Universidade Vigo GPSC atlanTTic research center for Telecommunication Technologies

MMFORWILD 2020

MultiMedia FORensics in the WILD (MMForWILD) 2020

A Walk on the Wild Side of Camera Attribution

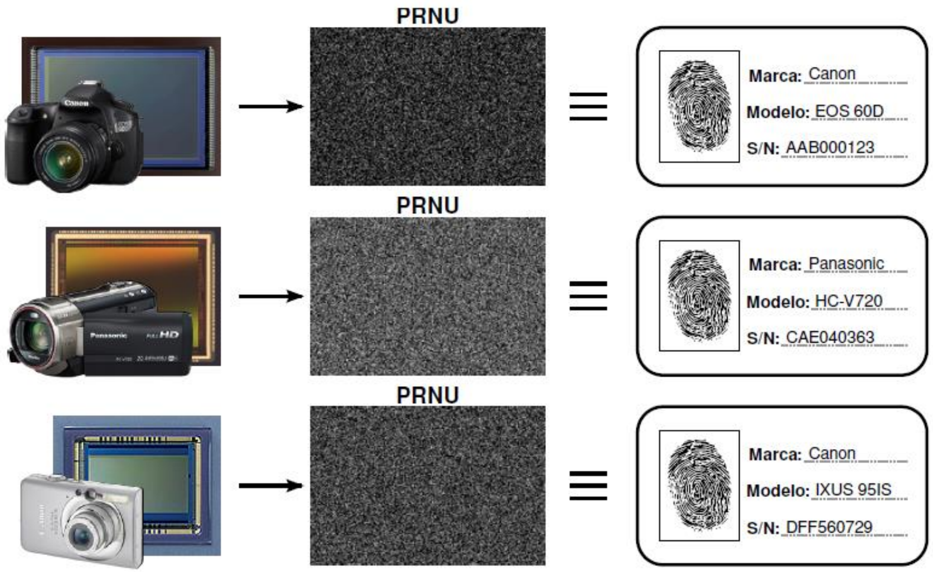
Fernando Pérez-González

AtlanTTic Research Center

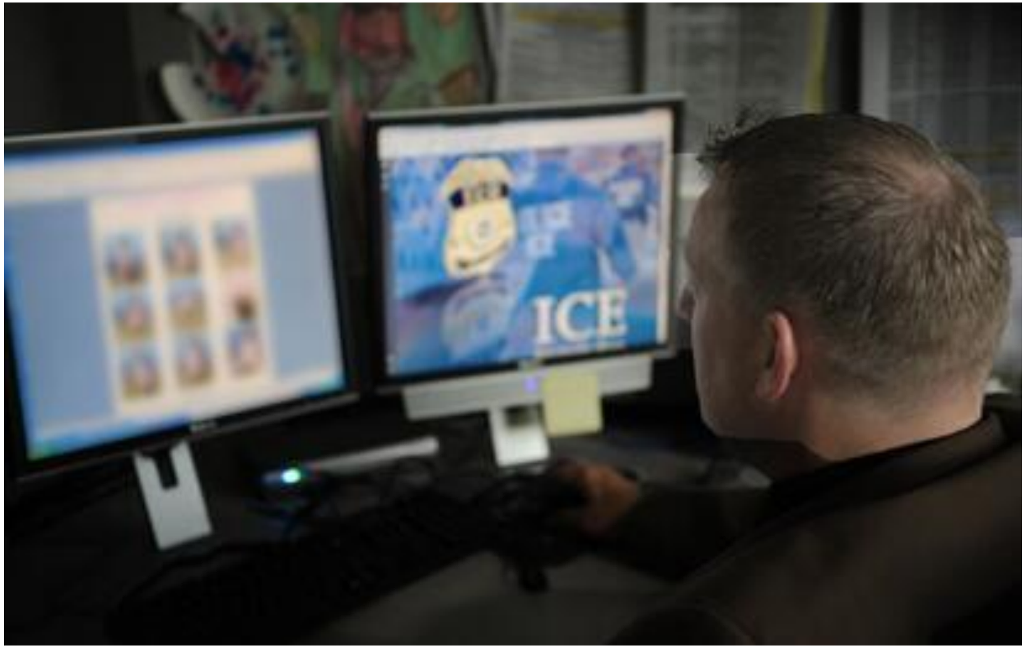
University of Vigo - Spain

Camera attribution with the PRNU

- Practically all (CMOS, CCD, etc.) have an intrinsic noise pattern: PRNU (Photo Response Non Uniformity)
  - **PRNU properties:** robustness, stability, universality
  - Can be used for forensic camera attribution due to its uniqueness.



Forensic uses (e.g. fight against child abuse)



Social Network Analysis for Law Enforcement

Digital onboarding

Biometric proof-of-life

Insurance damage reporting


atlanTTic

The two fundamental hypotheses in camera attribution

(1) The PRNU is a sort of mutliplicative noise:

PRNU   noise

Output pixel $\rightarrow$ $$y(i,j) \approx x(i,j) + x(i,j)k(i,j) + n(i,j)$$

Pristine image

or

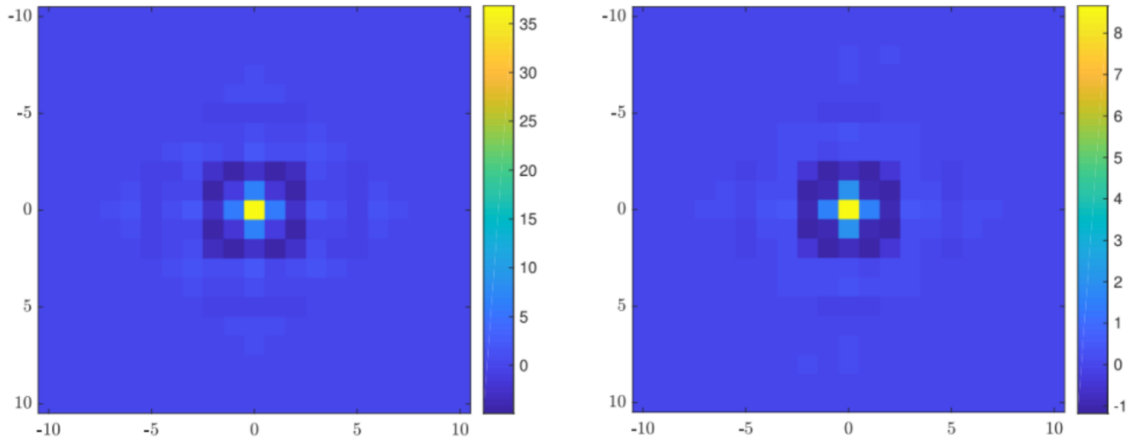
$$\mathbf{Y} \approx (\mathbf{1} + \mathbf{K}) \circ \mathbf{X} + \mathbf{N},$$

(2) The PRNU is unique, i.e., for any two devices with PRNUs $\mathbf{K}_1, \mathbf{K}_2$

$$\langle \mathbf{K}_1, \mathbf{K}_2 \rangle_F \ll ||\mathbf{K}_1||_F; \quad \langle \mathbf{K}_1, \mathbf{K}_2 \rangle_F \ll ||\mathbf{K}_2||_F$$

Additional hypothesis

3 The PRNU is zero-mean, Gaussian and nearly-white, i.e., for lags outside a small neighborhood of the origin $\mathcal{L}$ (s.t. $|\mathcal{L}| \ll N_1 \times N_2$) the autocorrelation is almost zero.



(a) $R_{\hat{k}}$, $L = 50$. (b) $R_{\hat{k}}$, $L = 1000$.

Estimated autocorrelation of the PRNU for a Nikon D60 camera

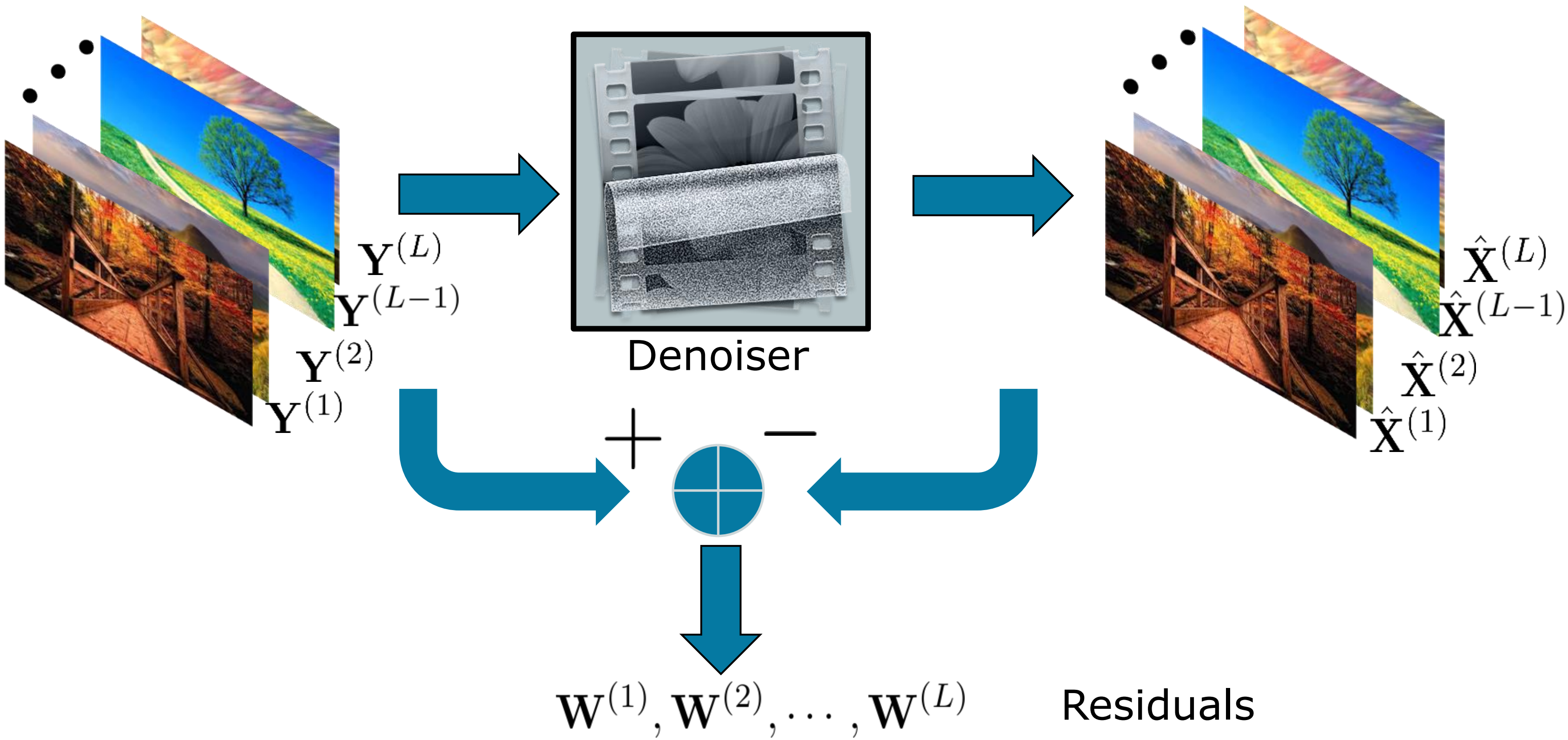
Camera attribution workflow

But sir, I heard something about residuals. Is that chemical residuals?

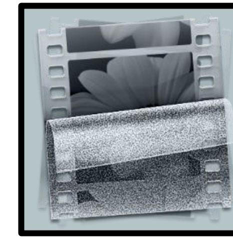
Extract residuals → Estimate PRNU

Extract residual → Measure PCE → Make decison

Residual computation [Lukás06]



Many options for the denoising



- [Mihcak99] wavelet-based (4-level 8-tap Daubechies QMF).
- [Kang 14] 8-neighbour context-adaptive interpolation (CAI).
- [Al-Ani15] similar-pixel opposite-sign PRNU in a neighborhood.
- [He13] content-adaptive guided filtering (CAGI).
- [Perona90] anisotropic diffusion.
- [Rudin94] total variation filtering.
- [Dabov 07] block-matching and 3D filtering (BM3D).
- [Alparone06] MMSE for multiplicative noise in the wavelet domain.
- …

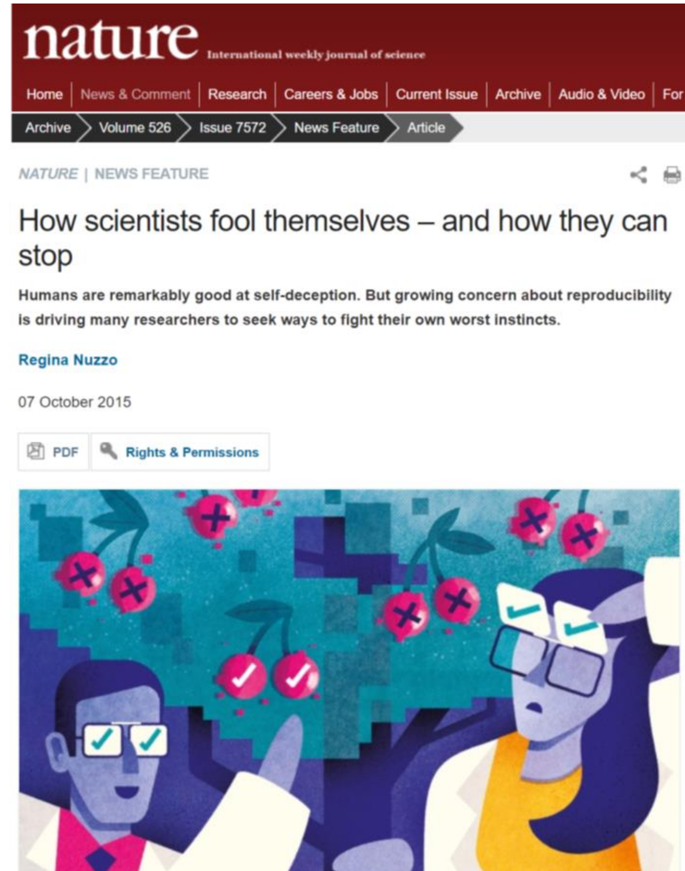
And several comparisons

- [Amerini09], [Cortiana11], [Al-Ani17]…
- **Main conclusion**: BM3D perfoms best but is computationally very expensive; Mihcak's is the most popular, but CAGI is worth exploring further.

	TPR @ 1e-03	EER	CPU time (ms)
BM3D	83.9%	4.9%	4273
Mihcak	70.9%	7.3%	1105
CAGI	70.5%	9.5%	138
TV	58.9%	8.0%	22
Similar pixel	51.2%	13.7%	920
CAI	24.3%	14.8%	4074

A note on asymmetric attention bias

- "Asymmetric attention to detail. Sometimes known as disconfirmation bias, this happens when we give expected results a relatively free pass, but we rigorously check non-intuitive results."


nature International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video | For Au

Archive > Volume 526 > Issue 7572 > News Feature > Article

NATURE | NEWS FEATURE

How scientists fool themselves – and how they can stop

Humans are remarkably good at self-deception. But growing concern about reproducibility is driving many researchers to seek ways to fight their own worst instincts.

Regina Nuzzo

07 October 2015

PDF | Rights & Permissions

PRNU estimation [Chen08]

- After the denoising the standard model goes like this:

$$\mathbf{W}^{(i)} = \mathbf{K} \circ \mathbf{X}^{(i)} + \mathbf{N}^{(i)}$$

- And if the noise is i.i.d. Gaussian, uncorrelated with $\mathbf{X}^{(i)}$ and $\mathbf{K}$ the Maximum Likelihood Estimator is

$$\hat{\mathbf{K}} = \frac{\sum_{i=1}^{L} \mathbf{W}^{(i)} \circ \hat{\mathbf{X}}^{(i)}}{\sum_{i=1}^{L} \hat{\mathbf{X}}^{(i)} \circ \hat{\mathbf{X}}^{(i)}}$$

Sample-wise division

- For residuals with different noise variances $\{\sigma_i^2\}_{i=1}^{L}$

$$\hat{\mathbf{K}} = \frac{\sum_{i=1}^{L} \mathbf{W}^{(i)} \circ \hat{\mathbf{X}}^{(i)} / \sigma_i^2}{\sum_{i=1}^{L} \hat{\mathbf{X}}^{(i)} \circ \hat{\mathbf{X}}^{(i)} / \sigma_i^2}$$

PRNU estimation

- Why then the simple averaging of residuals [Lukás06]

$$\hat{\mathbf{K}} = \frac{1}{L}\sum_{i=1}^{L}\mathbf{W}^{(i)}$$

is almost as good an estimate**?**

- Maybe we should revise our model for the “wild” case:

$$\mathbf{W}^{(i)} = \mathbf{K} \circ \mathbf{X}^{(i)} + \alpha(\mathbf{X}^{(i)} - \mathbb{E}\{\mathbf{X}^{(i)}\}) + \mathbf{N}^{(i)}$$

- Notice that multiplying by $\mathbf{X}^{(i)}$ also increases the ‘noise’ part.

PRNU detection [Goljan08], [Kang12]

- By far, the most popular detector is based on the PCE. Formally, given the test-image residual $\mathbf{W}_t$ and the estimated fingerprint $\hat{\mathbf{X}}_t \circ \hat{\mathbf{K}}$ it first computes the NCC

PCE: Peak to Correlation Energy

$$\rho(i,j) \doteq \frac{\langle \Delta_{i,j}(\mathbf{W}_t) - \mathbb{E}\{\mathbf{W}_t\}, \hat{\mathbf{X}}_t \circ \hat{\mathbf{K}} - \mathbb{E}\{\hat{\mathbf{X}}_t \circ \hat{\mathbf{K}}\}\rangle_F}{||\mathbf{W}_t - \mathbb{E}\{\mathbf{W}_t\}||_F \cdot ||\hat{\mathbf{X}}_t \circ \hat{\mathbf{K}} - \mathbb{E}\{\hat{\mathbf{X}}_t \circ \hat{\mathbf{K}}\}||_F}$$

with $\Delta_{i,j}$ the operator cyclic shift by (i,j).

So why the NCC alone works so well?

- Then, the Signed PCE (SPCE) is

$$\mathrm{SPCE}(\mathbf{W}_t, \hat{\mathbf{X}}_t \circ \hat{\mathbf{K}}) = \frac{\rho(0,0)}{\left(\frac{1}{N_1 \times N_2 - |\mathcal{L}|} \sum_{(i,j)\notin\mathcal{L}} \rho^2(i,j)\right)^{1/2}}$$

Simplifications

- The denominator of the SPCE estimates the std under H_0. But since $\Delta_{i,j}(\mathbf{W}_t) - \mathbb{E}\{\mathbf{W}_t\}$ and $\hat{\mathbf{X}}_t \circ \hat{\mathbf{K}} - \mathbb{E}\{\hat{\mathbf{X}}_t \circ \hat{\mathbf{K}}\}$ are uncorrelated for $(i,j) \notin \mathcal{L}$, we can approximate

$$\left(\frac{1}{N_1 \times N_2 - |\mathcal{L}|} \sum_{(i,j)\notin\mathcal{L}} \rho^2(i,j) \right)^{1/2} \approx 1$$

If $N_1 \times N_2 - |\mathcal{L}| \gg 1$

- Thus

$$\mathrm{SPCE} \approx \rho(0,0)$$

The importance of signal contamination

- Assume zero-mean residual and PRNU. In detection we must compute $\langle \mathbf{W}_t, \hat{\mathbf{X}}_t \circ \hat{\mathbf{K}} \rangle_F$. Remembering the new model proposed for the wild case:

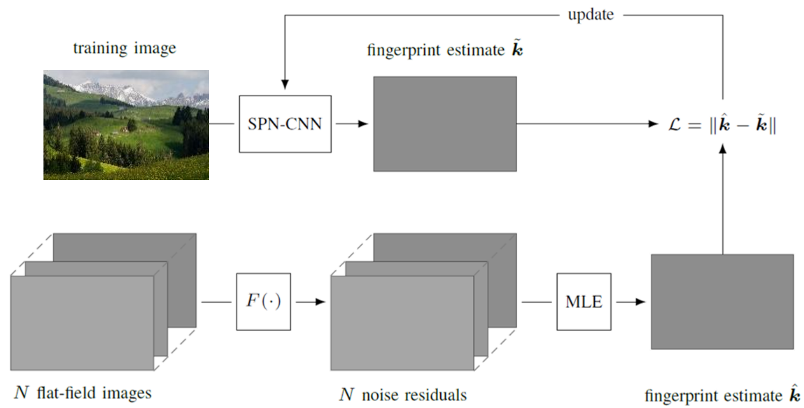
$$\mathbf{W}_t = \mathbf{K} \circ \mathbf{X}_t + \alpha(\mathbf{X}_t - \mathbb{E}\{\mathbf{X}_t\}) + \mathbf{N}_t$$

- The variance of the pure noise terms $\langle \mathbf{N}_t, \mathbf{X}_t \circ \hat{\mathbf{K}} \rangle_F$ depends on $\mathbb{E}\{X_t^2\}$
- But the variance of the leakage terms $\langle \alpha(\mathbf{X}_t - \mathbb{E}\{\mathbf{X}_t\}), \mathbf{X}_t \circ \hat{\mathbf{K}} \rangle_F$ depends on $\mathbb{E}\{X_t^4\} + \mathbb{E}^2\{X_t\}\mathbb{E}\{X_t^2\}$

Under some mild symmetry conditions

A takeaway

- An optimal denoiser (e.g., in MMSE sense) is not necessarily optimal for PRNU detection! (correlation of residual with $\mathbf{X}_t$ also counts)
- May explain why state-of-the-art DNN denoisers give no apparent advantage w.r.t. BM3D in this scenario [Kirchner19].
- And may explain the excellent performance of the SP-CNN denoiser in [Kirchner19] (albeit not suitable for wild scenarios):




Gee, whiz! Perhaps you should also check your FUNDAMENTAL hypotheses, sir!

1. The multiplicative dependence [Chen08]

$$X_0(i,j)$$

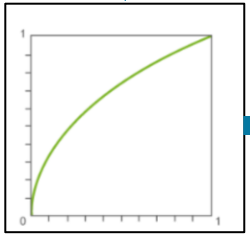
↓

Sensor

↓

$$Y_0(i,j) = G \cdot (1 + k(i,j))X_0(i,j) + N_0(i,j)$$

↓



Gamma correction

→

k is very small

First order Taylor expansion

→

Extra noise sources

↓

$$Y(i,j) = (1 + k(i,j))X(i,j) + N(i,j)$$

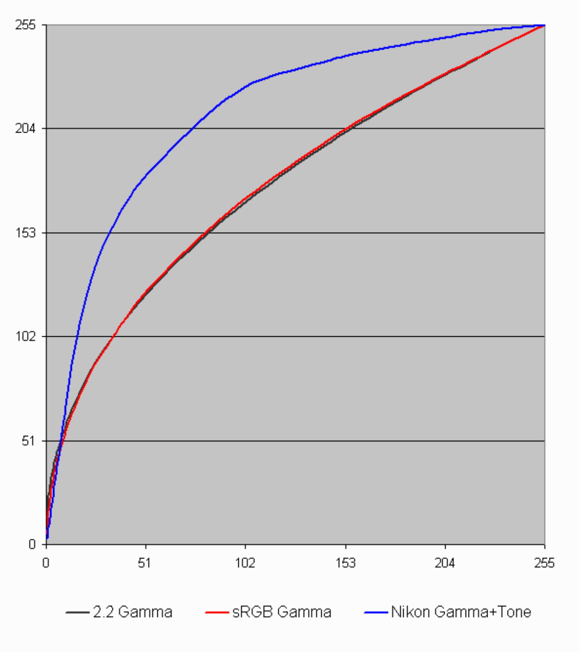
The gamma-response lemma [Pérez-González21]

- Let $y = h(x)$ be the (monotonic) camera response function. If the input is of the form $(1+k)x$ with $k \ll 1$, then the output is of the form $y(1+ck)$ for some constant c if and only if $h(x) = c_1 x^{\gamma}$, with c_1, γ constants.
- In other words: (1+PRNU) is multiplicative if and only if the camera response function is a pure gamma correction.
- Therefore, in general there is a function $g(\cdot)$ such that

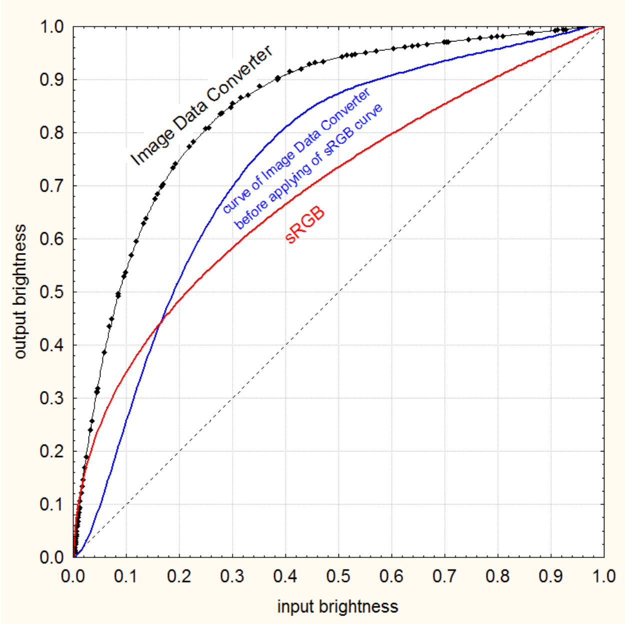
$$Y(i,j) = X(i,j) + k(i,j) \cdot g(X(i,j)) + N(i,j)$$

[Pérez-González21] Resolution for 2021: I promise I will write it and submit it.

Camera response functions



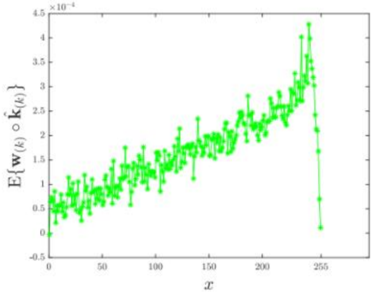
Nikon

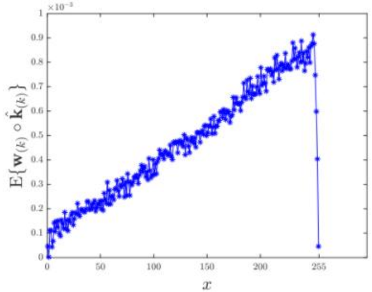


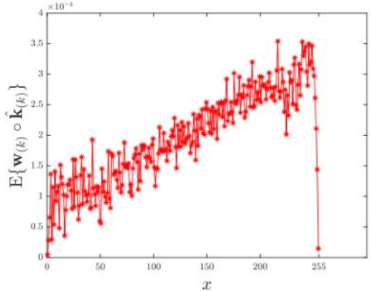
Sony NEX-5

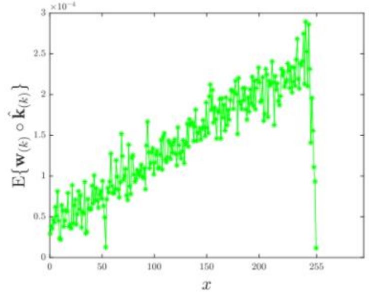
Function g(.)

Nikon D60 RAW

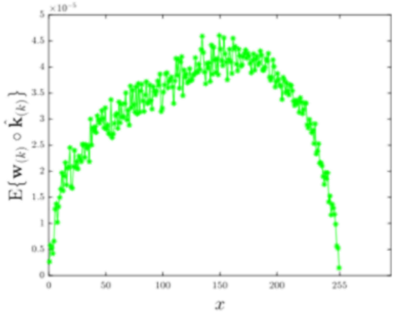

(i) TIFF-DC, $ph = (1, 1)$

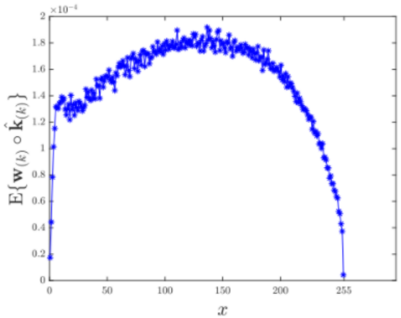

(j) TIFF-DC, $ph = (1, 2)$

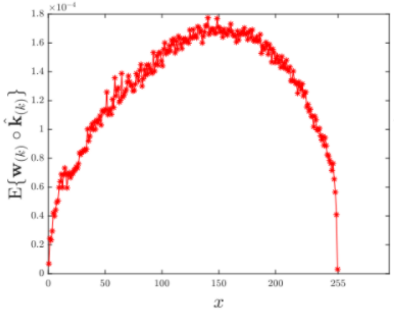

(k) TIFF-DC, $ph = (2, 1)$

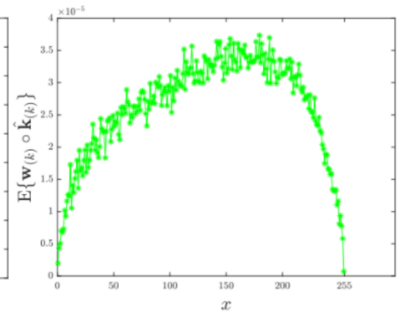

(l) TIFF-DC, $ph = (2, 2)$

Nikon D7000 TIFF


(e) G $ph = (1, 1)$.

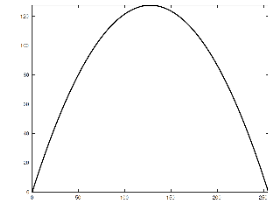

(j) B $ph = (1, 2)$.

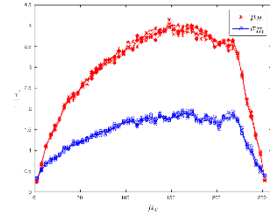

(c) R $ph = (2, 1)$.


(h) G $ph = (2, 2)$.

Extraction with g(.)

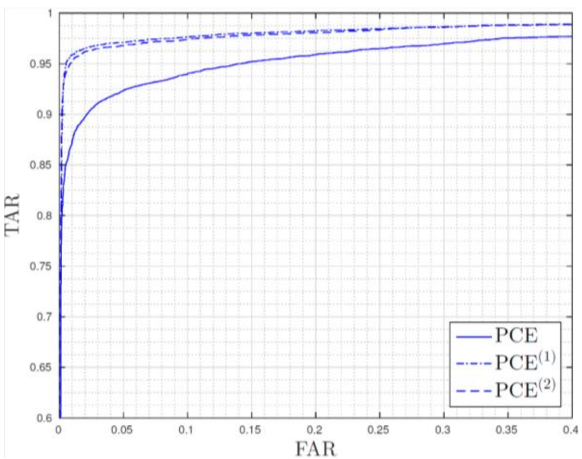
- Results with TIFF images, cropped to 512x512 patches.



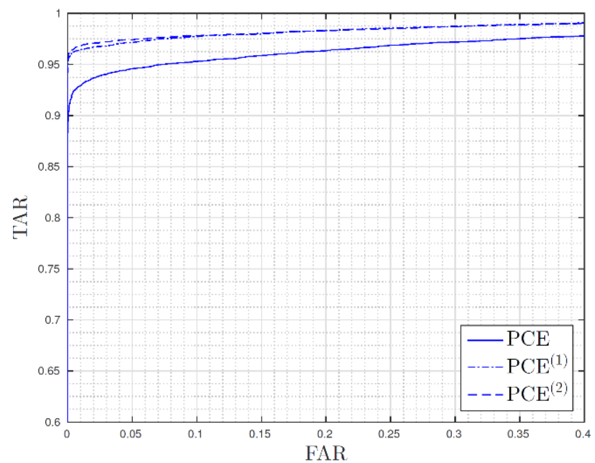


Estimation with parabola

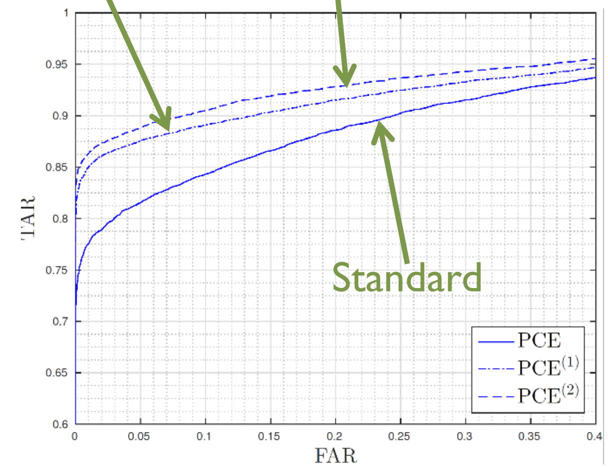
Estimation with g(x)



Nikon D3200

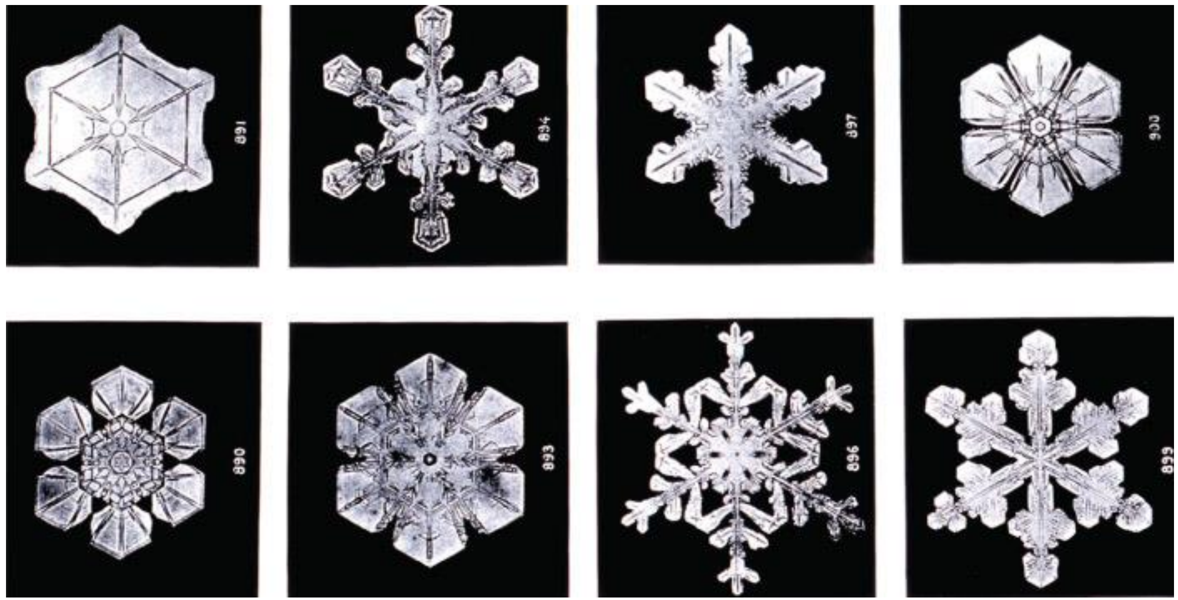


Canon D1100



Nikon D7000

2. The Snowflake Hypothesis



Sprinkle some s

PASPALEY

WIRED

POWERED BY THE DENTISTRY NETWORK DE RDH PI APEX360. ABOUT | CONTACT | NEWSLETTERS | ADVERTISE |

THE WALL STREET JOURNAL.

Subscribe Now | Sign In

$1 for 2 Months

Home World U.S. Politics Economy Business Tech Markets Opinion Arts Life Real Estate

BUSINESS How Budget Carriers Transformed the Airline ...

LIFE How Roger Federer Found His Groove, Again

MARKETS Ackerman's Take: Wells Fargo in Senate Hot Seat Again?

DRIVER'S SEAT

Forget Fingerprints: Car Seat IDs Driver's Rear End

By Yoree Koh

Jan 18, 2012 1:00 pm ET

If Shigeomi Koshimizu has his way, sometime in the not too distant future car owners may control their vehicles by the seat of their pants.

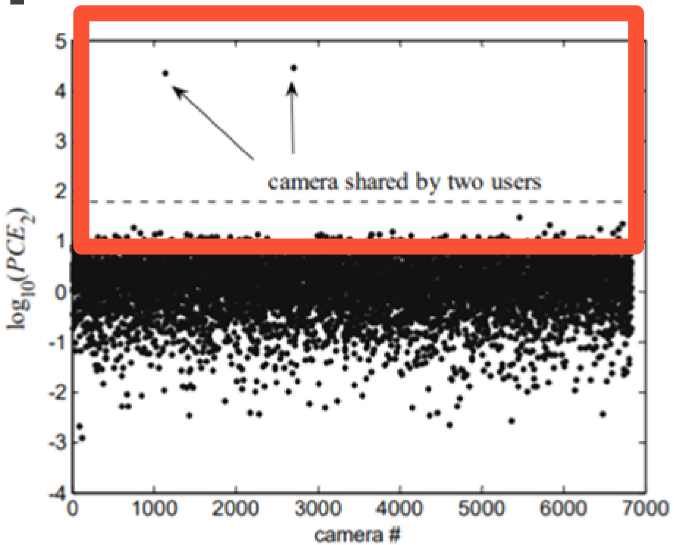
Literally.

Are FINGERprints really unique?

- US lawyer, Brandon Mayfield, mistakenly detained by FBI in connection with Madrid bombings (March 2004).
- An FBI supercomputer positively identified one of the Madrid fingerprints on a bag of detonators as Mayfield's.
- FBI maintained its certainty despite Spanish authorities denied the match.
- Actually, the fingerprints corresponded to an Algerian man.

Is the PRNU a Snowflake?

- [Goljan09] large-scale analysis with flickr images.
- Database in the wild: possibly several cameras from same user; images with digital zoom…
- Images per camera in interval [60,200] x ~7,000 cameras.
- A few cameras found to be identical.



Study in [Iuliani 20]

- VISION dataset: 35 devices, 11 brands + Control dataset: 23 smartphones, 17 different models + Flickr dataset: same models as Ctrl dataset and 31 additional models.
- No collisions reported on VISION, but on fingerprints with Ctrl dataset (PRNUs estimated with 5 flat images), yes:

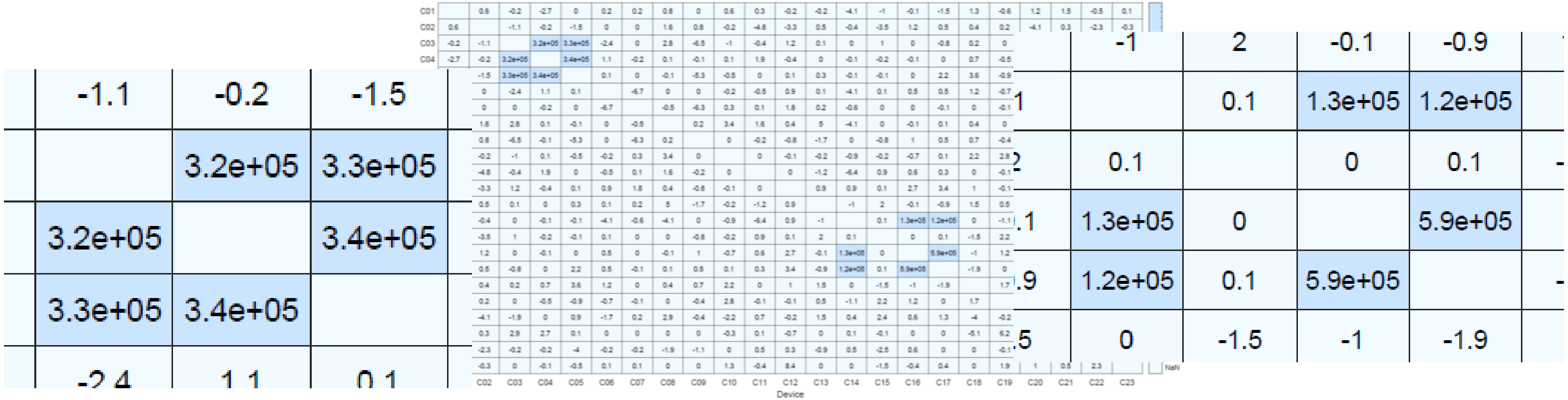
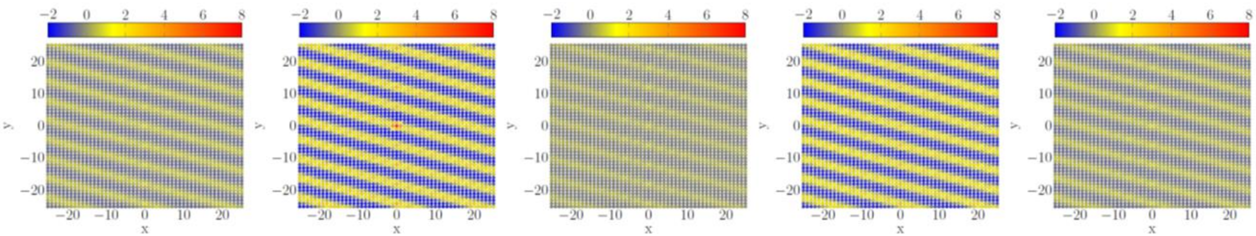


Figure 2: PCE statistics computed among different camera fingerprints in the Control dataset.

Study in [Iuliani 20]

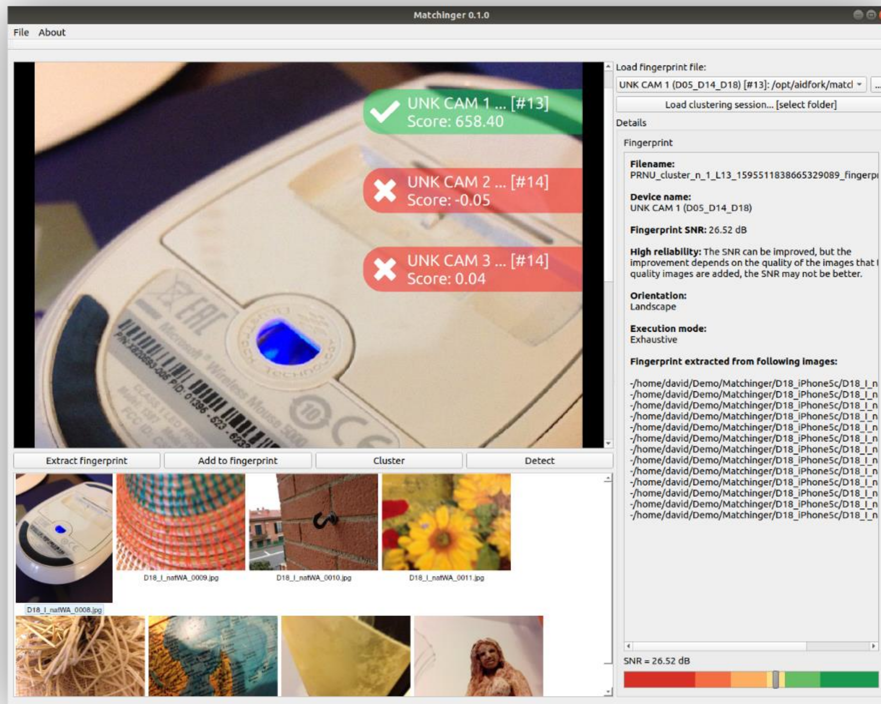
- Standard artifacts are removed (by Zero-meaning and Wiener filtering)
- "For the widely adopted PCE threshold of 60, false positive rates larger than 1% were observed for popular devices belonging to Huawei, Samsung, Nokia, and Xiaomi."
- [Gloe12] had found diagonal artifacts not entirely removable with Wiener filtering for a Nikon CoolPix S710 cameras (Dresden dataset)



(f) $c_{0;S710}$ (g) $c_{1;S710}$ (h) $c_{2;S710}$ (i) $c_{3;S710}$ (j) $c_{4;S710}$

Xcorrs of an image from c1;s710 with PRNUs of other s710's

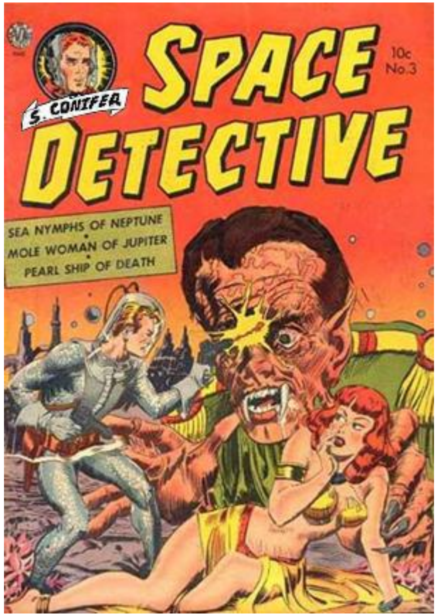
Our own experience: Matchinger



Talking about Wild? The future is Wilder!

The future is wilder

- Images and videos are more and more subject to really wild conditions:
- (Strong) compressions.
- Cropping and scaling.
- Digital zooming.
- High dynamic range imaging.
- Camera stabilization.
- In-camera/software lens distortion correction.
- Photo effects.
- Multicamera imaging.
- …



Quasi-homomorphic transformations

- If

$\mathbf{X}_t$

$\mathcal{T}_{\boldsymbol{\theta}}(\cdot)$

$\mathcal{T}_{\boldsymbol{\theta}}(\mathbf{X}_t)$

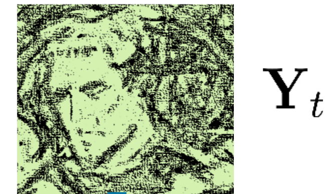
- Does

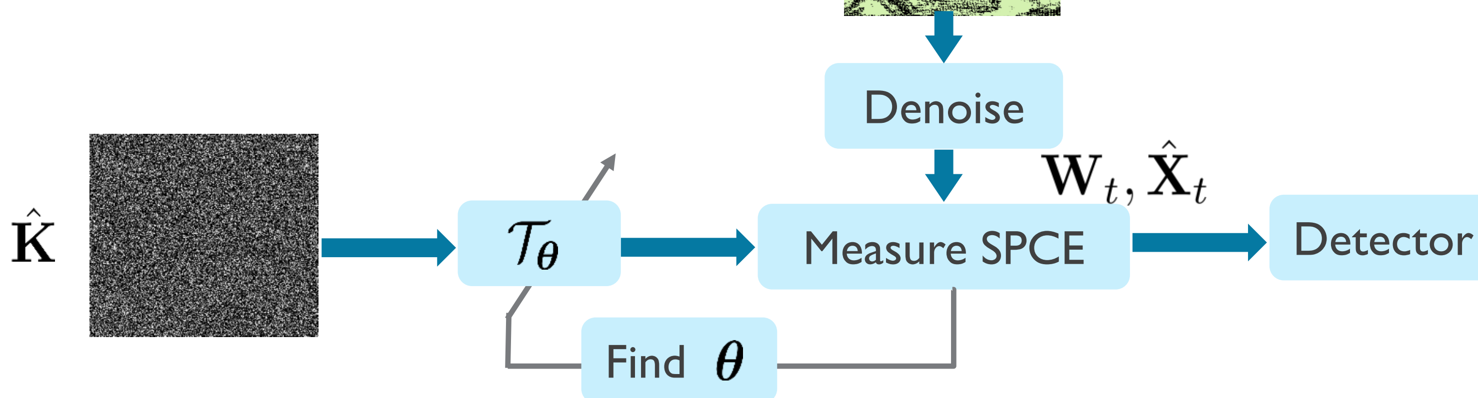
$$\mathcal{T}_{\boldsymbol{\theta}}(\mathbf{X}_t + \mathbf{X}_t \circ \mathbf{K}) \approx \mathcal{T}_{\boldsymbol{\theta}}(\mathbf{X}_t) + \mathcal{T}_{\boldsymbol{\theta}}(\mathbf{X}_t) \circ \mathcal{T}_{\boldsymbol{\theta}}(\mathbf{K}) \text{ ?}$$

Direct approach

- If so, given $\hat{\mathbf{K}}$, the detection statistic (GLRT) becomes:

$$\max_{\boldsymbol{\theta}} \ \mathrm{SPCE}\left(\mathcal{T}_{\boldsymbol{\theta}}(\hat{\mathbf{K}}) \circ \hat{\mathbf{X}}_t, \mathbf{W}_t\right)$$



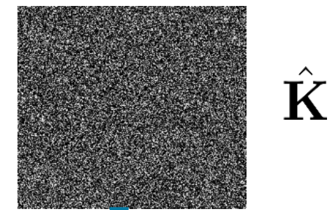


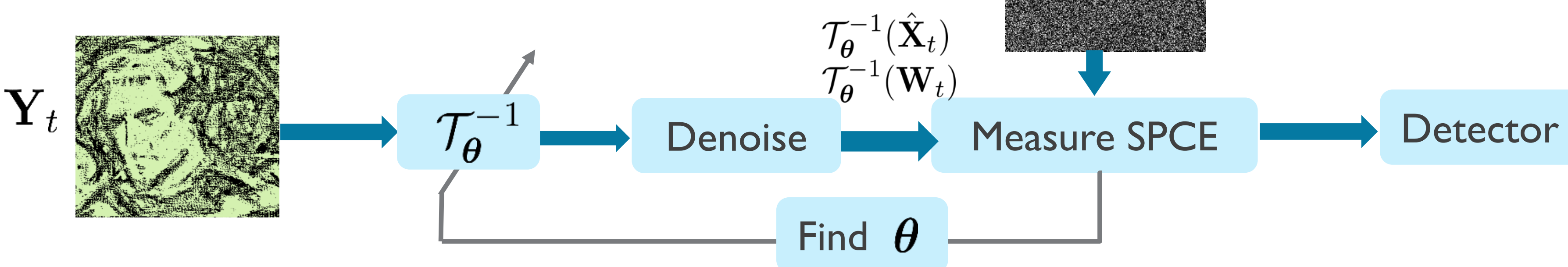
- The main challenge is to find efficient algorithms for searching the parameter space. Almost whiteness in $\mathbf{K}$ complicates things.

Inverse approach

- Based on the inverse transformation (provided it exists):

$$\max_{\boldsymbol{\theta}} \ \mathrm{SPCE}\left(\mathcal{T}_{\boldsymbol{\theta}}^{-1}(\mathbf{W}_t), \hat{\mathbf{K}} \circ \mathcal{T}_{\boldsymbol{\theta}}^{-1}(\hat{\mathbf{X}}_t)\right)$$





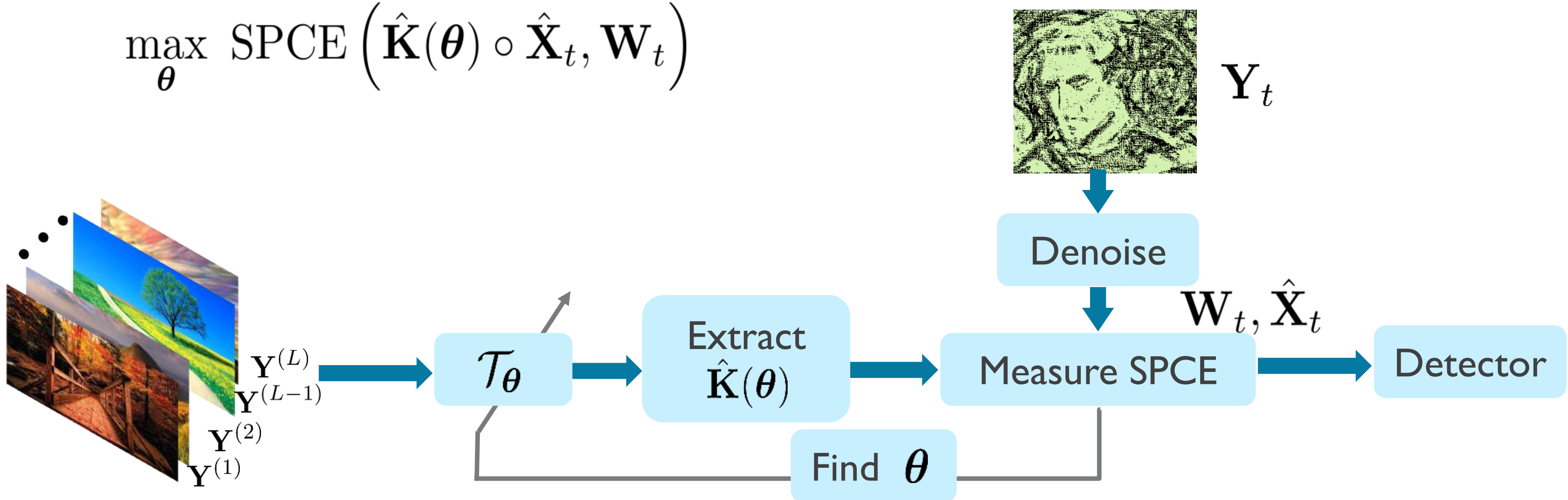
- This approach requires 1) $\mathcal{T}_{\boldsymbol{\theta}}^{-1}(\cdot)$ to be quasi-homomorphic, and 2) if $\mathbf{Y}_t = \mathbf{W}_t + \hat{\mathbf{X}}_t$, then the denoising of $\mathcal{T}_{\boldsymbol{\theta}}^{-1}(\mathbf{Y}_t)$ yields $\mathcal{T}_{\boldsymbol{\theta}}^{-1}(\hat{\mathbf{X}}_t)$ and $\mathcal{T}_{\boldsymbol{\theta}}^{-1}(\mathbf{W}_t)$.

Non-homomorphic case

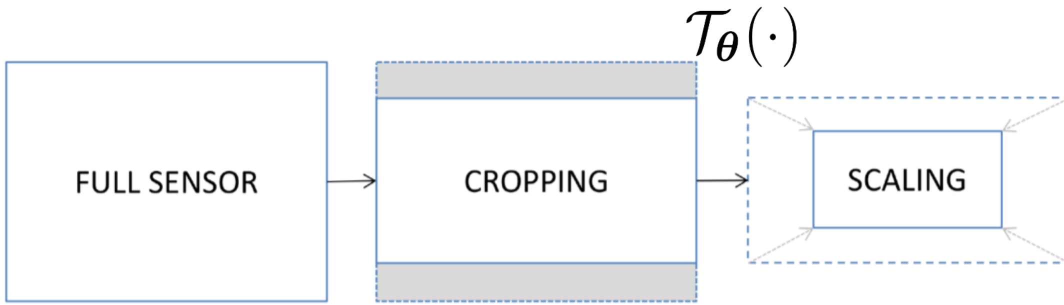
- In this case, it is much more effective (and expensive) to compute the PRNU from the residuals of transformed images

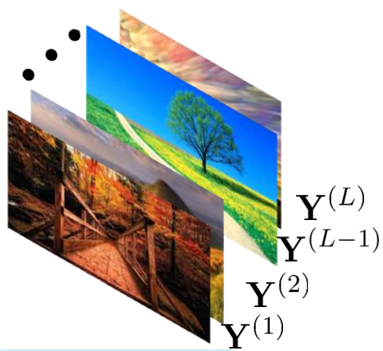
$$\hat{\mathbf{K}}(\boldsymbol{\theta}) \leftarrow \{\mathcal{T}(\mathbf{Y}^{(i)})\}_{i=1}^{L}$$

$$\max_{\boldsymbol{\theta}} \ \mathrm{SPCE}\left(\hat{\mathbf{K}}(\boldsymbol{\theta}) \circ \hat{\mathbf{X}}_t, \mathbf{W}_t\right)$$



Example: Non-stabilized video, mixed-media



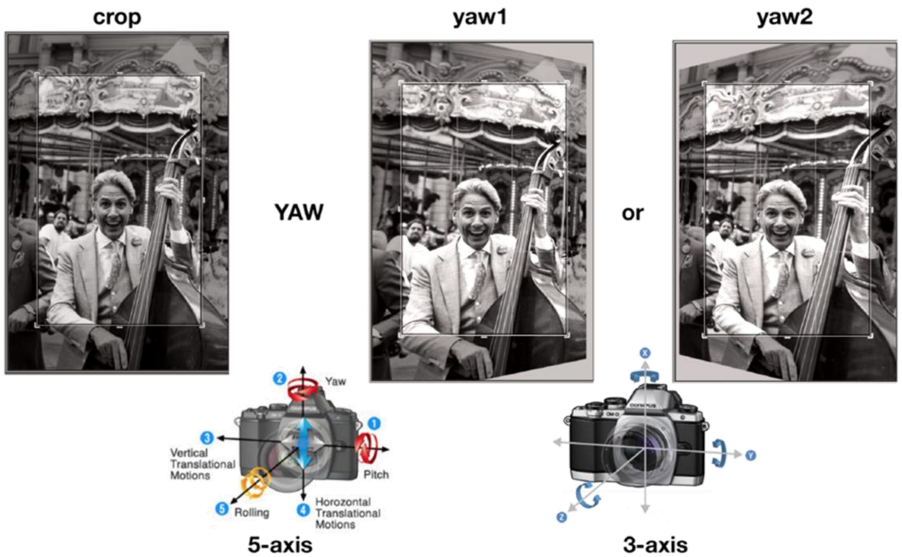


Extract PRNU → $\mathcal{T}_{\boldsymbol{\theta}}$ → Measure SPCE

Extract PRNU → Measure SPCE

Top Image: [Iuliani19]

Example: Stabilized video



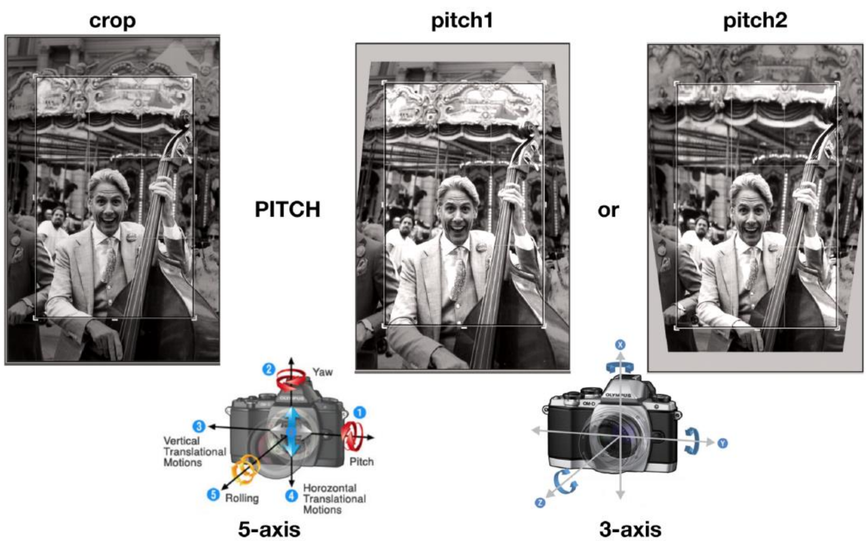


Image credit: A. Montibeller, M.S. Thesis, University of Trento.

Stabilized video

- [Chuang11]: Use the B frames. (not just I and P)
- [Taspinar16] proposed a pure brute-force approach.
- [Iuliani19]: Use still image PRNU as reference and find θ for each frame. Apply $\mathcal{T}_{\theta}^{-1}$ to register the frame. Use registered frames (with a minimum PRNU strength) to estimate video PRNU.
- [Mandelli20]: Find best frame for reference PRNU.
- [Taspinar20]: Integrate several consecutive frames to speed up calculations.

But mind the AUC!

- AUC does not reflect what happens for low FPRs.

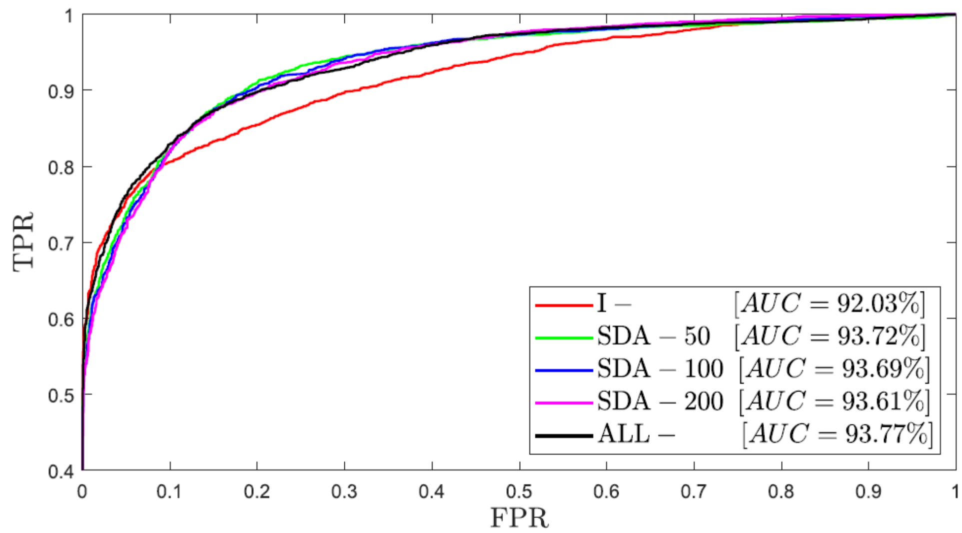
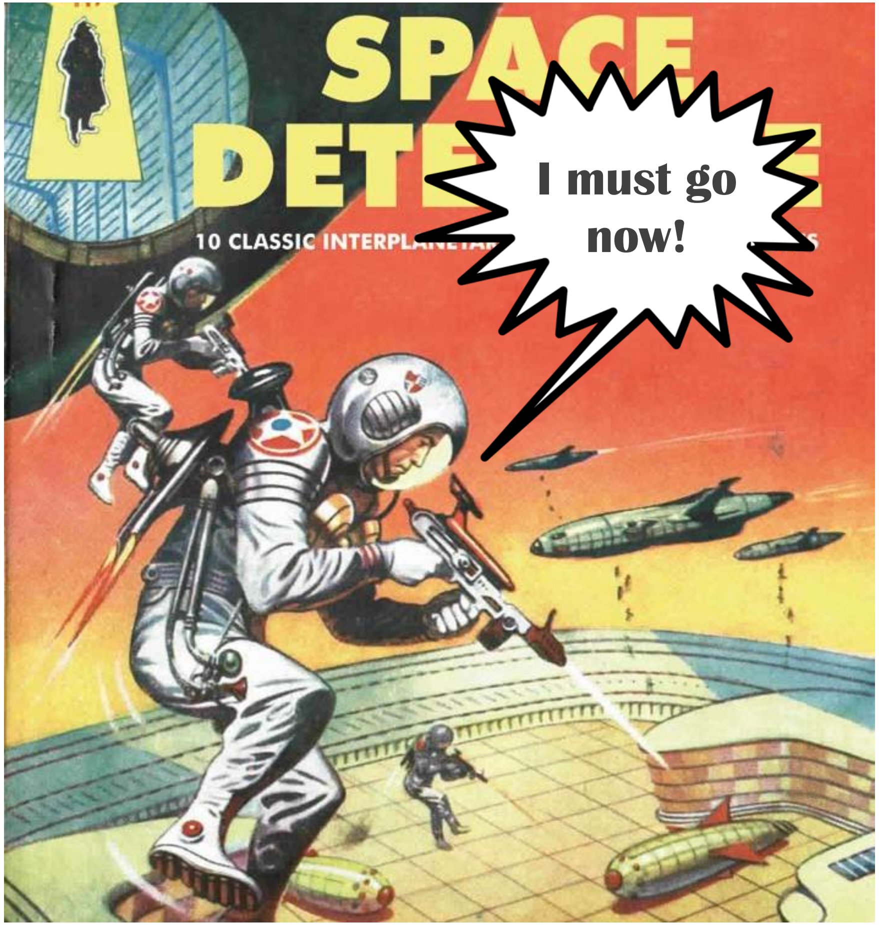
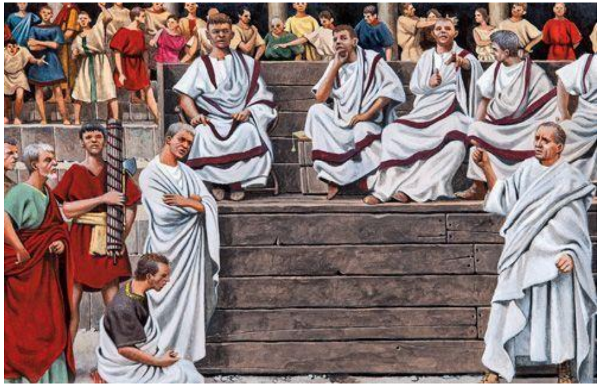


Fig. 12. ROC curves for 4-minute stabilized videos taken from VISION dataset.

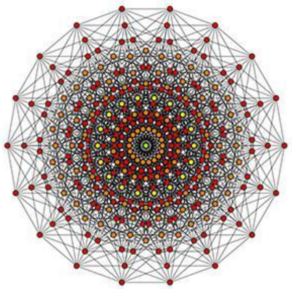
Image: [Taspinar20]


SPACE
DETE
10 CLASSIC INTERPL
I must go now!

Final thoughts



- We need to deepen our understanding and strengthen our hypotheses.
- Forensic ← Forensis ← Forum.
- We need more unbiased (meta)analyses, large-scale tests, and up-to-date databases.
- We need fresh approaches to address the curse of dimensionality, e.g., reinforcement learning.
- We need… to beat the future.




JUST
ONE
MORE
THING....

Thank you!

Fernando Pérez-González

fperez@gts.uvigo.es

References

- [Mihcak99] M. K. Mihcak, I. Kozintsev, and K. Ramchandran, "Spatially adaptive statistical modeling of wavelet image coefficients and its application to denoising," in *ICASSP 99*.
- [Kang 14] X. Kang, J. Chen, K. Lin, and A. Peng, "A context-adaptive SPN predictor for trustworthy source camera identification," *EURASIP J. Image Video Process.*,2014.
- [Al-Ani15] M. Al-Ani, F. Khelifi, A. Lawgaly, and A. Bouridane, "A novel image filtering approach for sensor fingerprint estimation in source camera identification," in *Proc. IEEE Conf. Adv. Video Signal Based Surveill. (AVSS)*, 2015.
- [Perona 90] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," IEEE Trans. Pattern Anal. Mach. Intell., vol. 12, no. 7, pp. 629–639, Jul. 1990.
- [Rudin 94] L. I. Rudin and S. Osher, "Total variation based image restoration with free local constraints," in *ICIP 94*.
- [Dabov07] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," IEEE TIP 2007.
- [Alparone06] L. Alparone, F. Argenti and G. Torricelli, "MMSE filter-ing of generalised signal-dependent noise in spatial andshift-invariant wavelet domain,"Signal Processing Journal", 2006.

- [Amerini09] I. Amerini, R. Caldelli, V. Cappellini, F. Picchioni, A. Piva, "Analysis of denoising filters for photoresponse non uniformity noise extraction in source camera identification", in International Conference on Digital Signal Processing, 2009.
- [Cortiana11] A. Cortiana, V. Conotter, G. Boato, and F. G. B. De Natale, "Performance comparison of denoising filters for source camera identification," Proc. SPIE, vol. 7880, p. 778007, Jan. 2011.
- [Al-Ani17] M. Al-Ani and F. Khelifi, On the SPN Estimation in Image Forensics: A Systematic Empirical Evaluation, IEEE TIFS 2017.
- [He13] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2013.
- [Kang12] X. Kang, Y. Li, Z. Qu, and J. Huang, "Enhancing source camera identification performance with a camera reference phase sensor pattern noise," IEEE TIFS, 2012.
- [Iuliani19] M. Iuliani, M. Fontani, D. Shullani, and A. Piva. "A hybrid approach to video source identification", Sensors, 2019.
- [Taspinar16] S. Taspinar, M. Mohanty, and N. Memon, "PRNU based source attribution with a collection of seam-carved images," in Proc. ICIP 2016.

- [Chuang 11] W.-H. Chuang, H. Su, and M. Wu, "Exploring compression effects for improved source camera identification using strongly compressed video," in Proc. ICIP 2011.
- [Mandelli20] S. Mandelli, P. Bestagini, L. Verdoliva, and S. Tubaro, "Facing device attribution problem for stabilized video sequences," IEEE Trans. Inf. Forensics Security, 2020.
- [Gloe12]: T. Gloe, S. Pfenning, M. Kirchner, "Unexpected Artefacts in PRNU-Based Camera Identification: A 'Dresden Image Database' Case-Study", MMSec 2012.
- [Iuliani20]: M. Iuliani, M. Fontani, and A. Piva, "A leak in PRNU based source identification? Questioning fingerprint uniqueness", ArXiv 2020.
- [Goljan08] M. Goljan and J. Fridrich, "Camera identification from cropped and scaled images". In Security, Forensics, Steganography, and Watermarking of Multimedia Contents X, SPIE 2008
- [Lukás06] J. Lukás, J. Fridrich, and M. Goljan, "Digital camera identification from sensor pattern noise," IEEE Trans. Inf. Forensics Security, 2006.
- [Nuzzo15] R. Nuzzo, "How scientists fool themselves–and how they can stop". Nature 2015.
- [Goljan09] M. Goljan, J. Fridrich, and T. Filler, "Large scale test of sensor fingerprint camera identification," Proc. SPIE, 2009.