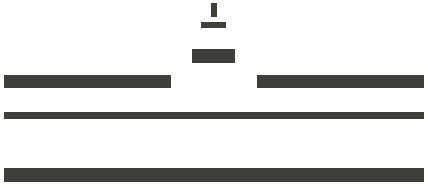


Westfälische
Wilhelms-Universität
Münster

Steganalysis in Technicolor

Boosting WS Detection of Stego Images from CFA-Interpolated Covers

Matthias Kirchner and Rainer Böhme
University of Münster

IEEE ICASSP | Florence, Italy | May 8, 2014

Are Steganographers Colorblind?

- research community seems to live in a monochromatic world where grayscale images abound

Are Steganographers Colorblind?

- research community seems to live in a monochromatic world where grayscale images abound
- a more colorful world raises questions about steganographic security

Are Steganographers Colorblind?

- research community seems to live in a monochromatic world where grayscale images abound
- a more colorful world raises questions about steganographic security

1 how plausible are grayscale images?

Are Steganographers Colorblind?

- research community seems to live in a monochromatic world where grayscale images abound
- a more colorful world raises questions about steganographic security

1. how plausible are grayscale images?
2. how much can steganalysts gain from color information?

Are Steganographers Colorblind?

- research community seems to live in a monochromatic world where grayscale images abound
- a more colorful world raises questions about steganographic security

1. how plausible are grayscale images?
2. how much can steganalysts gain from color information?

- this work: WS steganalysis and bilinear CFA interpolation

WS Steganalysis

[Fridrich & Goljan, 2004; Ker & Böhme, 2008]

- estimates the embedding rate $\hat{p}$ of uniform LSB replacement embedding in grayscale images

$$\hat{p} = \frac{2}{n} \sum_{i=1}^{n} (-1)^{x_i^{(p)}} \left(x_i^{(p)} - \hat{x}_i^{(o)} \right)$$

$\mathbf{x}^{(o)}$	cover object $\in \mathbb{Z}^n$
$\mathbf{x}^{(p)}$	stego object $\in \mathbb{Z}^n$
$\hat{\mathbf{x}}^{(o)}$	cover estimate
p	embedding rate

WS Steganalysis

[Fridrich & Goljan, 2004; Ker & Böhme, 2008]

- estimates the embedding rate $\hat{p}$ of uniform LSB replacement embedding in grayscale images

$$\hat{p} = \frac{2}{n} \sum_{i=1}^{n} (-1)^{x_i^{(p)}} \left(x_i^{(p)} - \hat{x}_i^{(o)} \right)$$

$\mathbf{x}^{(o)}$	cover object $\in \mathbb{Z}^n$
$\mathbf{x}^{(p)}$	stego object $\in \mathbb{Z}^n$
$\hat{\mathbf{x}}^{(o)}$	cover estimate
p	embedding rate

- cover estimate: linear prediction from spatial neighboorhood

$$\mathcal{F}_{\text{KB8}} : \begin{matrix} -\frac{1}{4} & \frac{1}{2} & -\frac{1}{4} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ -\frac{1}{4} & \frac{1}{2} & -\frac{1}{4} \end{matrix} \qquad \mathcal{F}_{\text{LS8}} : \begin{matrix} b & a & b \\ a & 0 & a \\ b & a & b \end{matrix}$$

WS Steganalysis

[Fridrich & Goljan, 2004; Ker & Böhme, 2008]

- estimates the embedding rate $\hat{p}$ of uniform LSB replacement embedding in grayscale images

$$\hat{p} = \frac{2}{n} \sum_{i=1}^{n} w_i \left(-1\right)^{x_i^{(p)}} \left(x_i^{(p)} - \hat{x}_i^{(o)}\right)$$

$\mathbf{x}^{(o)}$	cover object $\in \mathbb{Z}^n$
$\mathbf{x}^{(p)}$	stego object $\in \mathbb{Z}^n$
$\hat{\mathbf{x}}^{(o)}$	cover estimate
p	embedding rate
$\mathbf{w}$	vector of weights

- cover estimate: linear prediction from spatial neighboorhood

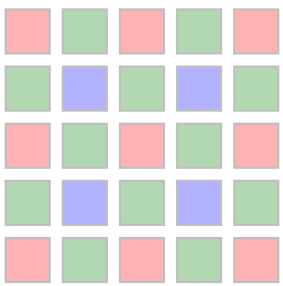
$$\mathcal{F}_{\text{KB8}} : \begin{matrix} -\frac{1}{4} & \frac{1}{2} & -\frac{1}{4} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ -\frac{1}{4} & \frac{1}{2} & -\frac{1}{4} \end{matrix} \qquad \mathcal{F}_{\text{LS8}} : \begin{matrix} b & a & b \\ a & 0 & a \\ b & a & b \end{matrix}$$

- enhanced variants for various assumptions about cover source and embedding strategies

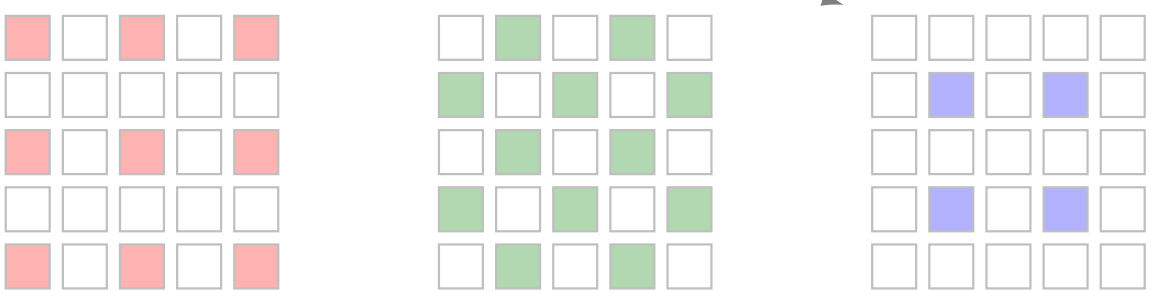
CFA Interpolation

- typical digital images are captured with a color filter array

Bayer pattern



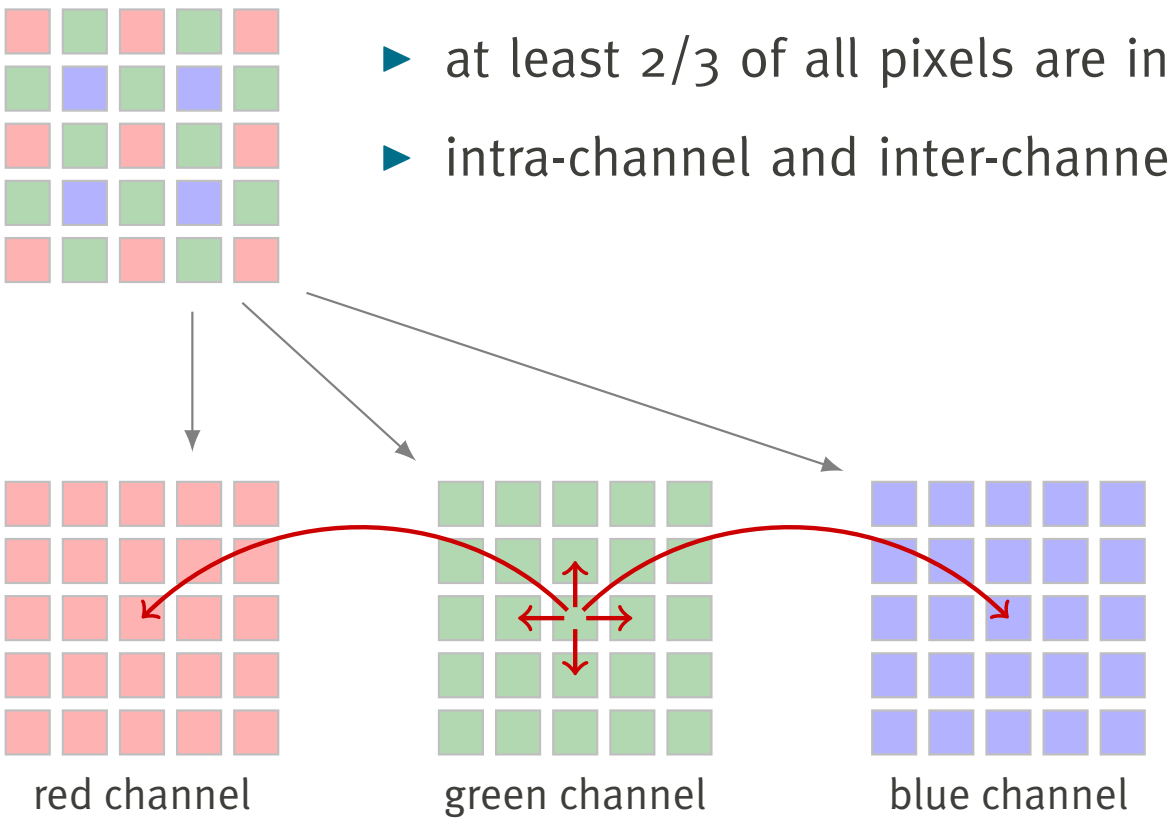
- at least 2/3 of all pixels are interpolated



CFA Interpolation

- typical digital images are captured with a color filter array

Bayer pattern

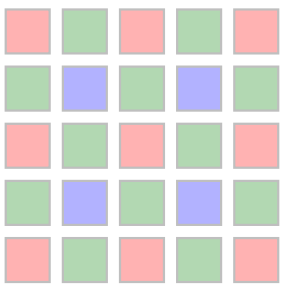


- at least 2/3 of all pixels are interpolated
- intra-channel and inter-channel dependencies

CFA Interpolation

- typical digital images are captured with a color filter array

Bayer pattern

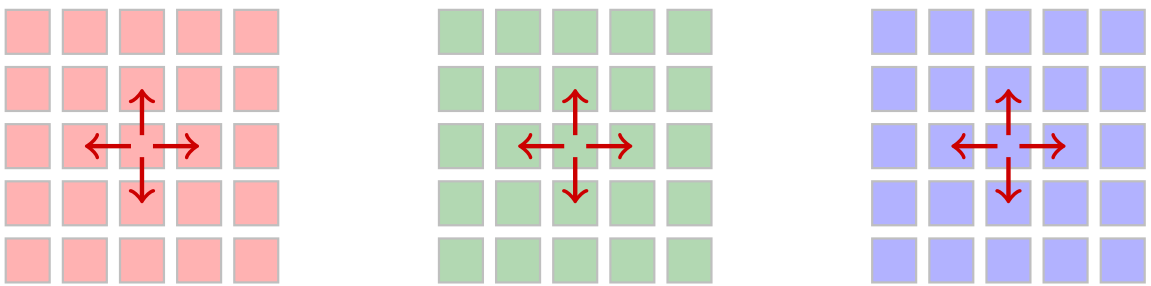


- at least 2/3 of all pixels are interpolated
- intra-channel and inter-channel dependencies

bilinear interpolation:

$$\mathcal{F}_{\text{green}}: \begin{pmatrix} 0 & \frac{1}{4} & 0 \\ \frac{1}{4} & 1 & \frac{1}{4} \\ 0 & \frac{1}{4} & 0 \end{pmatrix}$$

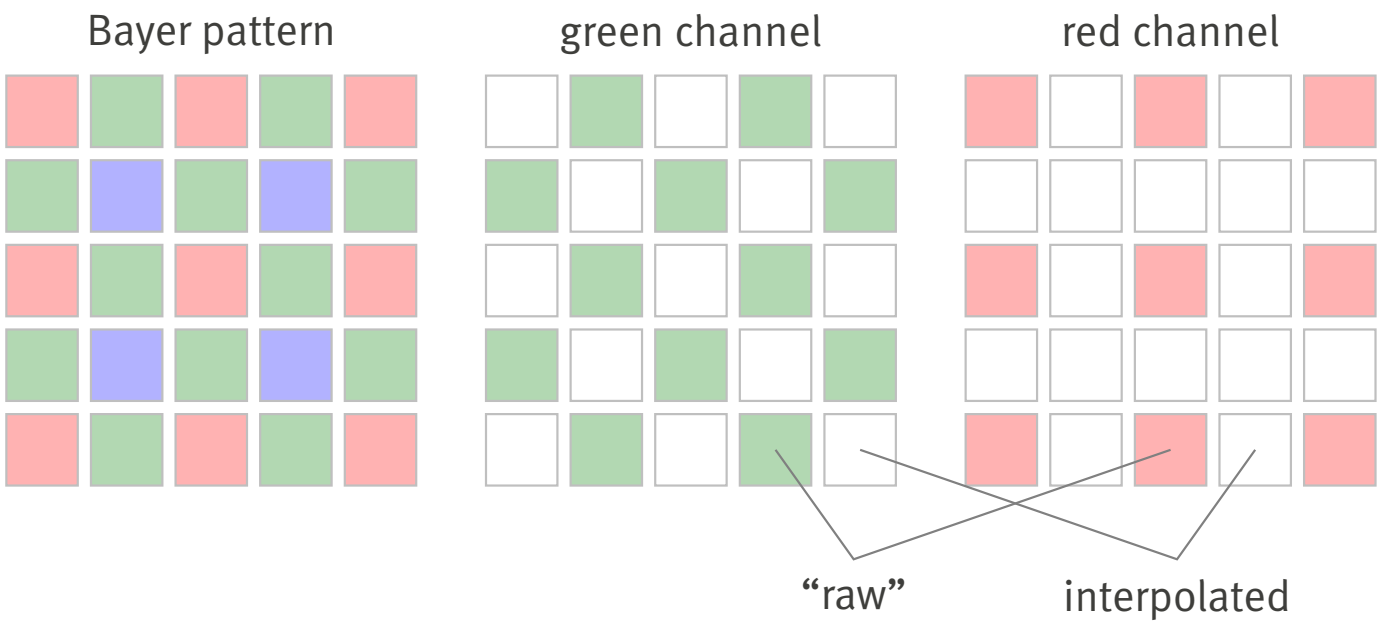
$$\mathcal{F}_{\text{red}}: \begin{pmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{2} & 1 & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{pmatrix}$$



red channel | green channel | blue channel

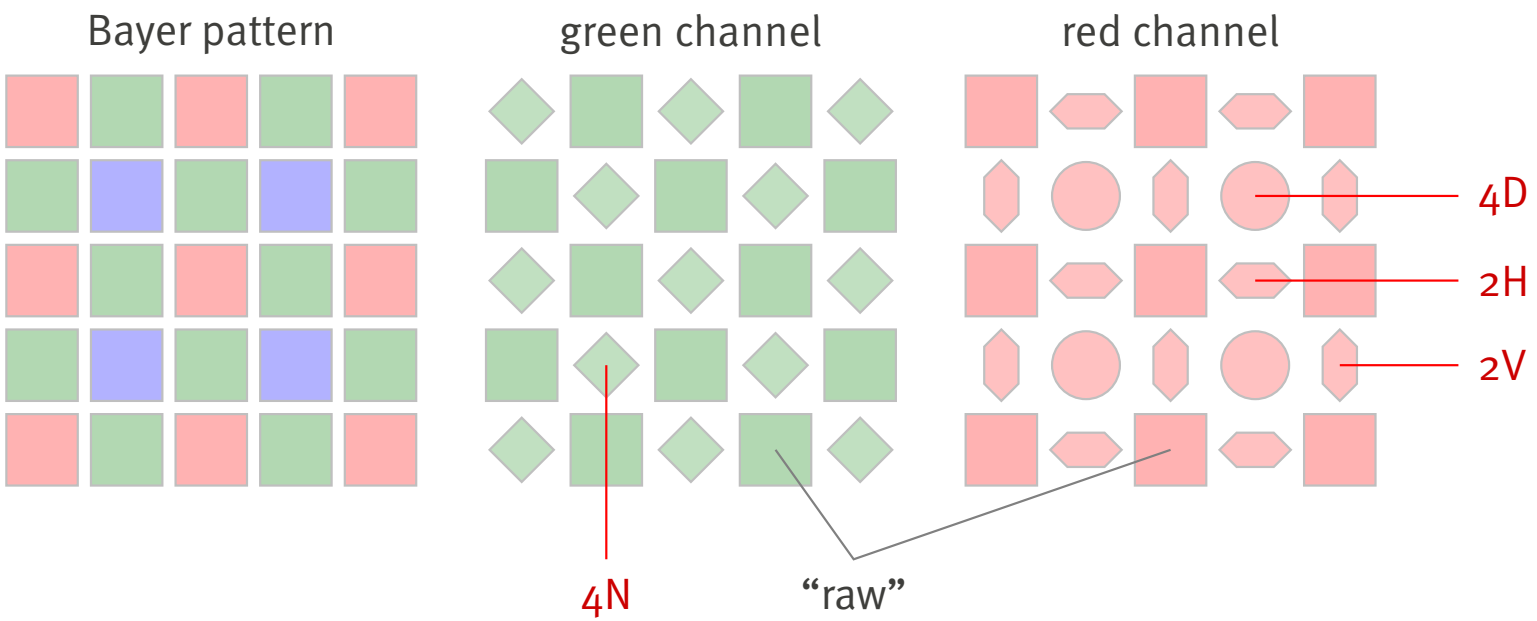
Spatial Neighborhood Categories

- CFA configuration defines characteristic pixel neighborhood categories



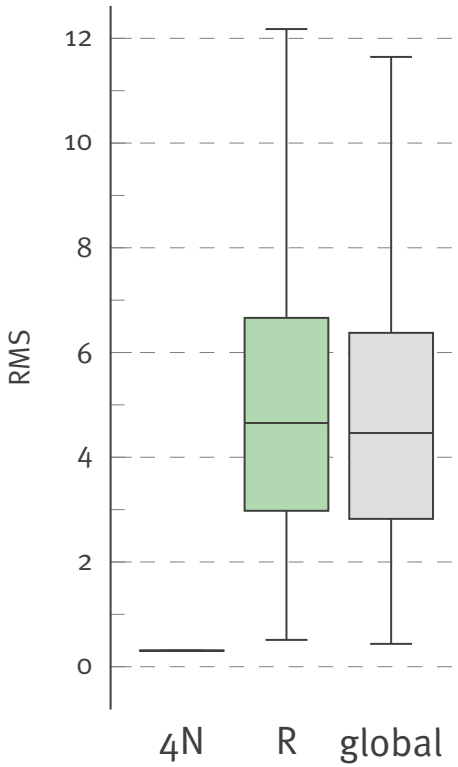
Spatial Neighborhood Categories

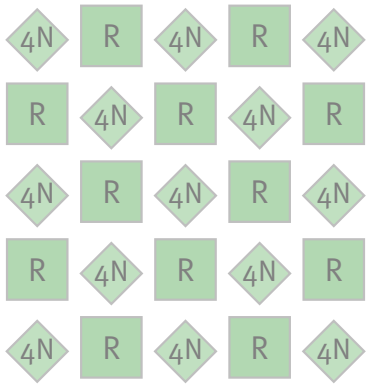
- CFA configuration defines characteristic pixel neighborhood categories
- $\mathbf{x}_{\{C\}} = (x_i \mid i \in C) \quad C \in \{R, 4N, 4D, 2H, 2V\}$



Spatial Neighborhood Predictability

- OLS linear prediction from all 3×3 neighborhoods ($\mathcal{F}_{\mathrm{LS8}}$) per category
- prediction error

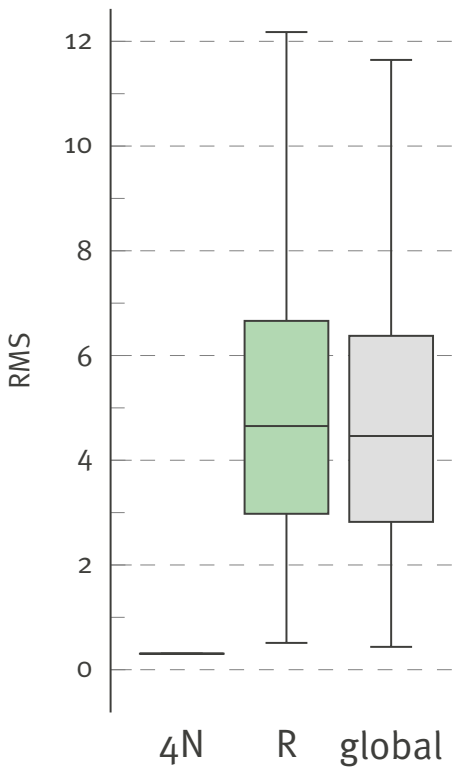


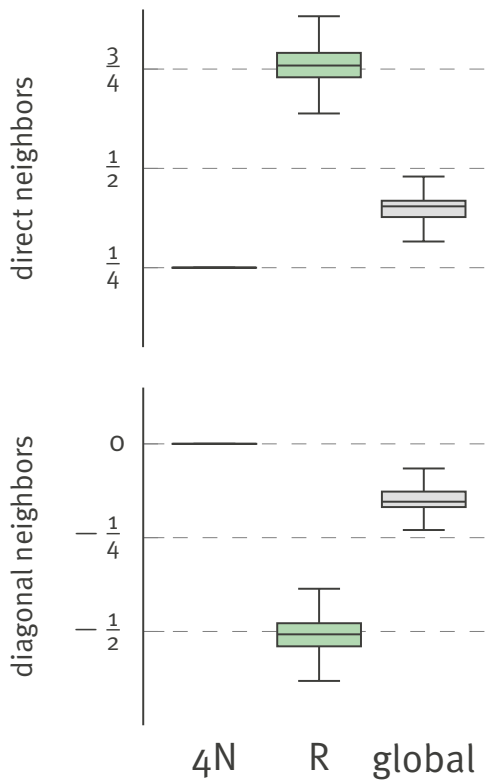


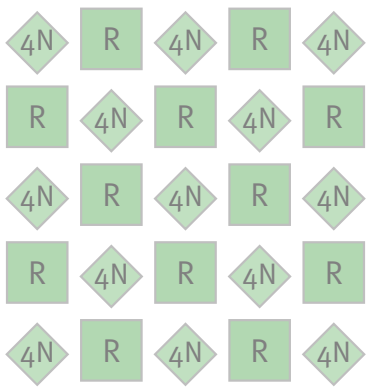
bilinear interpolation, green channel,
7408 images (size: 512×512)

Spatial Neighborhood Predictability

- OLS linear prediction from all 3×3 neighborhoods ($\mathcal{F}_{\mathrm{LS8}}$) per category
- prediction error
- estimated coefficients



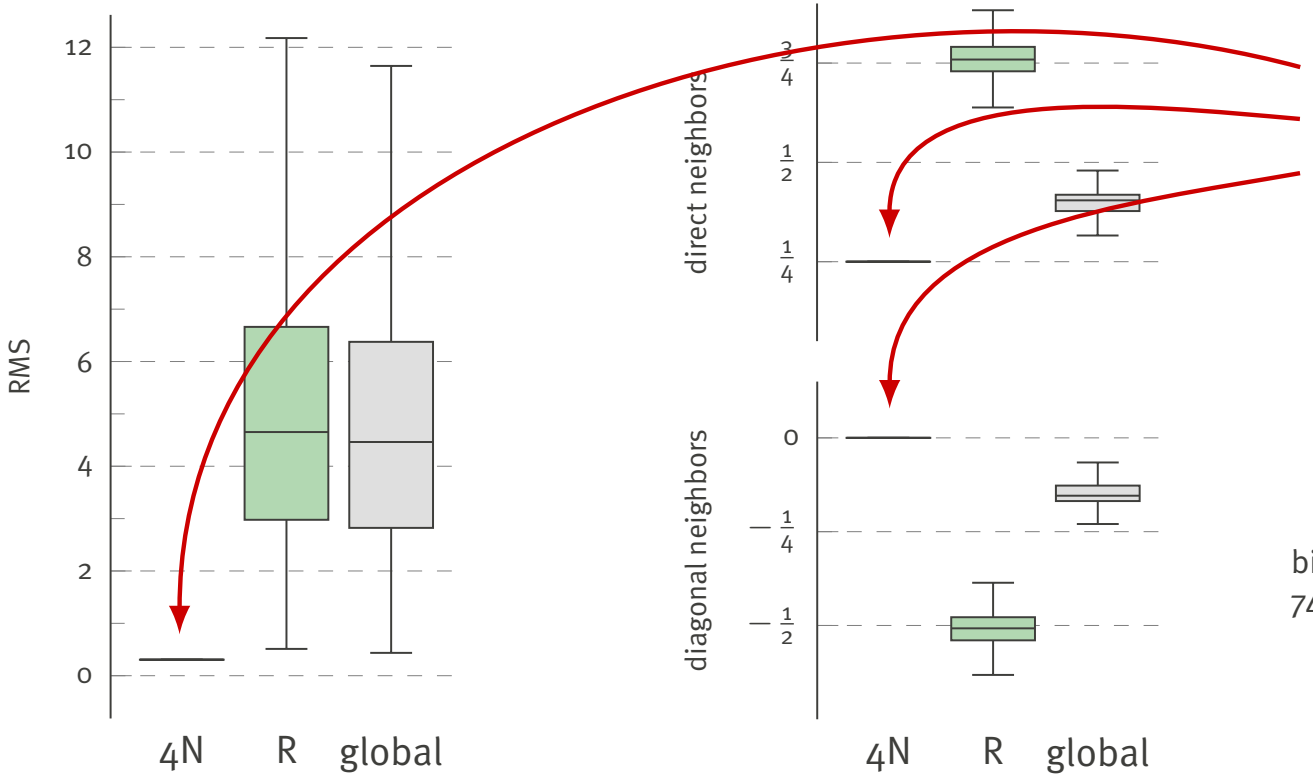




bilinear interpolation, green channel,
7408 images (size: 512×512)

Spatial Neighborhood Predictability

- OLS linear prediction from all 3×3 neighborhoods ($\mathcal{F}_{\mathrm{LS8}}$) per category
- prediction error
- estimated coefficients



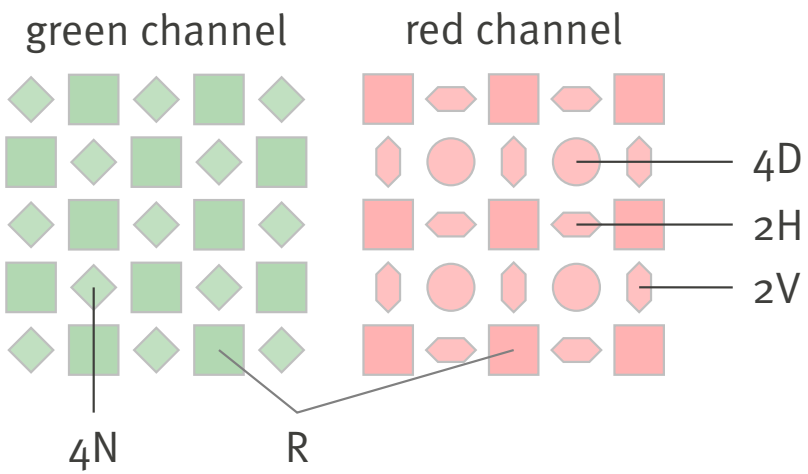
"perfect" predictability of interpolated pixels

bilinear interpolation, green channel, 7408 images (size: 512×512)

CFA-Inspired Linear Pixel Predictors

- different neighborhood categories, C, call for tailored predictors

  option 1: estimated from image per category, $\mathcal{F}_{\text{LS8}}\left(\mathbf{x}^{(p)}_{\{C\}}\right)$



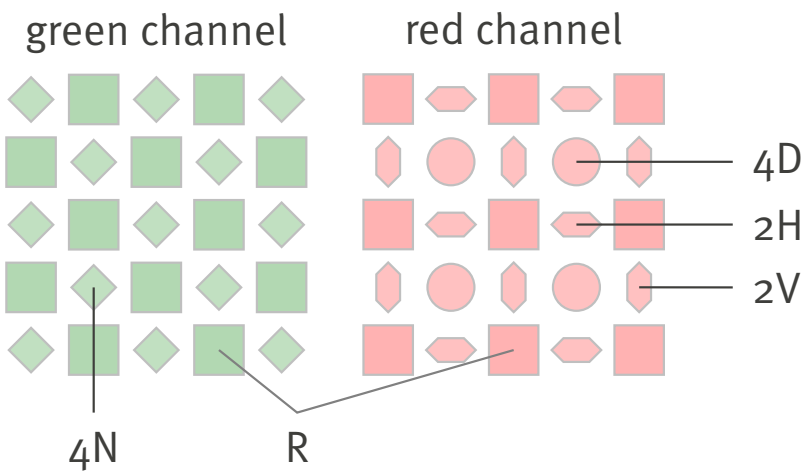
CFA-Inspired Linear Pixel Predictors

- different neighborhood categories, C, call for tailored predictors

  option 1: estimated from image per category, $\mathcal{F}_{\text{LS8}}\left(\mathbf{x}^{(p)}_{\{C\}}\right)$

  option 2: fixed pre-set filter kernels, $\mathcal{F}_{C}\left(\mathbf{x}^{(p)}_{\{C\}}\right)$

bilinear interpolation:



$$\mathcal{F}_{4\text{N}} = \begin{bmatrix} 0 & \frac{1}{4} & 0 \\ \frac{1}{4} & 0 & \frac{1}{4} \\ 0 & \frac{1}{4} & 0 \end{bmatrix} \qquad \mathcal{F}_{2\text{H}} = \begin{bmatrix} 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 \end{bmatrix}$$

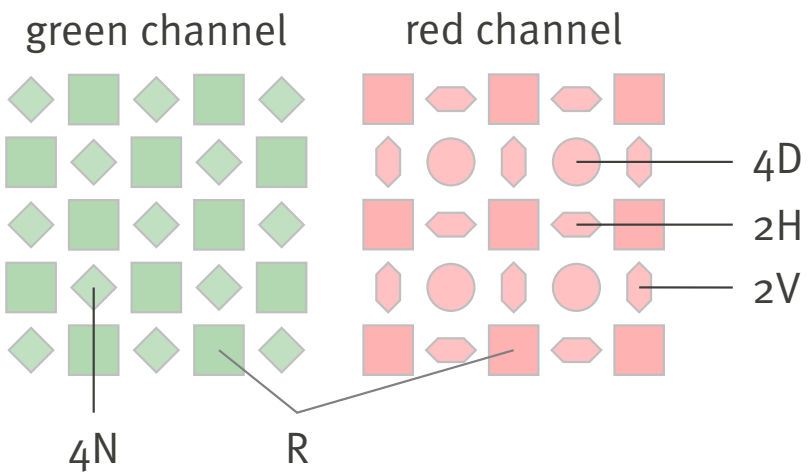
$$\mathcal{F}_{4\text{D}} = \begin{bmatrix} \frac{1}{4} & 0 & \frac{1}{4} \\ 0 & 0 & 0 \\ \frac{1}{4} & 0 & \frac{1}{4} \end{bmatrix} \qquad \mathcal{F}_{2\text{V}} = \begin{bmatrix} 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \end{bmatrix}$$

CFA-WS Steganalysis

- utilization of CFA neighborhood relations in individual color channels

$$\hat{p}_C = \frac{2}{|\{C\}|} \sum_{\{i \in C\}} (-1)^{x_i^{(p)}} \left(x_i^{(p)} - \mathcal{F}_C\left(\mathbf{x}^{(p)}\right)_i \right)$$

$\mathbf{x}^{(p)}$	stego object $\in \mathbb{Z}^n$
$\mathcal{F}_C$	predictor for category $C \in \{R, 4N, 4D, 2H, 2V\}$
p	embedding rate

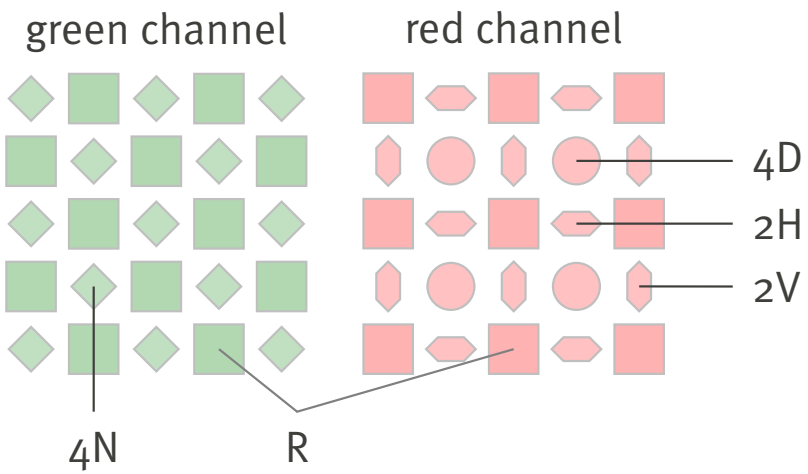


CFA-WS Steganalysis

- utilization of CFA neighborhood relations in individual color channels

$$\hat{p}_C = \frac{2}{|\{C\}|} \sum_{\{i \in C\}} w_i \, (-1)^{x_i^{(p)}} \left(x_i^{(p)} - \mathcal{F}_C\left(\mathbf{x}^{(p)}\right)_i \right)$$

$\mathbf{x}^{(p)}$	stego object $\in \mathbb{Z}^n$
$\mathcal{F}_C$	predictor for category $C \in \{R, 4N, 4D, 2H, 2V\}$
p	embedding rate
$\mathbf{w}$	vector of weights



- aggregation to a combined estimate $\hat{p}$ is equivalent to assigning weights

Experimental Setup

- 10,000 BOSSBase grayscale images (size: 512×512), sampled onto a Bayer grid to apply plain bilinear CFA interpolation

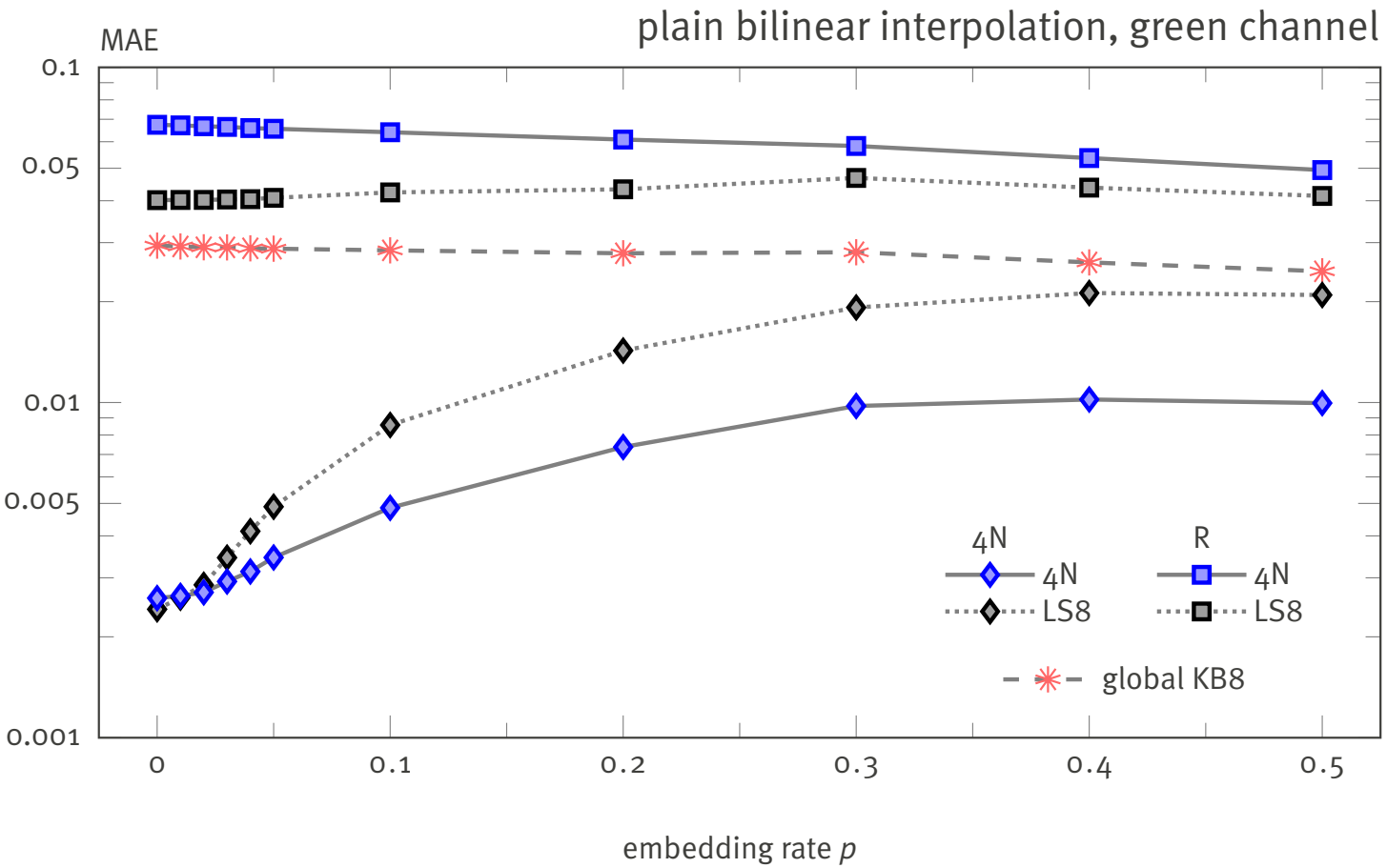
Experimental Setup

- 10,000 BOSSBase grayscale images (size: 512×512), sampled onto a Bayer grid to apply plain bilinear CFA interpolation
- 3,400 raw color image patches (size: 512×512) from the Dresden Image Database (Nikon D70), each processed with `dcraw` bilinear interpolation and proprietary content-adaptive Adobe Lightroom interpolation

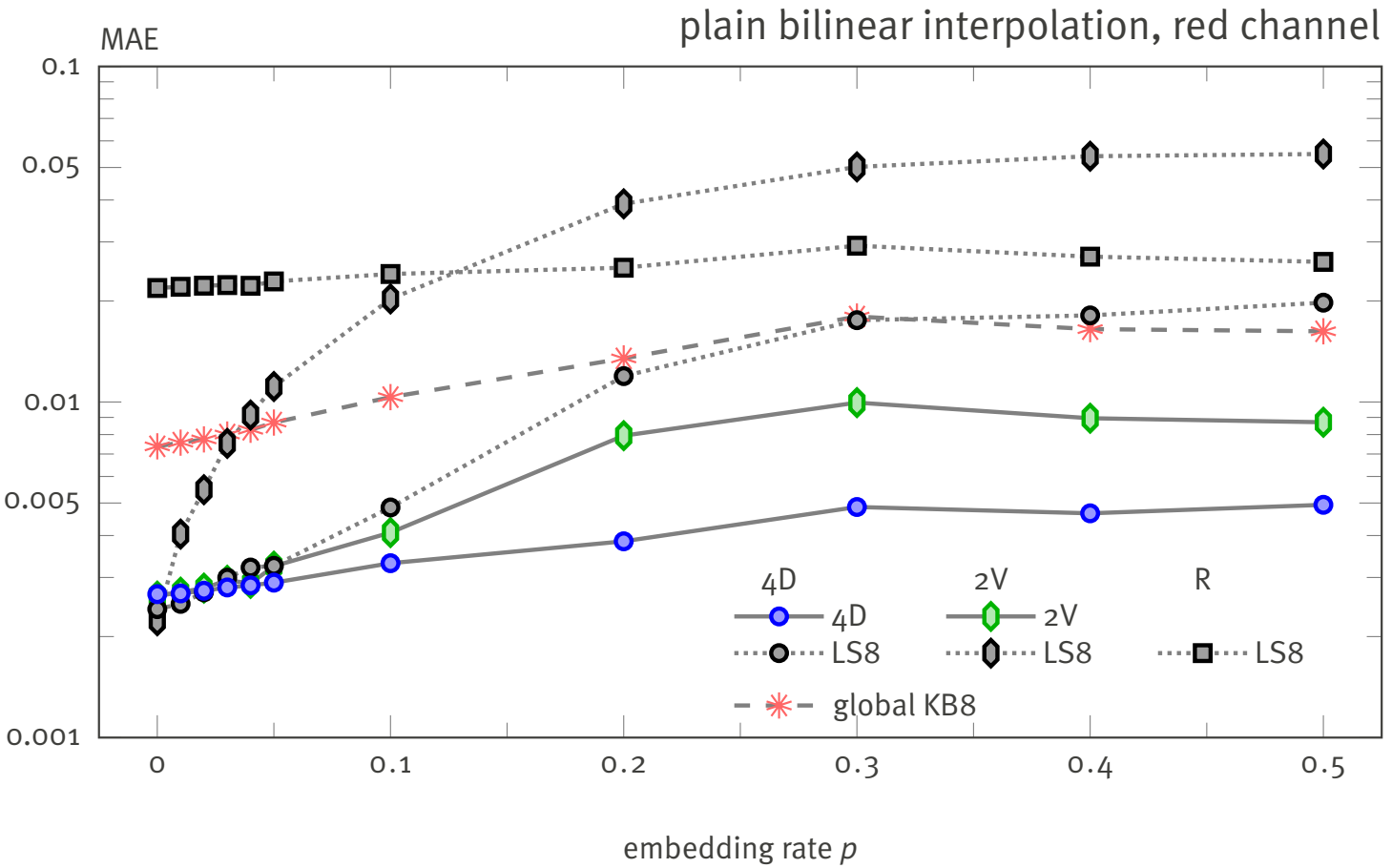
Experimental Setup

- 10,000 BOSSBase grayscale images (size: 512×512), sampled onto a Bayer grid to apply plain bilinear CFA interpolation
- 3,400 raw color image patches (size: 512×512) from the Dresden Image Database (Nikon D70), each processed with `dcraw` bilinear interpolation and proprietary content-adaptive Adobe Lightroom interpolation
- uniform LSB embedding
- exclude covers with $>5\,\%$ flat blocks (size: 3×3)

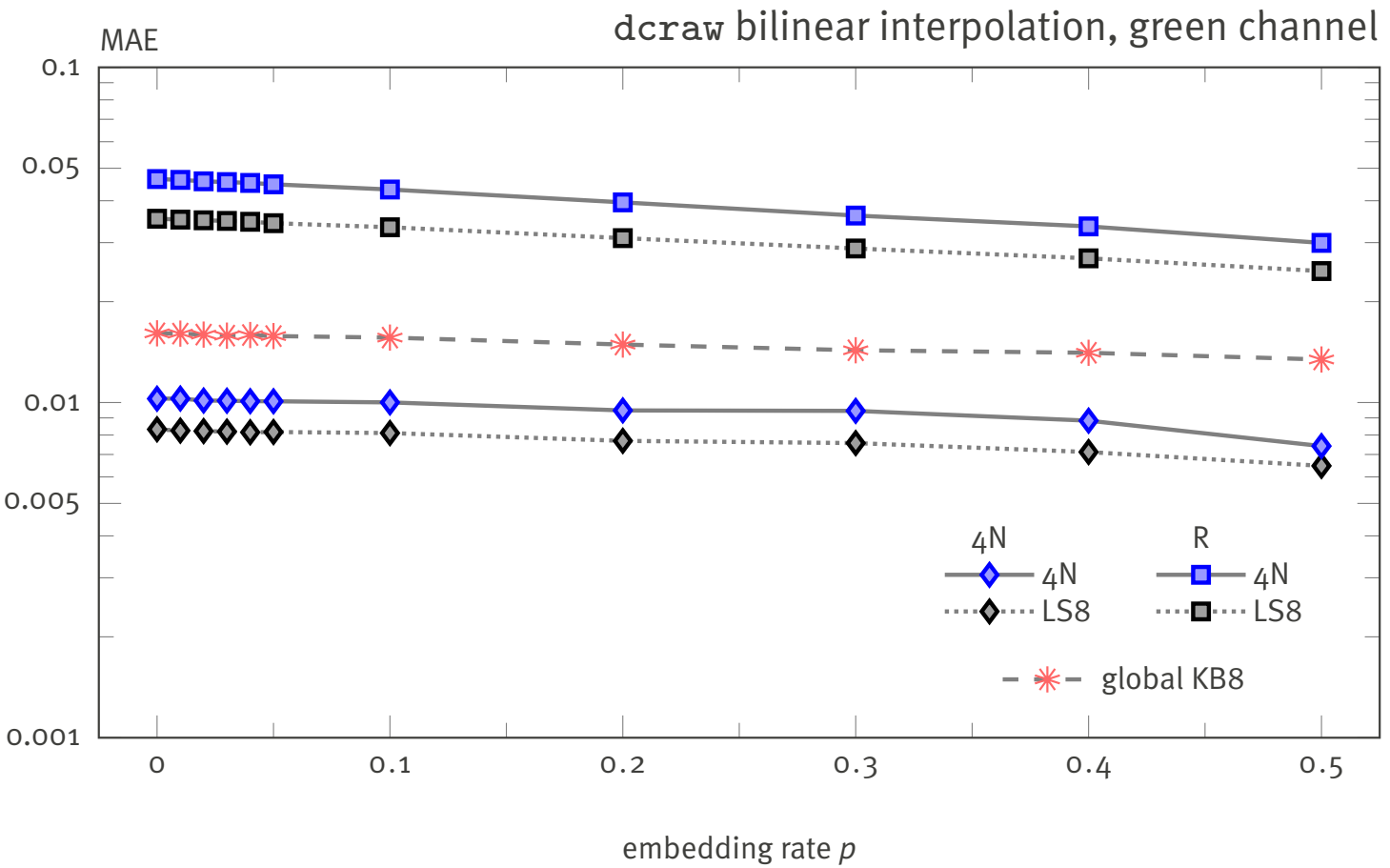
Steganalysis Results (I) – Estimation



Steganalysis Results (II) – Estimation



Steganalysis Results (III) – Estimation



Steganalysis Results (IV) – Detection

- increasing mismatch between CFA modelling assumptions and reality lets $\mathcal{F}_{\text{KB8}}$ gain advantage

($p = 0.01$)	bilinear				adaptive	
	plain N=7,408		`dcraw` N=3,316		Lightroom N=3,166	
$\mathcal{F}$	FP_{50}	EER	FP_{50}	EER	FP_{50}	EER
Standard WS (KB8)						
$\mathcal{F}_{\text{KB8}}$	0.35	0.41	0.23	0.33	0.08	0.19
$\mathcal{F}_{\text{LS8}}$	0.38	0.43	0.26	0.34	0.09	0.20
Proposed CFA-WS (4N)						
$\mathcal{F}_{\text{4N}}$	0.01	0.05	0.15	0.20	0.19	0.30
$\mathcal{F}_{\text{LS8}}$	0.01	0.04	0.10	0.16	0.11	0.23

green channel

Summary and Outlook

- color image steganography is hard

Summary and Outlook

- color image steganography is hard—and largely unexplored

Summary and Outlook

- color image steganography is hard—and largely unexplored
- good news: substantial steganalysis performance boosts hinge on sufficient information about CFA interpolation function

Summary and Outlook

- color image steganography is hard—and largely unexplored
- good news: substantial steganalysis performance boosts hinge on sufficient information about CFA interpolation function
- bad news: image forensics research

Summary and Outlook

- color image steganography is hard—and largely unexplored
- good news: substantial steganalysis performance boosts hinge on sufficient information about CFA interpolation function
- bad news: image forensics research
- good news: counter-forensics research

Summary and Outlook

- color image steganography is hard—and largely unexplored

- good news: substantial steganalysis performance boosts hinge on sufficient information about CFA interpolation function
- bad news: image forensics research
- good news: counter-forensics research
- join forces?

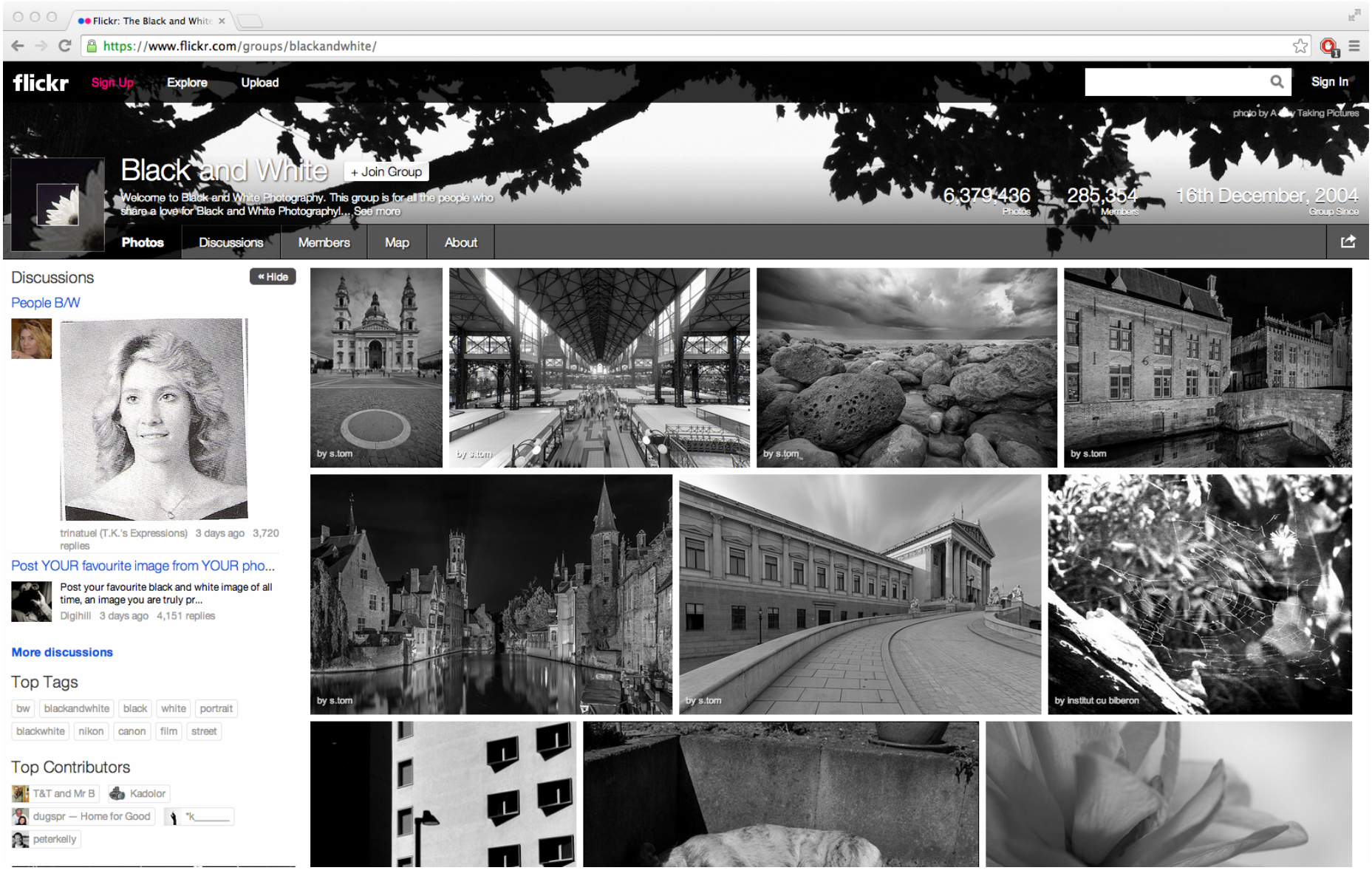
Summary and Outlook

- color image steganography is hard—and largely unexplored

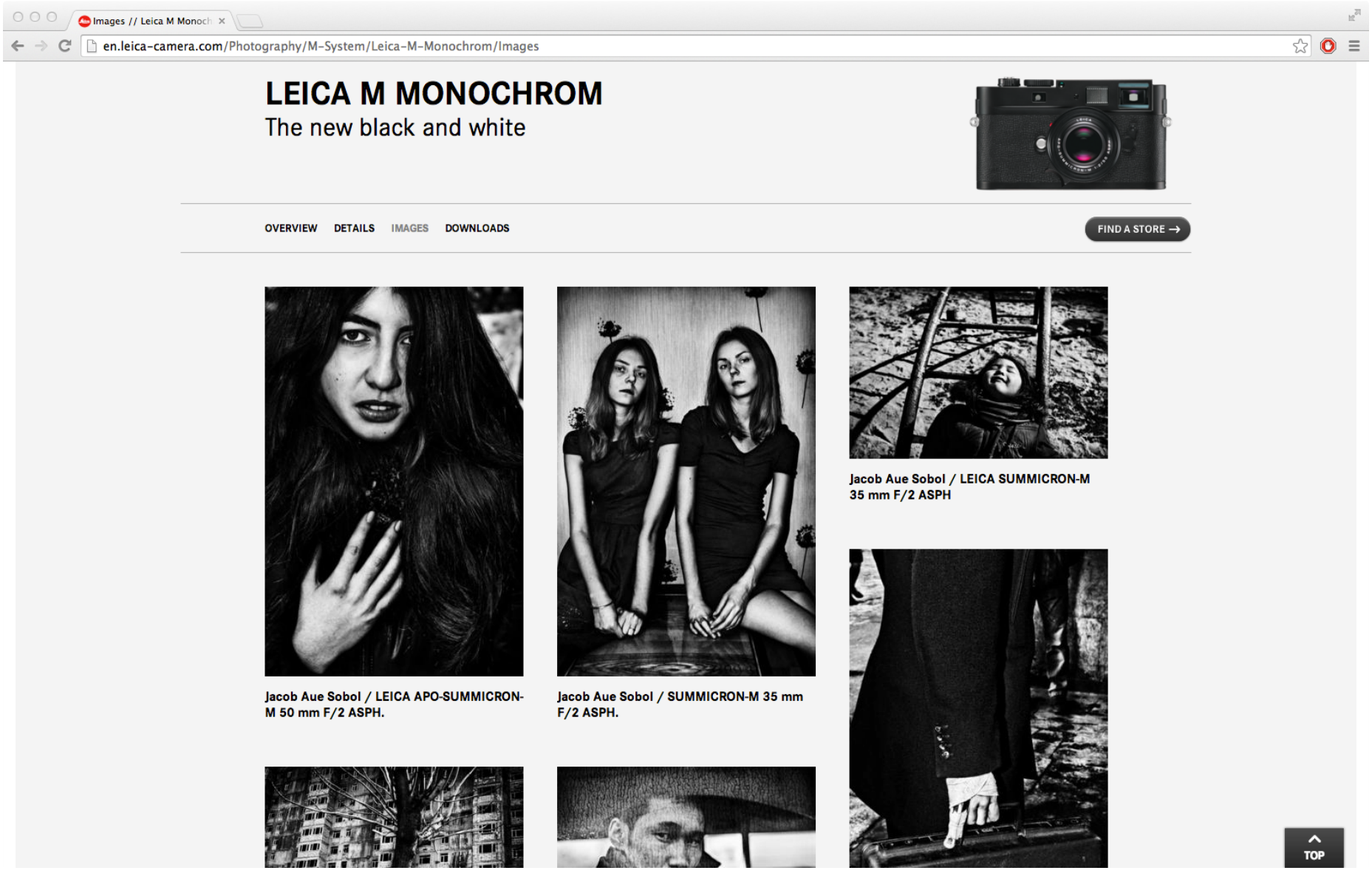
- good news: substantial steganalysis performance boosts hinge on sufficient information about CFA interpolation function
- bad news: image forensics research
- good news: counter-forensics research
- join forces?

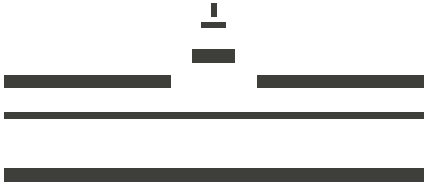
- come up with plausible communication channels for grayscale images?

Plausible Grayscale Images?



Plausible Grayscale Images?





Westfälische
Wilhelms-Universität
Münster

Steganalysis in Technicolor*

Boosting WS Detection of Stego Images from CFA-Interpolated Covers

Thanks for your attention!

*Similarities with existing companies and trademarks are accidental.