SOURCE CAMERA MODEL IDENTIFICATION BASED MACHINE LEARNING APPROACH

PhD student

AMEL TUAMA

Supervised by

MARC CHAUMONT **FRÉDÉRIC COMBY**

CONTENTS

- Digital Image Forensics.
- Methods of Camera Identification.
- Proposed Method.
- Experiments and Results.
- Conclusions and Perspectives.

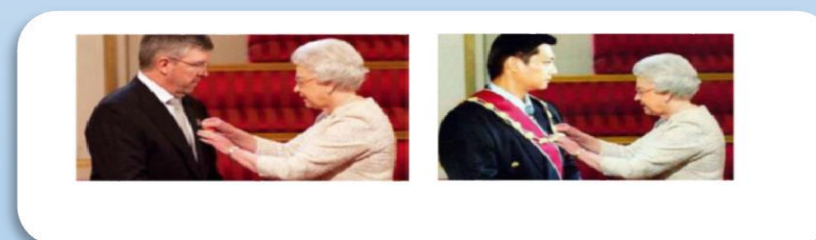
DIGITAL IMAGE FORENSICS

Digital Image Forensics

Camera Identification

Forgery Detection

Image of which camera?

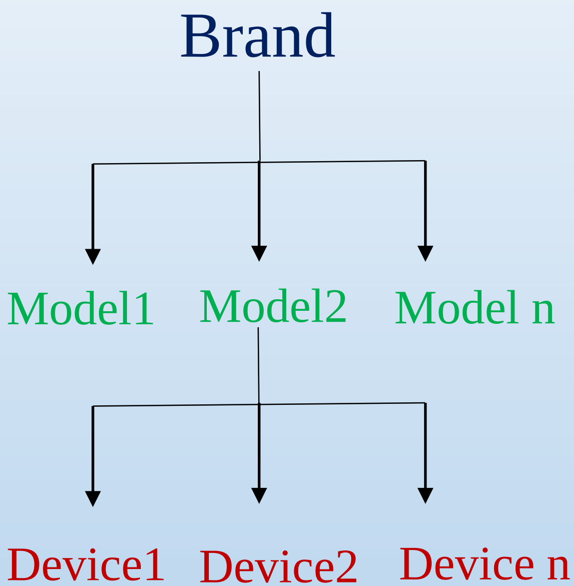


a-Original image b-Tampered image

Why Source Identification?

- In tracing the history of an image, identifying the device used for its acquisition.
- In a court of law, the origin of a particular image can represent crucial evidence.

Brand, Model, Device



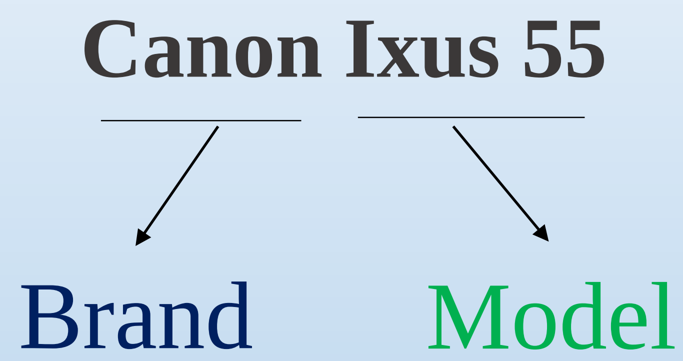
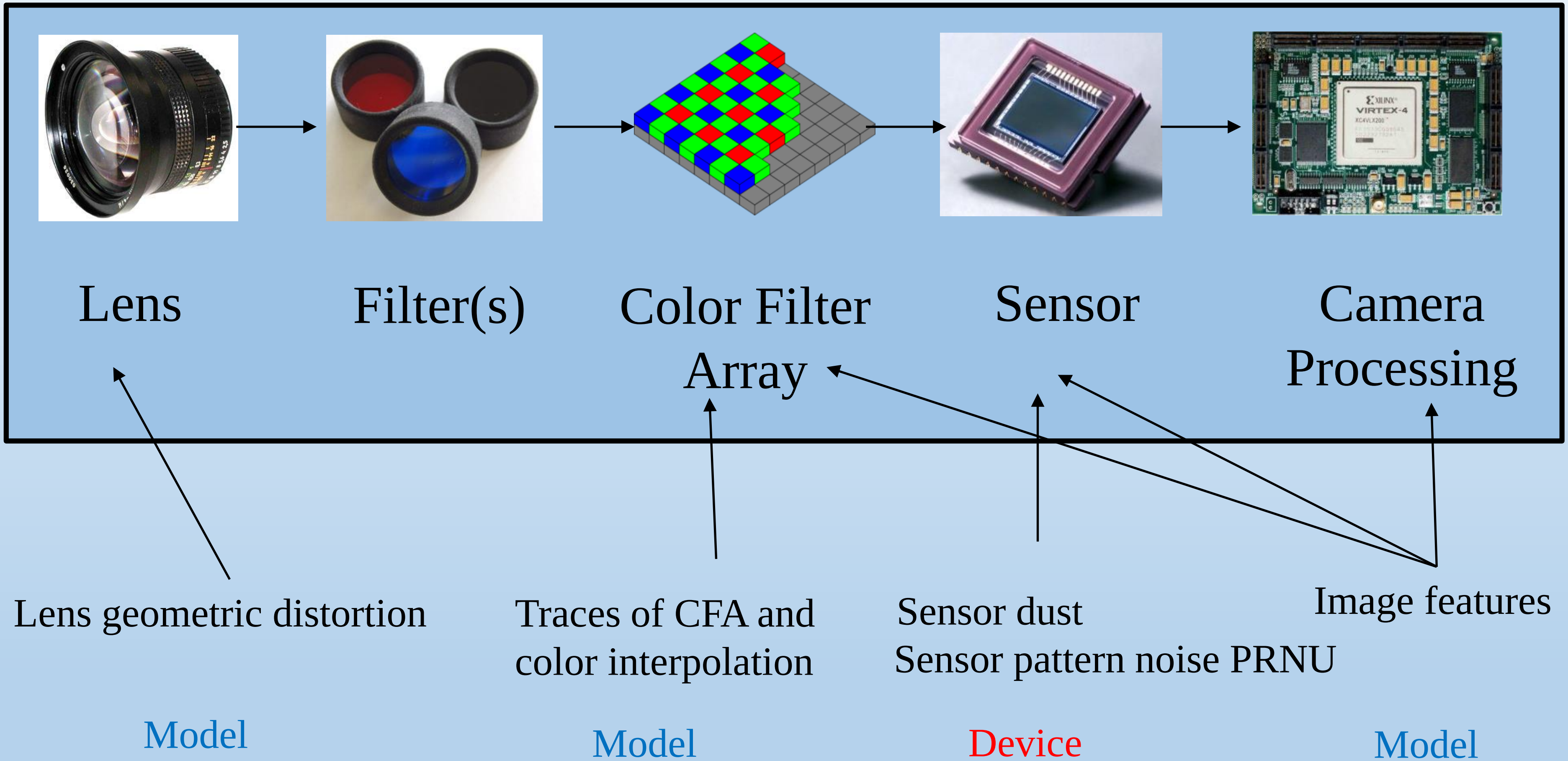


Image Acquisition Pipeline



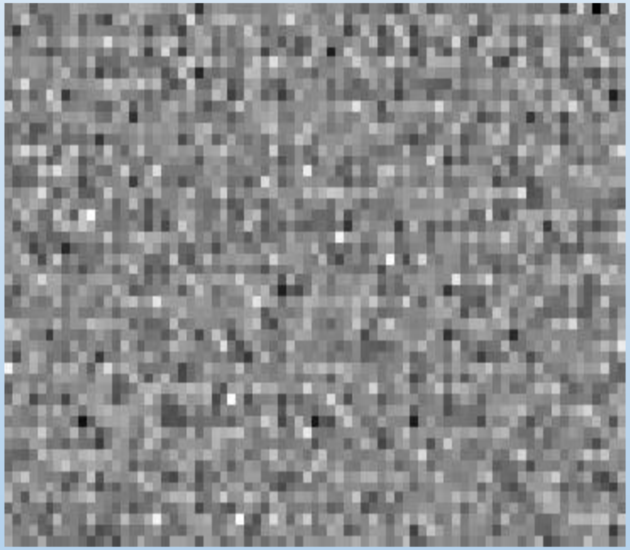
Methods Based Correlations

- **PRNU** Photo Response Non Uniformity is a major source of pattern noise.
- It is a reliable method for identifying individual source camera device because it is unique for each sensor [3].

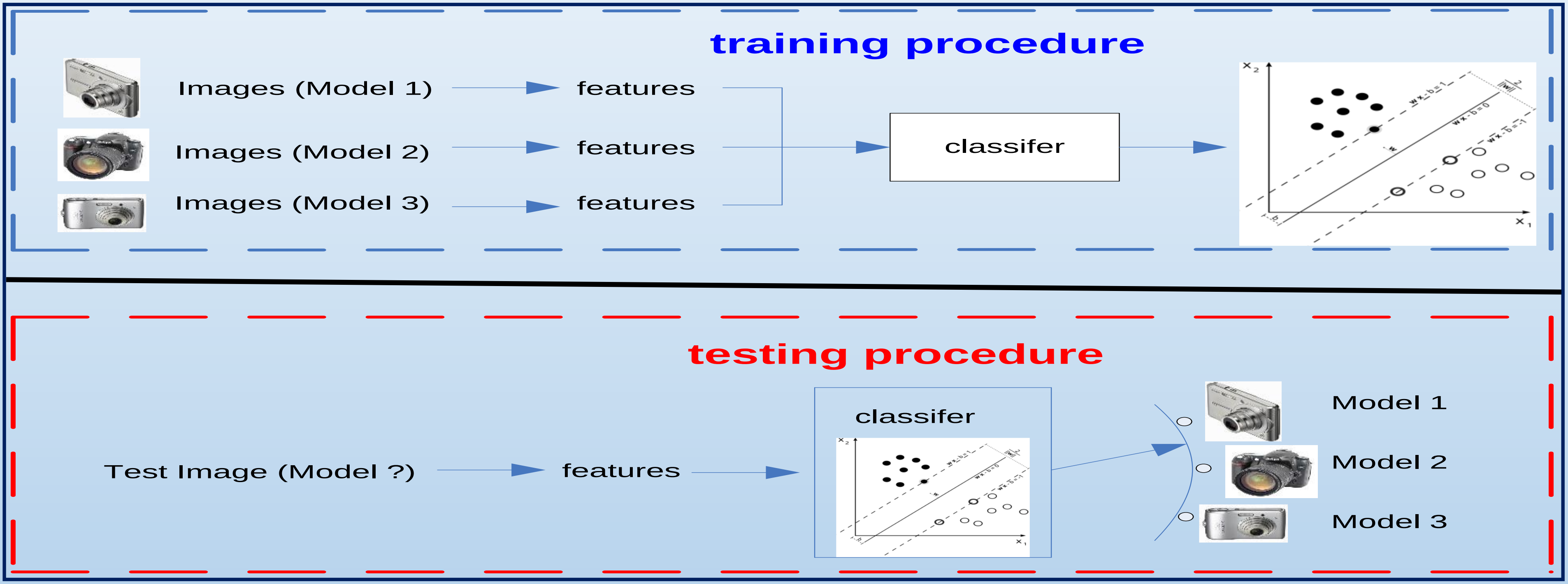
$$N = I - F(I)$$

$$K_c = \frac{\sum NI}{\sum I^2}$$

$$\rho(N, K_c) = \frac{(N-\overline{N}).(K_c-\overline{K_c})}{\|N-\overline{N}\|.\|K_c-\overline{K_c}\|}$$



Proposed Method



Proposed Method

- Principle:

  - Extract a polluted sensor noise POL-PRNU.

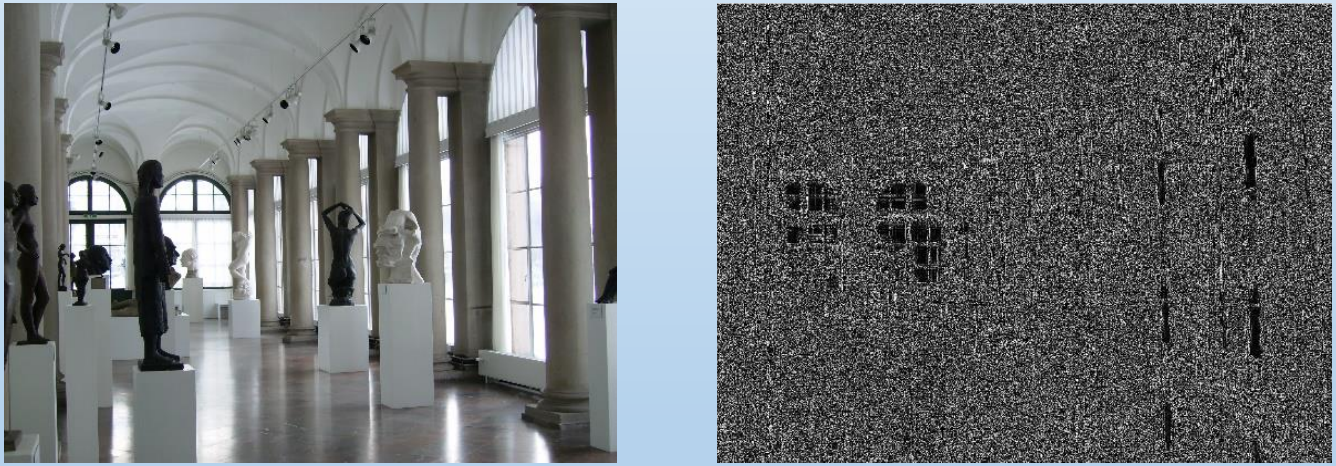
  - Extract Two sets of features from POL-PRNU.

POL-PRNU Extraction

- POL-PRNU polluted sensor noise is the sensor noise contaminated with other types of noise like image contents[3]:

$$N = I - F(I),$$

where ***I*** is the image**,** ***F(I)*** is a wavelet based denoising filter.



Fig(1) Sample image with its residual noise

Feature set (1)
High order statistics from POL-PRNU

- High order statistic features among neighboring pixels from the POL-PRNU[4] .

$$\mathbf{R} \longleftarrow \boldsymbol{trunc_T}\ \mathbf{(round(L/q)),}$$

where $trunc_T$ minimizes the residual range with T϶{-T,…,T}, round(x) gives nearest integer value of x, L is the linear pattern of POL-PRNU, q϶{1,1.5,2}.

- The horizontal and vertical co-occurrences matrix ***C*** is given by:

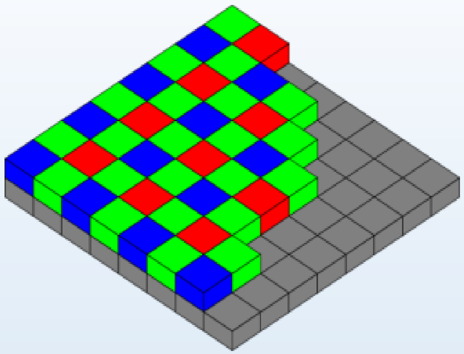
$$\boldsymbol{C_d^h = \frac{1}{Z} \left| \{ (i,j) \mid \quad R_{i,j} = d1 ,\ R_{i,j+1} = d2 ,\ R_{i,j+2} = d3 ,\ R_{i,j+3} = d4 \} \right| ,}$$

where ***Z*** is the normalization factor, $\boldsymbol{R_{i,j} \epsilon \mathbb{N}}$, ***d = (d1,…,d4)*** $\in \{-\boldsymbol{T}, \dots, \boldsymbol{T}\}^{\mathbf{4}}$ with ***T=2.***

Equivalently we can compute the vertical co-occurrences matrix.

- This step gives 10764 Features

Feature set (2)
Traces of color dependencies in CFA interpolation



- The CFA & interpolation introduces specific correlations between the samples of a color image. Bayer array is the most frequently used.
- Compute the normalized cross correlation between the combination of the POL-PRNU of the color channels {C1,C2}∈{RR,GG,BB,RG,RB,GB} and their shifts $\Delta 1 \in \{0, \dots, 3\}$, $\Delta 2 \in \{0, \dots, 3\}$[5]:
- This step gives 96 features.

$$\rho(C1,C2,\Delta)=\frac{\sum_{i,j}(C1_{i,j}-\overline{C1})(C2_{i-\Delta 1,j-\Delta 2}-\overline{C2})}{\sqrt{\sum_{i,j}\left(C1_{i,j}-\overline{C1}\right)^2 \sum_{i,j}\left(C2_{i-\Delta 1,j-\Delta 2}-\overline{C2}\right)^2}}.$$

Experiments & Results

- We used 14 camera model from Dresden database[6].
- We used 1400 images for training, 100 images for each camera model.
- We used 1400 images for testing, 100 images for each camera model.
- We have (10764 + 96 = 10860) features.
- SVM classifier with radial basis function (RBF): $K(x_i, x_j) = e^{-\gamma|x_i - x_j|^2}$, γ>0.
- Apply grid search and cross-validation to choose the best parameters c and γ.
- Use best parameters c and γ to train and test SVM [7].

Experiments & Results

Total Classification Accuracy **97.81 %**

CAMERA MODEL	Agfa Photo DC-733s	Agfa Photo DC-830i	Agfa Photo Sensor 530s	Canon Ixus 55	Fujifilm FinePix J50	Kodak M1063	Nikon D200 Lens A/B	Olympus M1050SW	Panasonsic DMC-FZ50	Praktica DCZ 5.9	Samsung L74wide	Samsung NV15	Sony DSC-H50	Sony DSC-W170
CLASSIFICATION ACCURACY (%)	**96.93**	**97.92**	**98.78**	**99.57**	**98.35**	**98.21**	**99.07**	**98.92**	**99**	**97.97**	**99.91**	**97.57**	**93**	**93.94**

Table 1. Identification accuracies for the fourteen camera models

Method	**Total Accuracy**
CFA	**86.93**
CO-OCCURRENCES	**96.91**
FILLER et al. [5]	**88.23**
PROPOSED METHOD (CFA+CO-OCCUR)	**97.81**

Table 2. Comparison of the proposed method with other approaches.

Conclusions

- Source Camera Identification is the process of deciding which camera has been used to capture a particular image.
- Our method identifies camera model based machine learning approach.
- The results illustrate the efficiency of the proposed method since it provides 97.81% of accuracy compared to Filler's method 88.23%.

Perspectives and Future plan

- Enhance the proposed method by extracting more features increase the accuracy.
- Try a bigger database with more camera models.
- Test the data set with deep learning approach and CNN.

References

[1] K. S. Choi, E. Y. Lam, and K. K. Y. Wong, "Source camera identification using footprints from lens aberration*"*, *Proceedings of the SPIE 2006*.

[2] A. Emir Dirik , Husrev T. Sencar, Nasir Memon," Source camera identification based on sensor dust characteristics.", 2006

[3] J. Lukas, J. Fridrich, and M. Goljan, "Digital Camera Identification from Sensor Pattern Noise", IEEE Transactions on Information Forensics and Security 2006.

[4] J. Fridrich and J. Kodovsky, Rich models for steganalysis of digital images, IEEE Transactions on Information Forensics and Security, vol. 7, no. 3, pp. 868882, 2012.

[5] Filler T., Fridrich J., Goljan M.,"Using sensor pattern noise for camera model identification", 2008.

[6] Thomas Gloe, Rainer Böhme, 'The Dresden Image Database for Benchmarking Digital Image Forensics' SAC'10 March 22-26, 2010, Sierre, Switzerland.

[7] Chih-Chung Chang and Chih Jen Lin http://www.csie.ntu.edu.tw/~cjlin/libsvm

THANK YOU

Questions?