining of this text is organised as follows. Section II reviews image sensor output modelling as well as the PRNU estimation process. In Section III, the proposed approach for a practical identification procedure is presented in details. Section IV shows and discusses some experimental results. And in Section V, concluding remarks are drawn.

II. Device identification based on sensor noise pattern

This section presents the sensor output model adopted in this work as well as the strategy to estimate sensor fingerprints. In all expressions, boldface font (e.g., **I**) represents matrices, with $[i, j]$ denoting their element indices. If not stated otherwise, matrix operations are *element-wise*.

A. Sensor output model

In the formation process of a non computer generated digital image, real scene information passes through several stages, as wavelength filtering (usually by using a colour filter array), light-to-electrical current conversion, signal amplitude quantization, colour channel interpolation (demosaicing), and lossy coding (see Fig. 1). Such stages are common to most digital cameras, cellphones and camcorders spread worldwide. Each of them inserts some kind of distortion into the original image signal, which must be taken into account in modelling the image formation process.

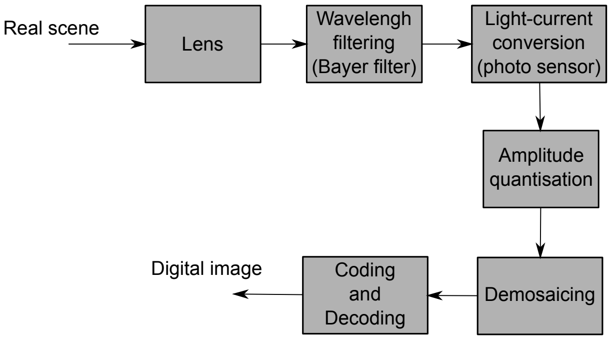

Fig. 1. Imaging chain.

A simplified model for a single-channel image $\mathbf{I}$ is given by [2]

$$\mathbf{I} = g^{\gamma} \left[\mathbf{Y} + \mathbf{YK} + \mathbf{\Omega}\right]^{\gamma} + \mathbf{\Theta} \tag{1}$$

where g represents the colour channel gain (different for each one); γ represents the gamma correction factor; $\mathbf{Y}$ denotes the incident light intensity in the absence of any noise or distortion; $\mathbf{K}$ is a zero-mean multiplicative factor responsible for the device PRNU (camera fingerprint); $\mathbf{\Omega}$ is a combination of several types of noise (dark current, shot noise, etc.); and $\mathbf{\Theta}$ models the combined distortion due to quantisation and/or lossy compression.

B. Camera fingerprint estimation

Based on expression 1, a maximum likelihood (ML) estimator for the PRNU can be derived as [2]

$$\hat{\mathbf{K}} = \frac{\sum_{k=1}^{d} \mathbf{W}_k \mathbf{I}_k}{\sum_{k=1}^{d} (\mathbf{I}_k)^2} \tag{2}$$

where d is the total number of sample images used; $\mathbf{I}_k$ represents the k-th sample image; and $\mathbf{W}_k$ corresponds to the residual noise of $\mathbf{I}_k$, which in turn is given by

$$\mathbf{W}_k = \mathbf{I}_k - \hat{\mathbf{I}}_k^{(0)} \tag{3}$$

where the $\hat{\mathbf{I}}_k^{(0)}$ is a denoised version of $\mathbf{I}_k$ (for details, see [5]).

The larger the total number of sample images (d), the better the PRNU estimation. For acceptable quality estimation, d typically ranges from 30 to 50 [2].

The PRNU ML-estimation ($\hat{\mathbf{K}}$) often contains some undesirable artefacts, which are caused by colour interpolation, lossy compression, on-sensor signal transfer and sensor design choices. Such artefacts are of two types: periodic and non-periodic. The periodic artefacts can be mitigated by subtracting the averages from every row and column of $\hat{\mathbf{K}}$. While the non-periodic artefacts can be alleviated by a Wiener filter applied to the frequency domain [6].

For colour images, the PRNU ML-estimation procedure is repeated for each colour channel and then the corresponding estimations can be combined to obtain an overall camera PRNU.

C. Image fingerprint estimation

The estimation of the PRNU from a single image (i.e. the test image we want to analyse) is performed by taking the corresponding output of a high-pass filter. In this work, we use the technique based on wavelet decomposition that is presented by Mihcak in [2], [5].

III. Proposed approach

This section presents a practical procedure to perform device identification in open scenarios. The proposed approach comprises a classifier based on a neural network trained by an extreme learning machine algorithm (ELM) with model selection accomplished by a repeated double cross validation (RDCV) strategy. From this point on, we consider the image under investigation has the same dimensions of the estimated camera PRNU and does not have undergone any kind of geometrical transformation, such as scaling or rotation.

A. Statistical similarity metric

By noise pattern detection, we mean the procedure that evaluate the connection between suspect camera and test image by assessing the statistical similarity between their estimated PRNUs. This can be represented by a binary hypothesis testing problem as

$$\begin{aligned} H_0 &: \mathbf{K}_1 \neq \mathbf{K}_2 \\ H_1 &: \mathbf{K}_1 = \mathbf{K}_2 \end{aligned} \tag{4}$$

where $\mathbf{K}_1$ and $\mathbf{K}_2$ are the camera and test image PRNUs, respectively.

The null hypothesis H_0 indicates the test image was not taken by the suspect camera (i.e. their fingerprints are different), while the alternative hypothesis H_1 says the opposite.

Two statistical metrics widely employed in PRNU-based techniques are the well-known sample *Pearson's correlation coefficient* (CC) and the so-called *peak-to-correlation energy* (PCE) [7]. In this work, a windowed CC is used, called *step window Pearson's correlation coefficient* (SWCC), whose definition follows.

$$\mathbf{r}[i,j] = \frac{\sum\limits_{(k,l)\in\mathcal{W}} \left(\mathbf{X}[k,l] - \bar{X}_{\mathcal{W}}\right)\left(\mathbf{Y}[k,l] - \bar{Y}_{\mathcal{W}}\right)}{\sqrt{\sum\limits_{(k,l)\in\mathcal{W}} \left(\mathbf{X}[k,l] - \bar{X}_{\mathcal{W}}\right)^2 \sum\limits_{(k,l)\in\mathcal{W}} \left(\mathbf{Y}[k,l] - \bar{Y}_{\mathcal{W}}\right)^2}} \tag{5}$$

where $r[i,j]$ represents the SWCC within a $N \times N$ window $\mathcal{W}$ with centre at $[i,j]$; $\mathbf{X}$ and $\mathbf{Y}$ are two samples, whose statistical similarity is being evaluated; $\bar{X}_{\mathcal{W}}$ and $\bar{Y}_{\mathcal{W}}$ denote the arithmetic mean of $\mathbf{X}$ and $\mathbf{Y}$, respectively, for a given $\mathcal{W}$.

There are some reasons to choose the SWCC: 1) local contributions of image content to PRNU term are better characterised; 2) multiple values of CC for a single image, instead of just one CC value per image, are obtained; 3) the computational complexity is lower when compared with PCE.

In general, independently of the statistical metric chosen, decision for H_0 or H_1 is performed by comparison against a threshold. Rigorously, two thresholds would be needed, one corresponding to the probability of false positive and other to the false negative. However, the determination of those thresholds is an issue, asking for training procedures or the availability of some *a priori* statistical model, which is not easy to obtain [8].

B. Dealing with open scenarios

As stated in Section I, two scenarios are possible for device identification situations. In the *closed scenario*, typical in research papers, there is a finite set of cameras and the test picture is known to be taken by one of those cameras. Thus, hypothesis H_1 is necessarily true for one of the cameras and decision is made in favour of the camera whose PRNU best correlates test image PRNU. In fact, no decision threshold is required at all. In the second context, known as *open scenario*, common in practical situations, just the suspect camera and the test image are available. In this case, a decision threshold is fundamental to choose between H_0 and H_1. Without such a threshold, a forensic technician cannot evaluate correctly the assessed statistical similarity of PRNUs. In addition, the determination of the decision threshold is not an easy task, as it varies from camera to camera and depends highly on a training set composed of true examples of H_0 and H_1 as well. Then, for open scenarios, the parameters of a device identification procedure are specific for a given suspect camera and the performance of such a procedure depends on the training set.

We consider the device identification task as a classification problem, where two classes are assumed: images acquired by the suspect camera (class 1); and images acquired by other cameras (class 0). Class 1 patterns can be obtained from the suspect camera itself, by taking as many pictures as necessary to derive a satisfying camera model. In turn, we propose that class 0 examples might come from public bases on internet. Some of them allow searching by camera make and model, as well as picture resolution, which is very useful in terms of saving time[1]. Class 0 examples in this work are all acquired by cameras of the same make and model of suspect camera, because we assume this is the worst situation for correct device identification.

Next we show a summary of the proposed steps to perform device identification in open scenarios:

1) Estimate suspect camera PRNU (see Section II-B).
2) Obtain class 1 patterns of the training set: by taking N pictures (random scenes) with the suspect camera.
3) Obtain class 0 patterns of the training set: from internet, download N pictures taken by different cameras of the same make and model of the suspect camera.
4) Feature extraction: for each image k of the training set.
   a) Estimate picture PRNU (see Section II-C).
   b) Calculate the SWCC (see Section III-A).
   c) Calculate the mean (μ_k) and variance (σ_k^2) of SWCC.
   d) Store the features (μ_k and σ_k^2).
5) Train the neural network classifier (see Sections III-C and III-D).
6) Test the image under investigation.

C. Extreme learning machine

In this section, we present the extreme learning machine (ELM) algorithm, chosen to train the neural network classifier. The ELM is a learning scheme for single-hidden layer feedforward neural networks (SLFNs) [4], which provides faster learning and good performance.

The ELM algorithm for a given training set $\aleph = \{(\mathbf{x}_k, \mathbf{t}_k)\,|\mathbf{x}_k \in \mathbf{R}^n, \mathbf{t}_k \in \mathbf{R}^m, k = 1, 2, \ldots, N\}$, activation function $g(\cdot)$, and hidden node number $\tilde{N}$ is:

1) Randomly assign input weight $\mathbf{w}_i$ vectors and bias b_i, with $i = 1, 2, \ldots, \tilde{N}$.
2) Calculate the hidden layer output matrix $\mathbf{H}$,

$$\mathbf{H} = \begin{bmatrix} g\left(\mathbf{w}_1 \cdot \mathbf{x}_1 + b_1\right) & \cdots & g\left(\mathbf{w}_{\tilde{N}} \cdot \mathbf{x}_1 + b_{\tilde{N}}\right) \\ \vdots & \ddots & \vdots \\ g\left(\mathbf{w}_1 \cdot \mathbf{x}_N + b_1\right) & \cdots & g\left(\mathbf{w}_{\tilde{N}} \cdot \mathbf{x}_N + b_{\tilde{N}}\right) \end{bmatrix} \tag{6}$$

where, $\mathbf{w}_i \cdot \mathbf{x}_k$ is the inner product between $\mathbf{w}_i$ and $\mathbf{x}_k$.

3) Calculate the output weight β

$$\beta = \mathbf{H}^{-1}\mathbf{T} \tag{7}$$

[1] Class 0 images of this work were downloaded from Flickr web page (http://www.flickr.com)

where $\mathbf{H}^{-1}$ is the inverse of $\mathbf{H}$ and $\mathbf{T} = [\mathbf{t}_1, \ldots, \mathbf{t}_N]^T$.

In order to define the number of hidden nodes which best performs, the repeated double cross validation (RDCV) approach is used.

D. Repeated double cross validation

The repeated double cross validation (RDCV) is a strategy suited for small data sets to estimate the optimum complexity of linear regression models, and the prediction errors for new cases [9]. A pseudo programming code for RDCV follows:

```
FOR $r = 1$ TO $nr$ ($nr$: number of runs)
1) Split data ($D$) randomly into $nf$ folds of approximately equal size.
2) FOR $f = 1$ TO $nf$ ($nf$: number of outer loop folds)
    a) Select the $f^{th}$ fold as the test set, $D_t = D(f)$.
    b) Select the others folds as the calibration set, $D_{cal} = D - D_t$.
    c) Split $D_{cal}$ into $ns$ folds of approximately equal size.
    d) FOR $s = 1$ TO $ns$ ($ns$: number of inner loop folds)
        i) Select the $s_{th}$ fold as the validation set, $D_v = D_{cal}(s)$.
        ii) Select the others folds as the training set, $D_{tr} = D_{cal} - D_v$.
        iii) Train the model $M$ with $D_{tr}$.
        iv) Apply the model to the validation set $D_v$.
        NEXT $s$
    e) Train the model $M$ with all calibration set $D_{cal}$.
    f) Apply the model to the test set $D_t$.
   NEXT $f$
NEXT $r$
```

After nr runs, a set of nr performance values for each pattern in D is available. Such a procedure is repeated for different numbers of hidden nodes in order to select the classifier that best performs.

IV. Experimental procedure and results

The intention behind the conception of the following experiments is to evaluate the performance of the proposed approach in terms of device identification in open scenarios. By now, we are not interested in finding either the best parameter to configure the RDCV or the best architecture for the ELM.

We assume two suspect cameras: a Sony-DSC-W210, serial number 6507300 (Sony_1); and a Fuji-FinePix JZ300, serial number OAQ38568 (Fuji_1). Following the steps described in Section III-B, we take, for each camera, 30 pictures of a cloudy sky to estimate the corresponding camera fingerprints, and other 100 pictures of random content to obtain class 1 patterns. All images are captured with full resolution (4000×3000) and high quality. From Flickr web page, a number of 100 images taken by cameras of the same model as Sony_1 (Sony_Flickr_1) and other 100 of the same model as Fuji_1 (Fuji_Flickr_1) are downloaded to serve as class 0 examples.

For the SWCC, we set the window size to be 128×128, with a step size of 64 pixels, which produces a SWCC set of 2745 values per image.

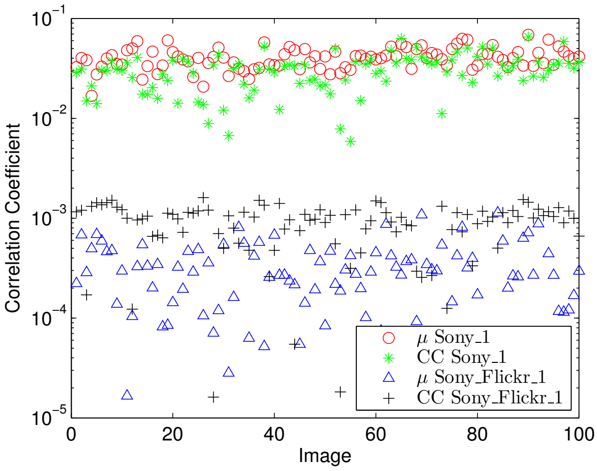

Fig. 2. Comparison between CC and the mean of SWCC for 100 images of each camera (Sony_1 and Sony_Fickr_1) with camera fingerprint estimated from Sony_1.

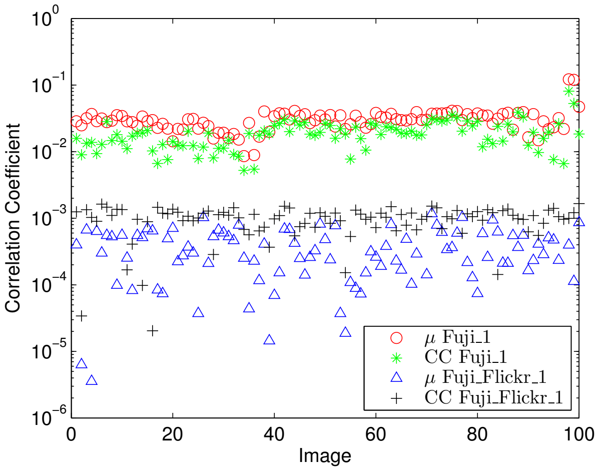

Fig. 3. Comparison between CC and the mean of SWCC for 100 images of each camera (Fuji_1 and Fuji_Fickr_1) with camera fingerprint estimated from Fuji_1.

In Figures 2 and 3, we present a comparison between the mean of SWCC and the traditional CC for suspect cameras Sony_1 and Fuji_1, respectively. For both suspect cameras, SWCC exhibits some evidence of superior class discrimination, when compared with CC. The authors speculate that windowing characterise better the influence of image content over PRNU estimation.

Figures 4 and 5 show the obtained success rates for the device identification method based on a SLFN classifier, which is optimised by the ELM algorithm in conjunction with the RDCV strategy, as a function of the number of hidden nodes. The experimental procedure considers as activation function $g(\cdot)$ the sigmoid and normalised outputs into $[-1, 1]$ interval. The RDCV is set to run 100 times, with 5 folds in the outer loop and 10 folds in the inner loop (see Section III-D). The number of hidden nodes ranges from 2 to 15.

For the results in Figure 6, we use pictures from Fuji_1 and Fuji_Flickr_1 to train a SLFN classifier with 6 hidden nodes. A set of 200 patterns (100 from class 0 and 100 from

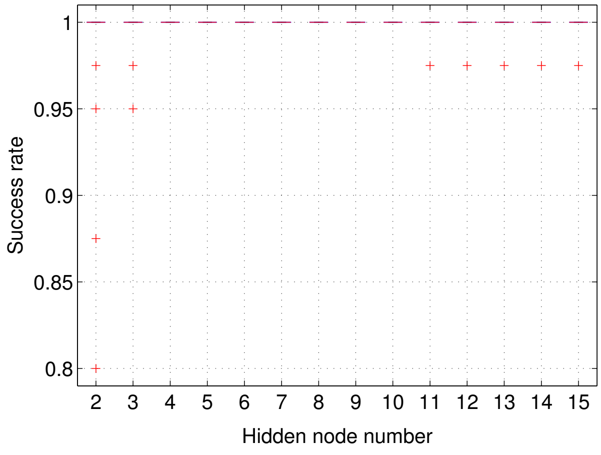

Fig. 4. Device identification success rate as a function of the number of hidden nodes in a SLFN optimised by ELM and RDCV (100 executions, 5 folds in the outer loop and 10 folds in the inner loop), with camera fingerprint estimated from Sony_1. In the box plots, red line and red cross denote median and outliers, respectively.

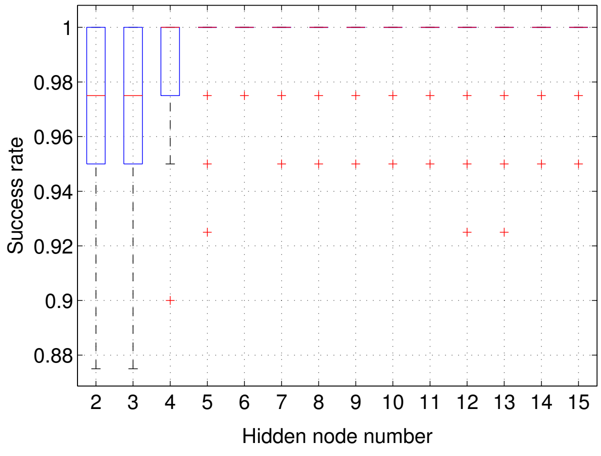

Fig. 5. Device identification success rate as a function of the number of hidden nodes in a SLFN optimised by ELM and RDCV (100 executions, 5 folds in the outer loop and 10 folds in the inner loop), with camera fingerprint estimated from Fuji_1. In the box plots, red line, red cross, black line and blue box denote median, outliers, minimum value and lower-to-upper quartile, respectively.

class 1) is randomly split into train (80%) and validation (20%) subsets. For testing, we use an ensemble of 350 images made up of: 100 pictures taken by two cameras (50 pictures each) of the same model of Fuji_1 (serial numbers OAQ38657 and OBQ39278), identified as Fuji_2 and Fuji_3; 50 images acquired with a camera of the same model of Sony_1 (serial number 6507323), named Sony_2; 100 images taken by Sony_1; and 100 pictures downloaded from Flickr web page, identified as Sony_Flickr_1. Figure 6 shows all test images as well as the decision boundary of the classifier in feature plane. Although, training phase is performed only with Fuji_1 and Fuji_Flickr_1 images, the classifier is able to cope with images obtained by cameras of different makes and models.

V. Conclusions

In this paper, the problem of device identification is addressed. Specifically, a procedure based on a SLFN classifier, optimised by ELM algorithm in conjunction with RDCV, is proposed as well as a strategy to select the training patterns in order to be applied to open scenario situations. Experimental results demonstrates the proposed method performs well. Further investigation should consider device identification for pictures subjected to geometric transformations and video signals as well.

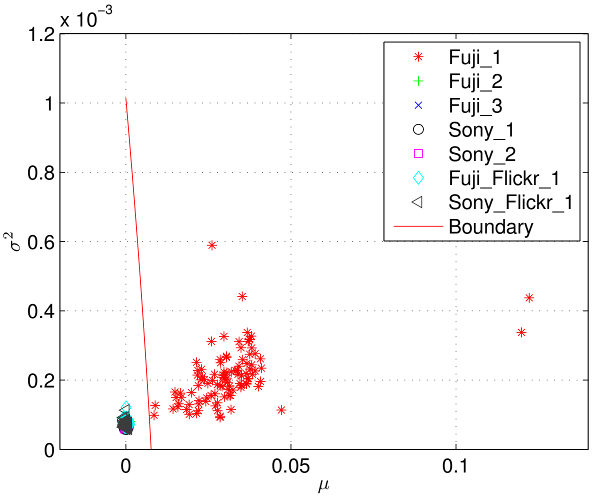

Fig. 6. Test images and decision boundary in the feature plane, with camera fingerprint estimated from Fuji_1.

Acknowledgements

The authors would like to thank PROEX/UFPA and the Scientific Forensic Centre Renato Chaves (Centro de Perícias Científicas Renato Chaves, CPCRC) for supporting this work.

References

[1] Judith A. Redi, Wiem Taktak, and Jean-Luc Dugelay, "Digital image forensics: a booklet for beginners," *Multimedia Tools Appl.*, vol. 51, no. 1, pp. 133–162, jan 2011.

[2] J. Fridrich, "Digital image forensics," *IEEE Signal Processing Magazine*, vol. 26, no. 2, pp. 16–37, March 2009.

[3] M. Goljan and J. Fridrich, "Camera identification from cropped and scaled images," in *Proceedings of the SPIE Eletronic Imaging, Forensics, Security, Steganography, and Watermarking of Multimedia Contents X*, 2008, pp. 28–30.

[4] Guang-Bin Huang, Qin-Yu Zhu, and Chee-Kheong Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, no. 1â3, pp. 489 – 501, 2006, ¡ce:title¿Neural Networks¡/ce:title¿ ¡ce:subtitle¿Selected Papers from the 7th Brazilian Symposium on Neural Networks (SBRN '04)¡/ce:subtitle¿ ¡xocs:full-name¿7th Brazilian Symposium on Neural Networks¡/xocs:full-name¿.

[5] M. K. Mihcak, I. Kozintsev, and K. Ramchandran, "Spatially adaptive statistical modeling of wavelet image coefficients and its application to denoising," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Phoenix, AZ, March 1999, vol. 6, pp. 3253–3256.

[6] Jessica Fridrich, *Digital Image Forensics: There is More to a Picture than Meets the Eye*, chapter Sensor Defects in Digital Image Forensics, Springer, May 2012.

[7] B. V. K. Vijaya Kumar and L. Hassebrook, "Performance measures for correlation filters," *Appl. Opt.*, vol. 29, no. 20, pp. 2997–3006, Jul 1990.

[8] H. B. Costa, R. F. Zampolo, D. M. Carmo, A. R. Castro, and E. P. Santos, "On the practical aspects of applying the prnu approach to device identification tasks," in *International Conference on Multimedia Forensics, Surveillance and Security*, Brasília (DF) - Brazil, September 2012.

[9] Peter Filzmoser, Bettina Liebmann, and Kurt Varmuza, "Repeated double cross validation," *Journal of Chemometrics*, vol. 23, no. 4, pp. 160–171, 2009.