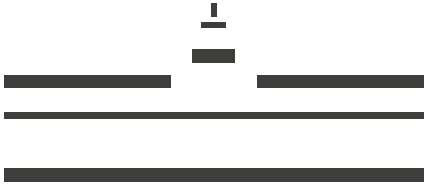


Westfälische
Wilhelms-Universität
Münster

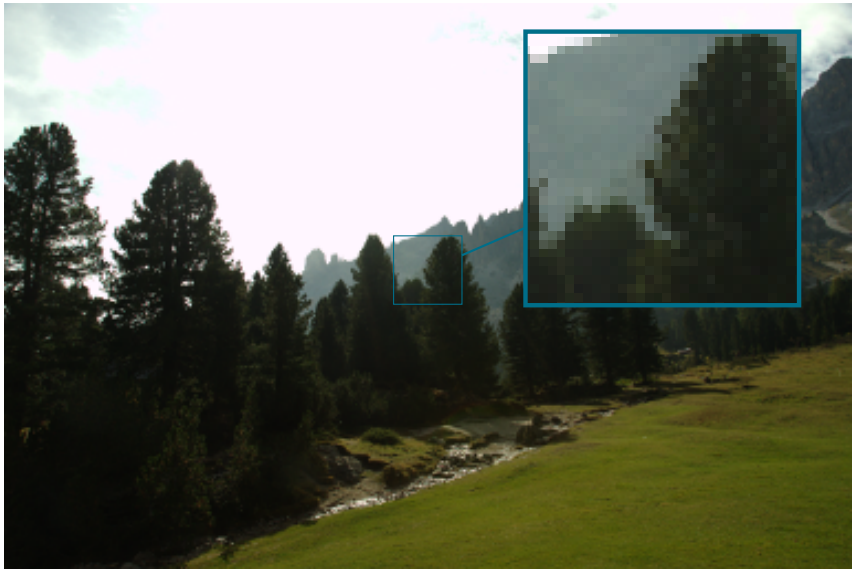
Forensics of High-Quality JPEG Images with Color Subsampling

Matthias Carnein, Pascal Schöttle, Rainer Böhme

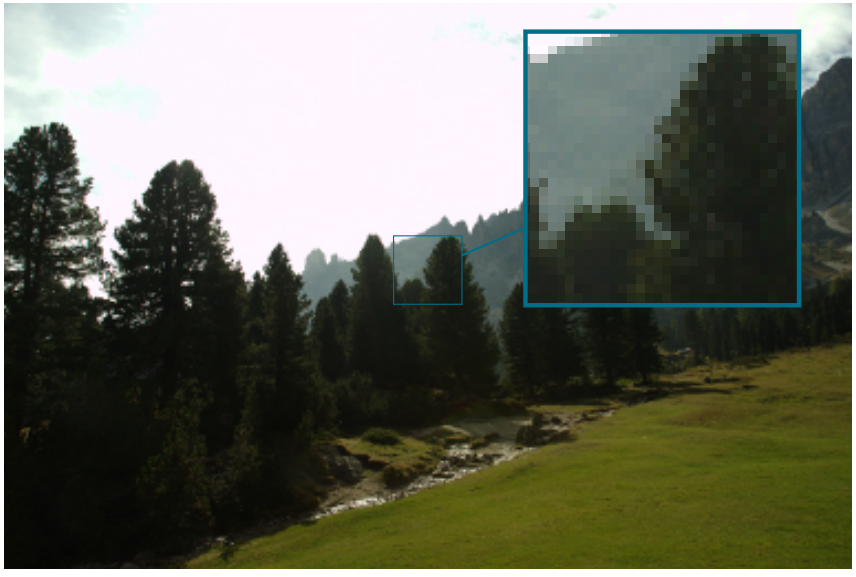
19th November 2015

Introduction

uncompressed



$Q = 100$



- Block convergence analysis can detect high quality JPEG compression
- Goal: Extend to color images

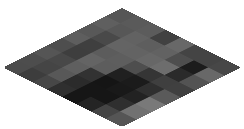
Image source: Dang-Nguyen et al. (2015)

Outline

1. Block Convergence
2. Extension To Color
3. Experimental Results
4. Conclusion

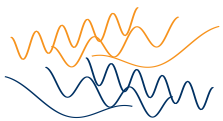
Outline

Grayscale JPEG



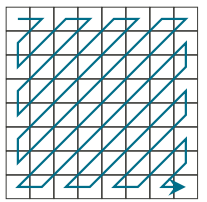
Block Splitting

Discrete Cosine Transformation



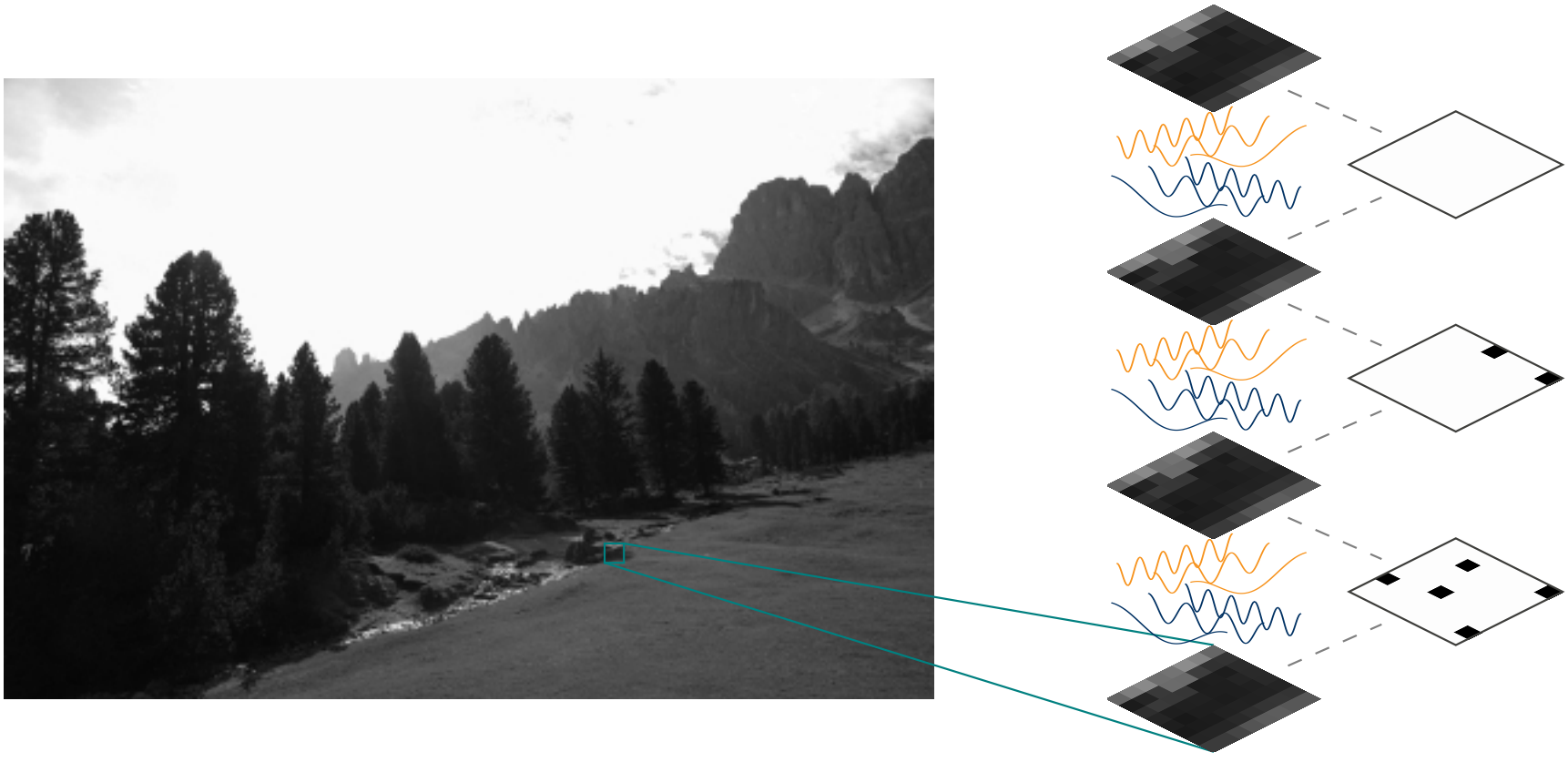
$$\left[\frac{y_{i,j}}{q_{i,j}}\right]$$

Quantization



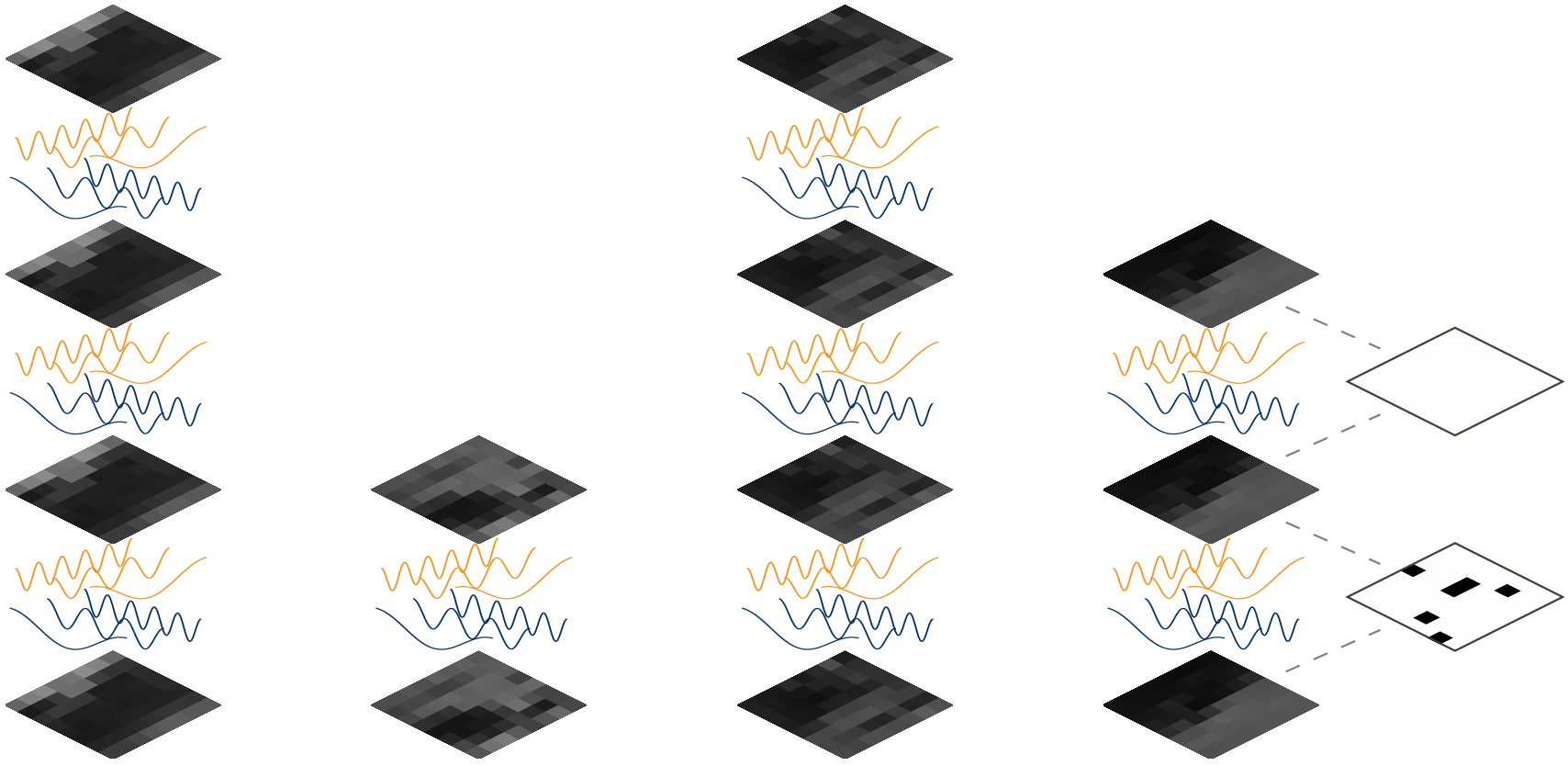
Entropy Coding

Block Convergence



Lai and Böhme (2013)

Block Convergence

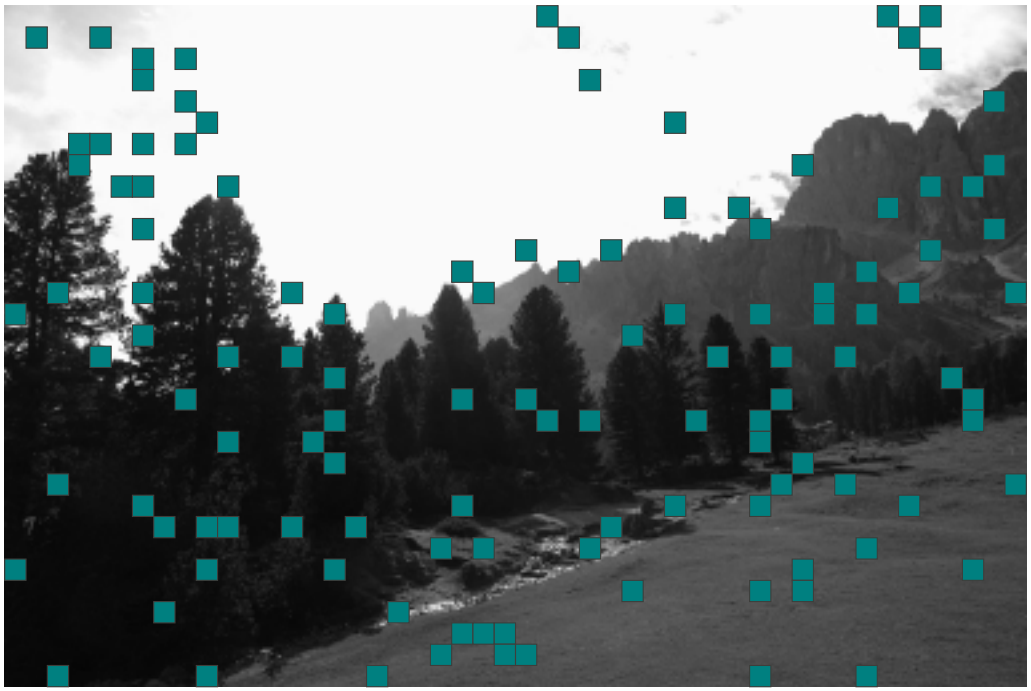


stable in $t = 2$ $t = 0$ $t = 2$ $t = 1$

Lai and Böhme (2013)

Ratio of Stable Blocks

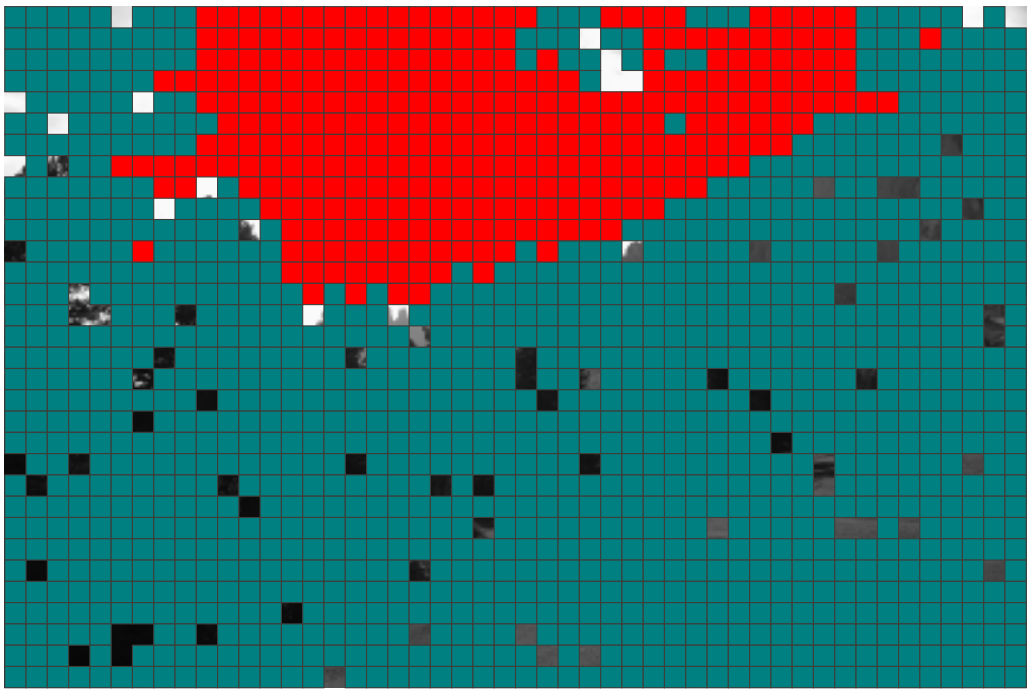
- Number of compressions: 1



Lai and Böhme (2013)

Ratio of Stable Blocks

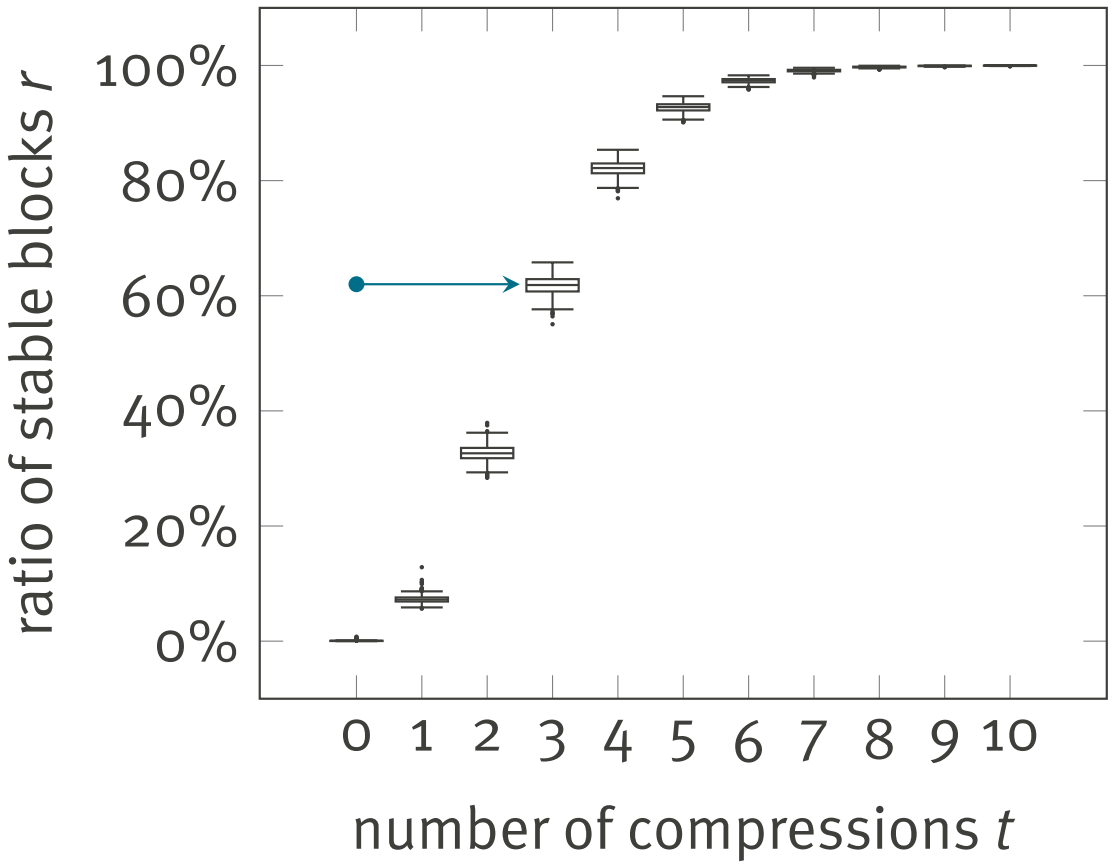
- Number of compressions: 5



- Ratio of stable blocks: $r = (\#stable - \#flat)/(\#total - \#flat)$

Lai and Böhme (2013)

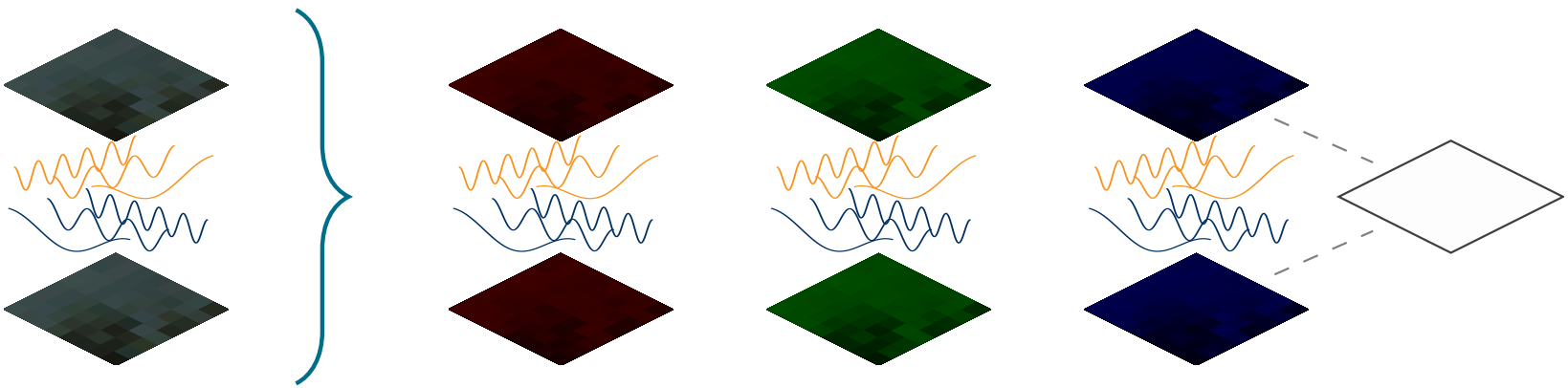
Distribution



Outline

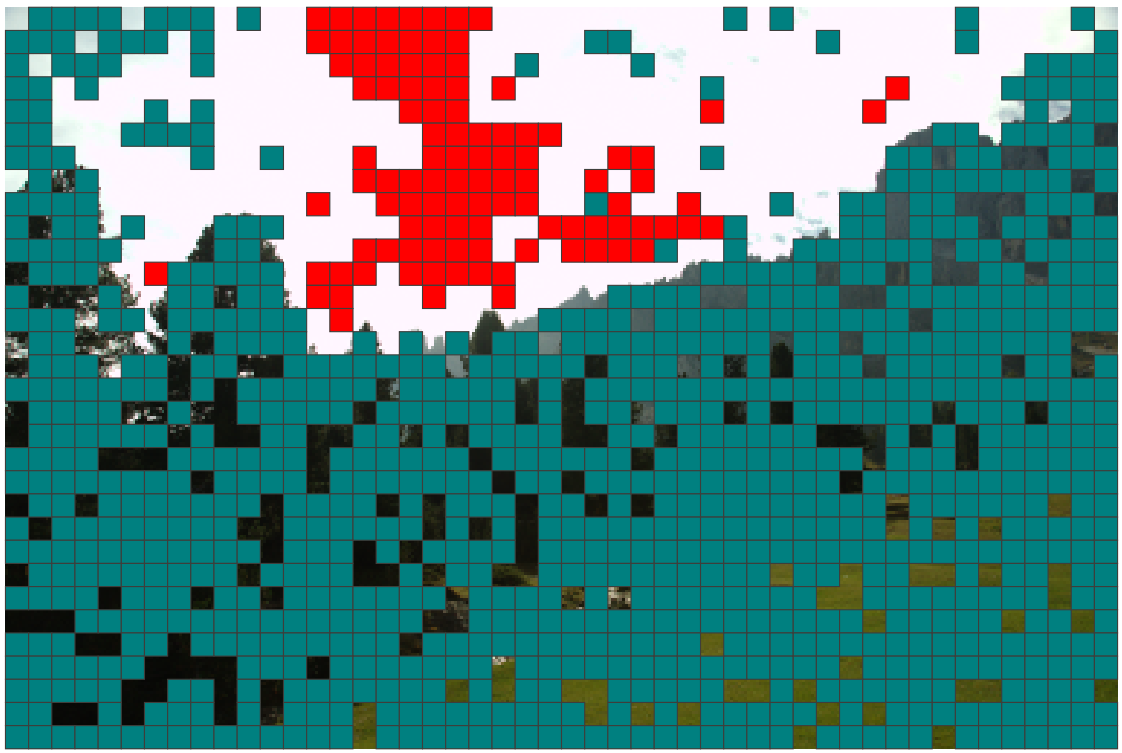
Block Convergence

- block is stable if no channel changes



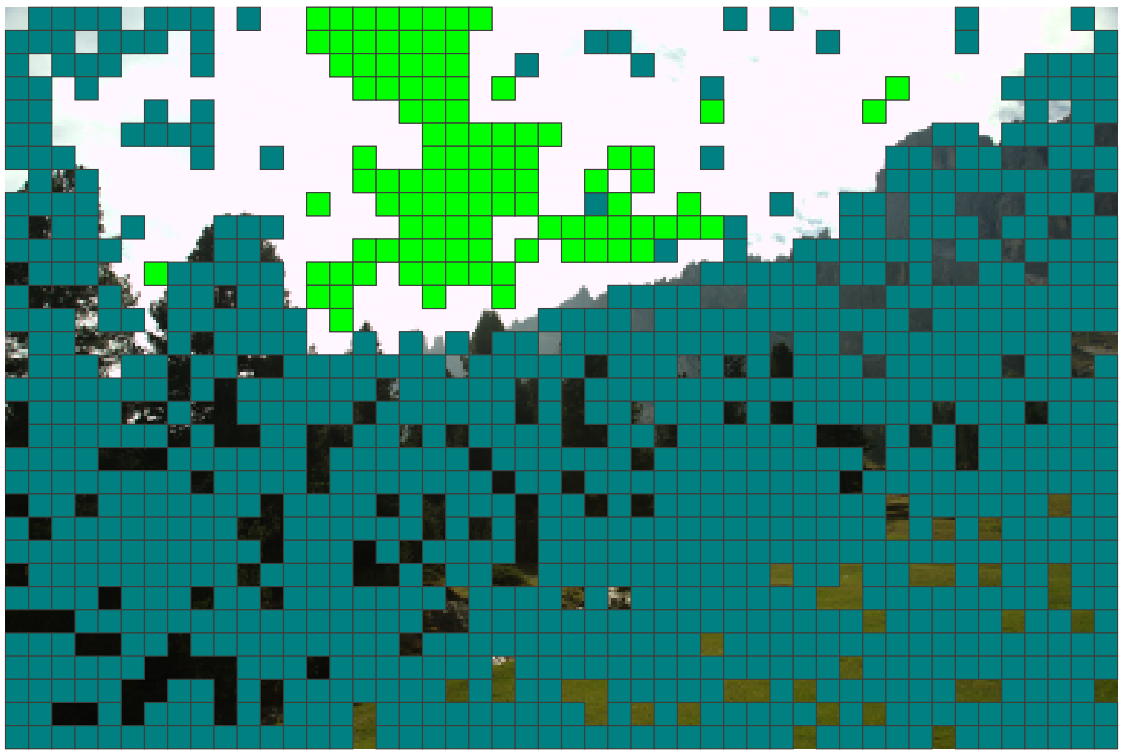
Ratio of Stable Blocks

- Number of compressions: 5



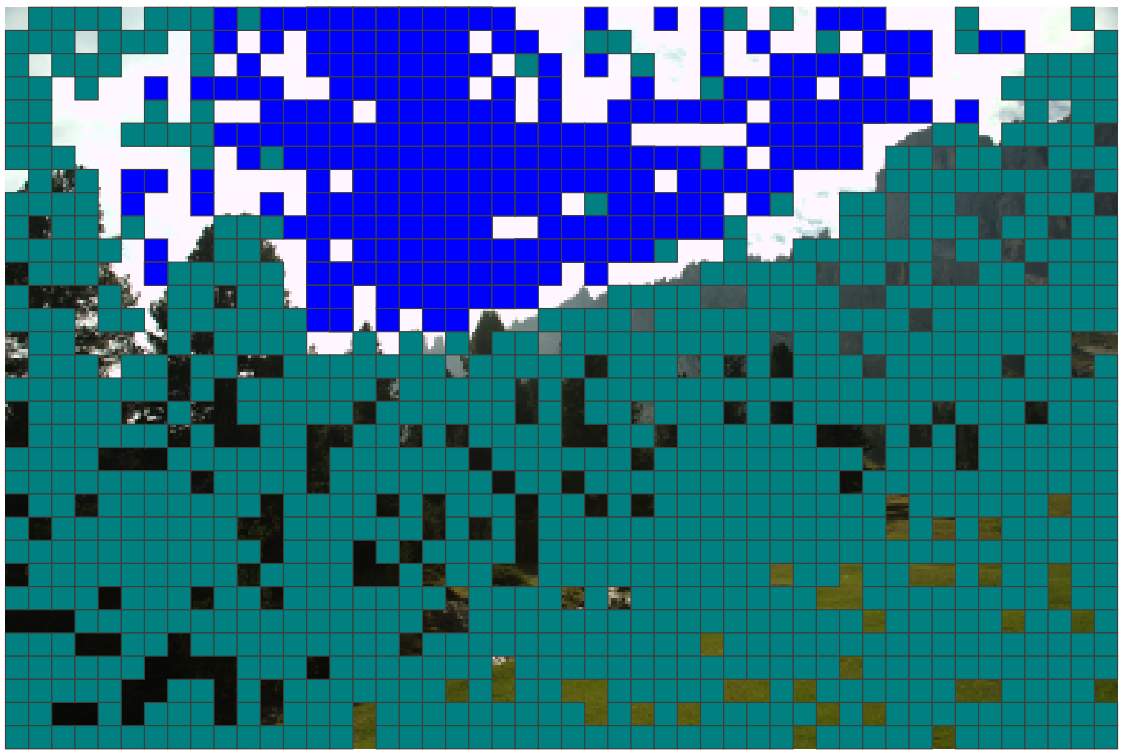
Ratio of Stable Blocks

- Number of compressions: 5

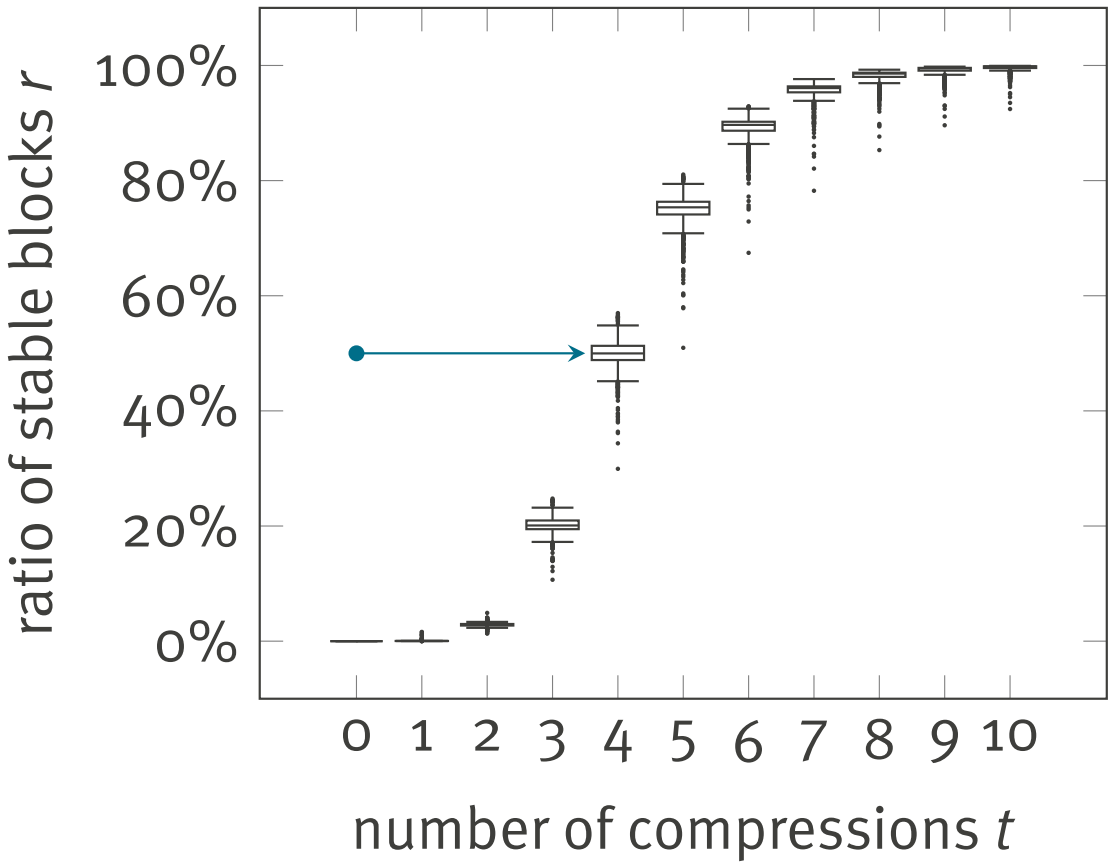


Ratio of Stable Blocks

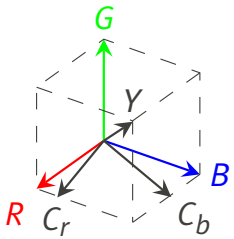
- Number of compressions: 5



Distribution

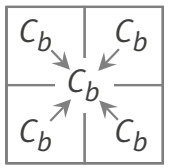


Color JPEG



Color Conversion

Subsampling



Block Splitting

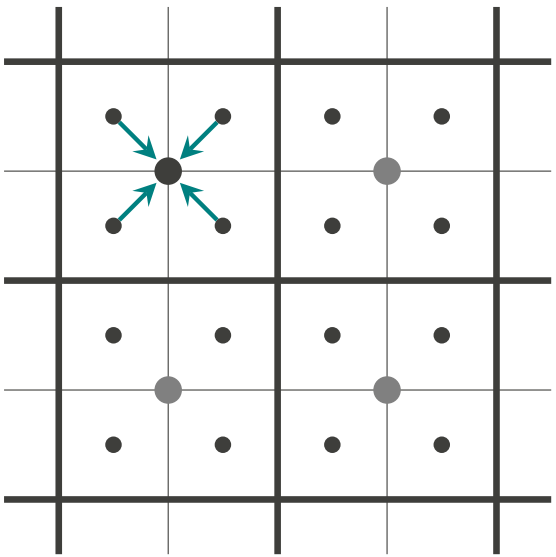
Discrete Cosine Transformation

Quantization

Entropy Coding

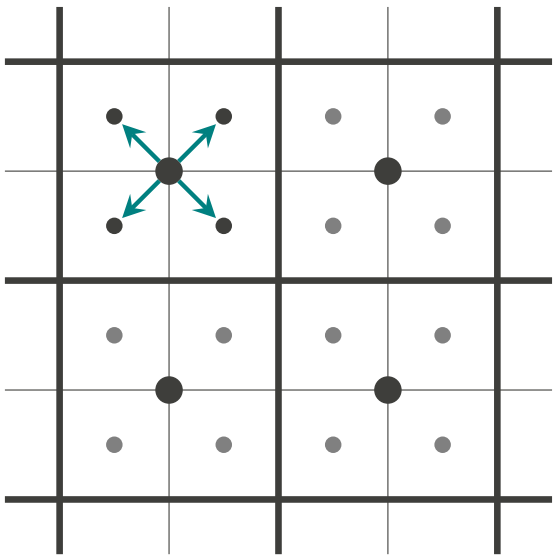
Simple Sub- and Upsampling

Subsampling



1/4

Upsampling

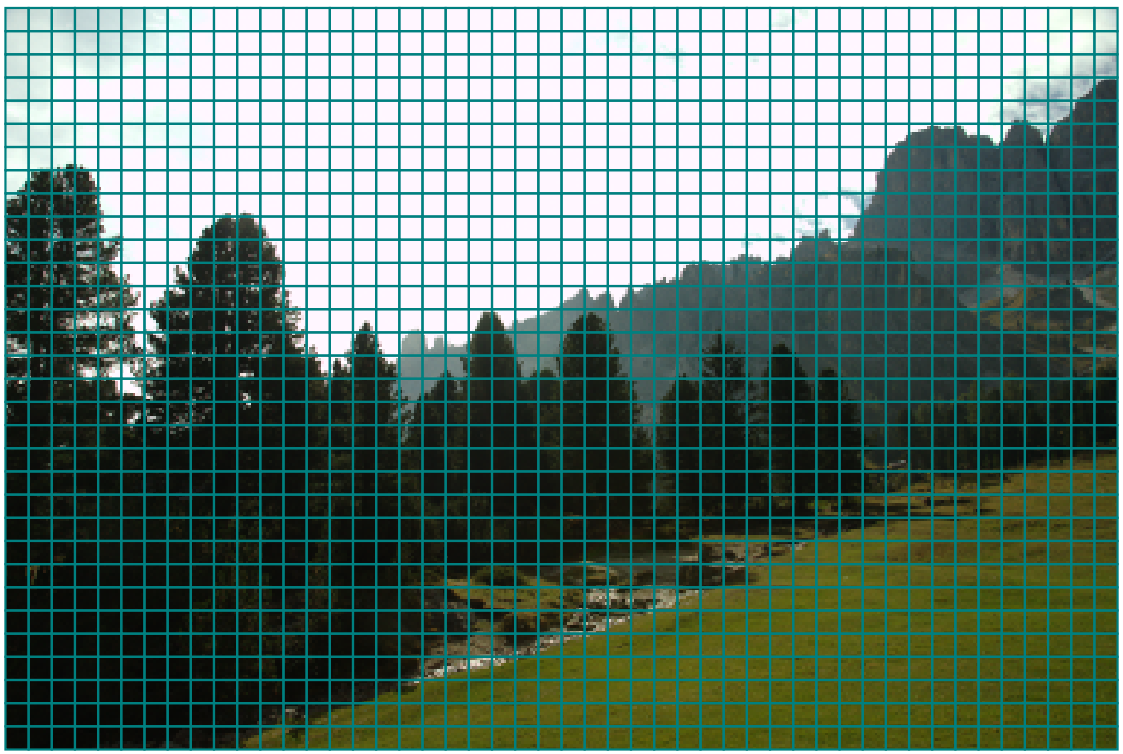


1

Independent JPEG Group (2015)

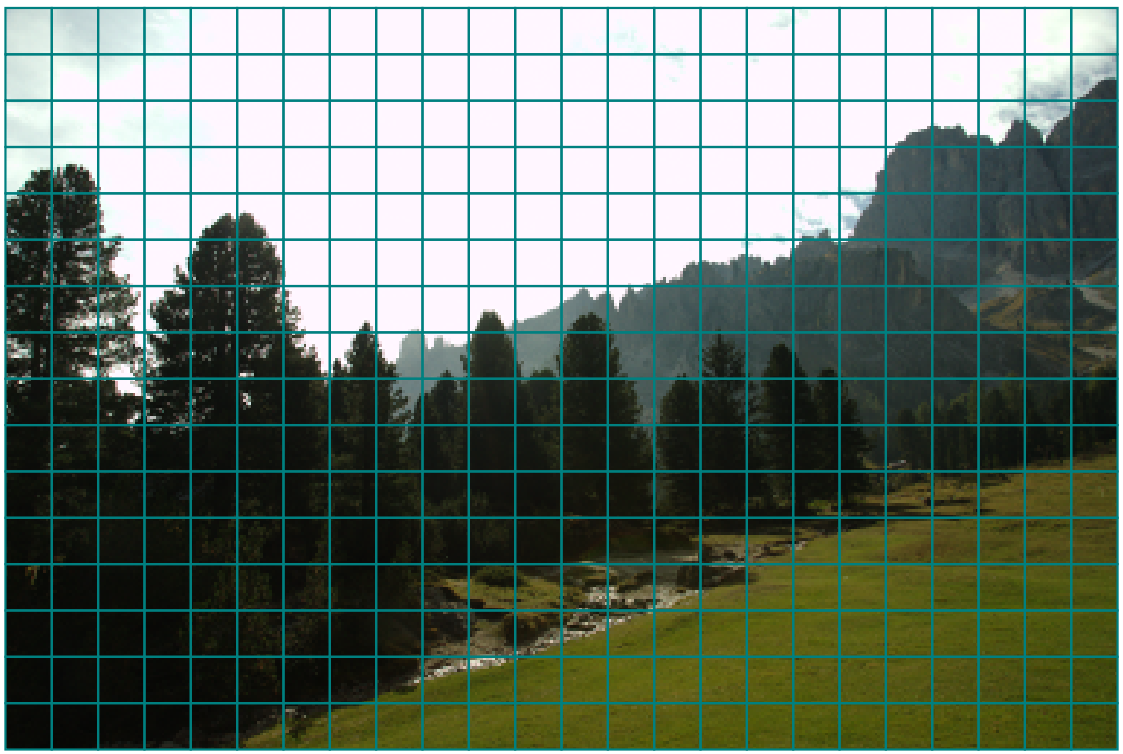
Macro-Blocks

- Observe convergence for area covered by chrominance block

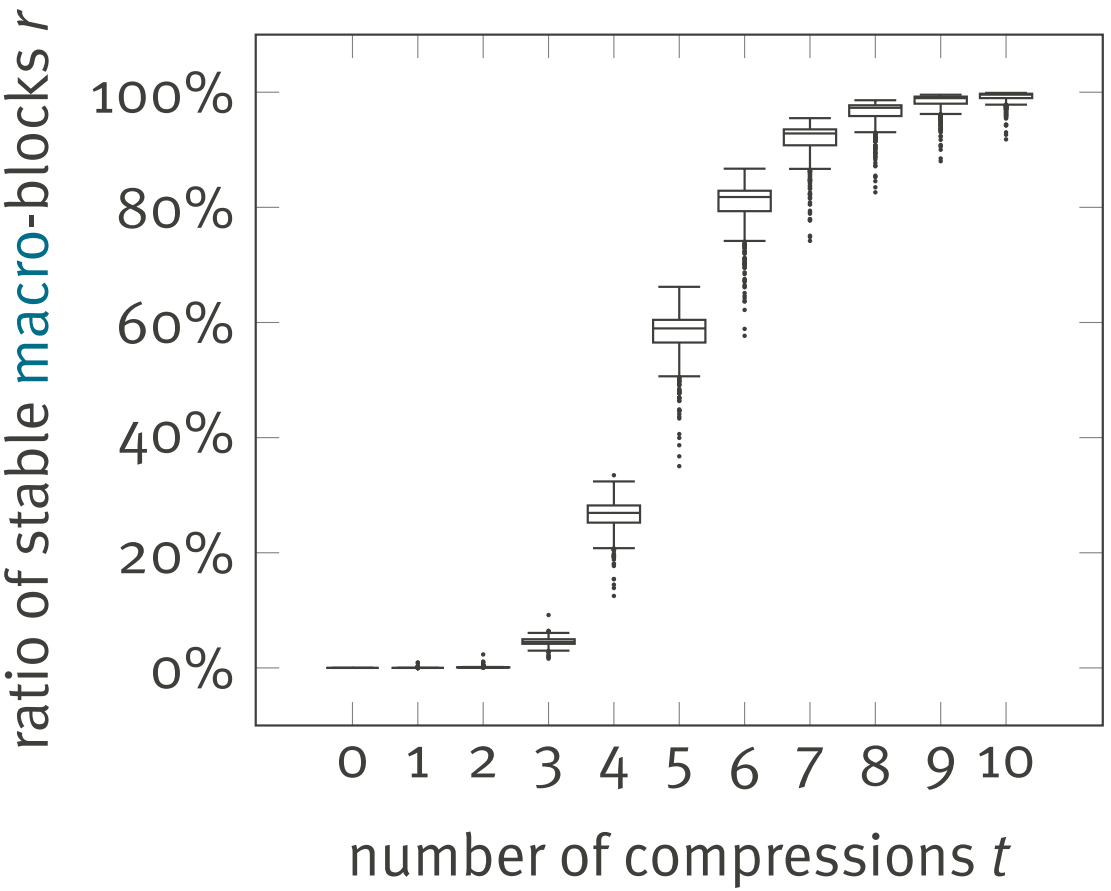


Macro-Blocks

- Observe convergence for area covered by chrominance block

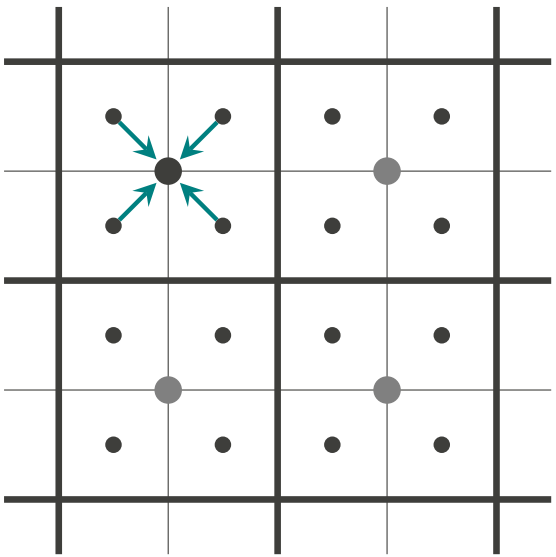


Distribution

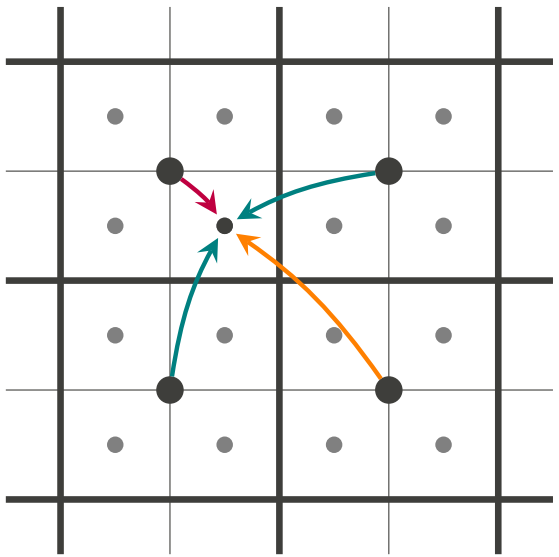


Fancy Upsampling

Subsampling



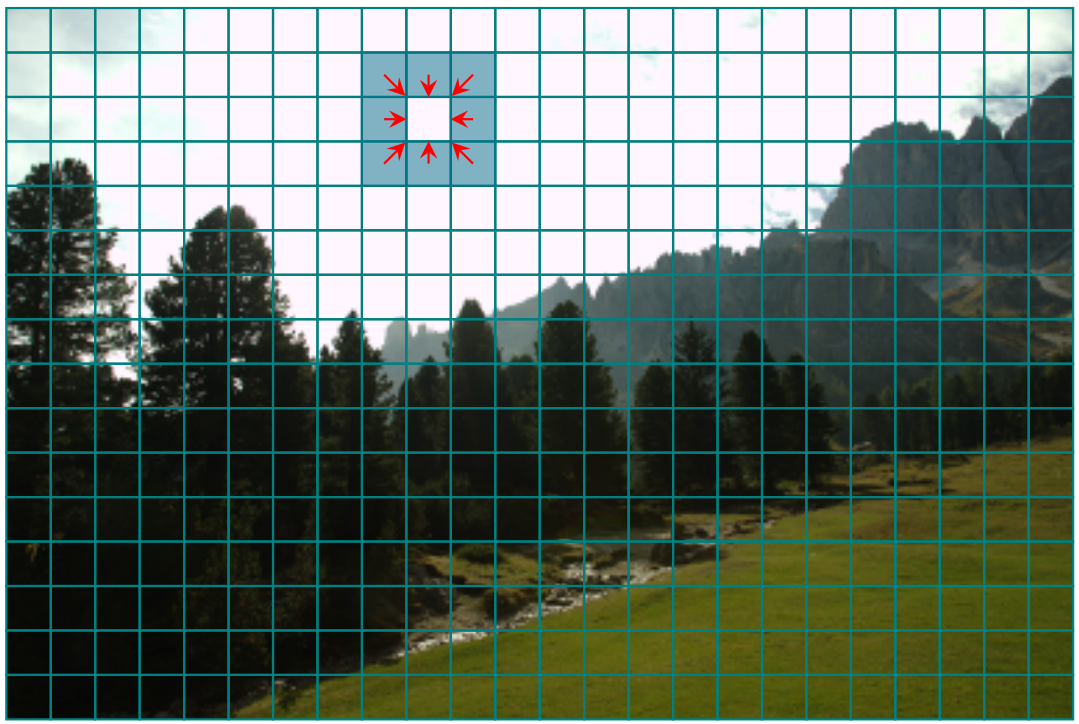
Upsampling



Independent JPEG Group (2015)

Spill-Over

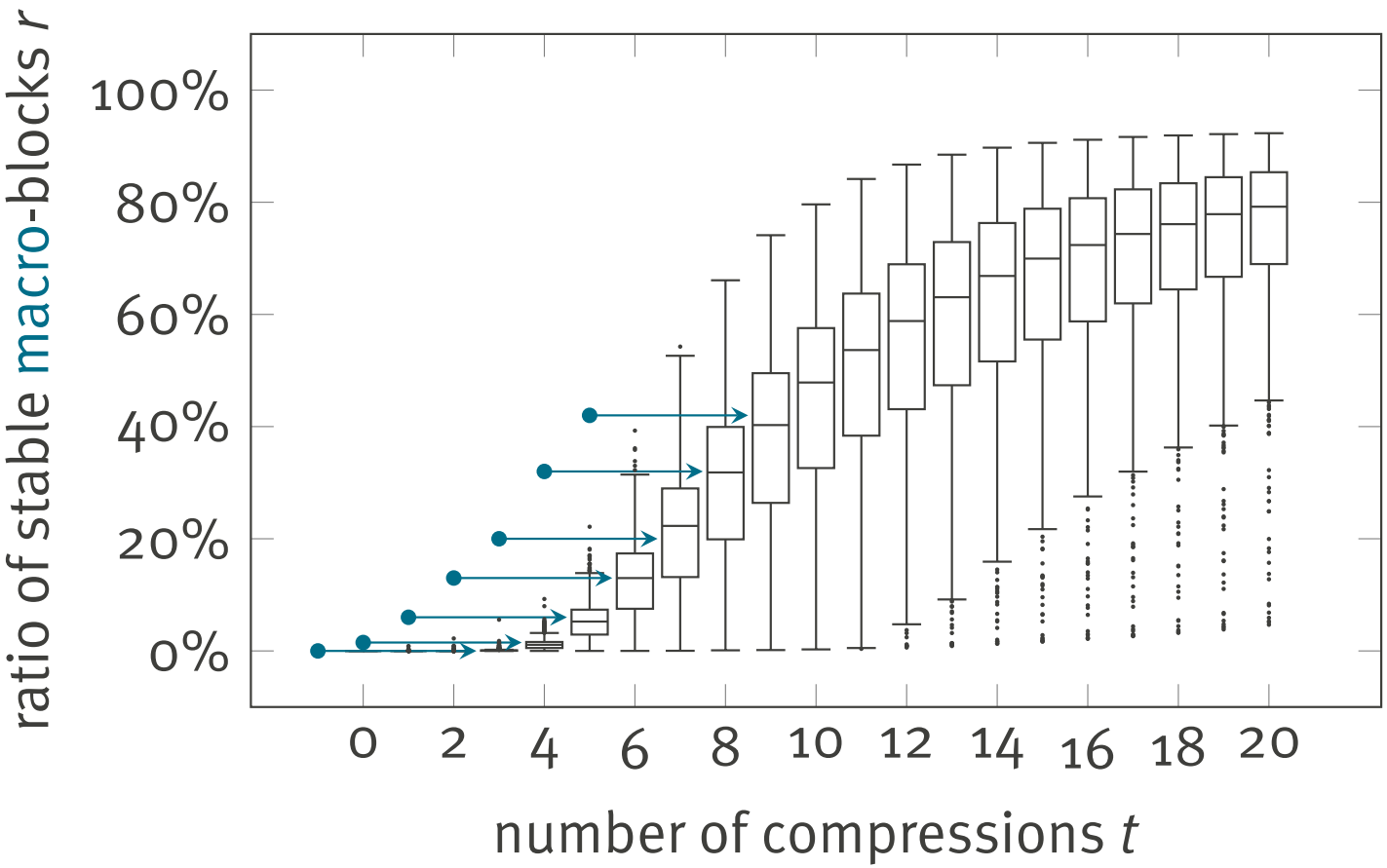
- fancy upsampling not confined within a macro-block
- ignore spill-over: stable if block does not change with compression



Spill-Over

- fancy upsampling not confined within a macro-block
- ignore spill-over: stable if block does not change with compression
- prevent spill-over: compress each block individually

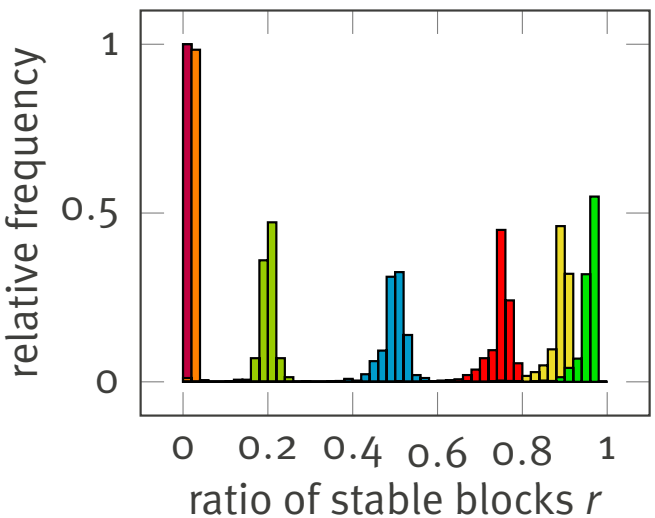
Distribution



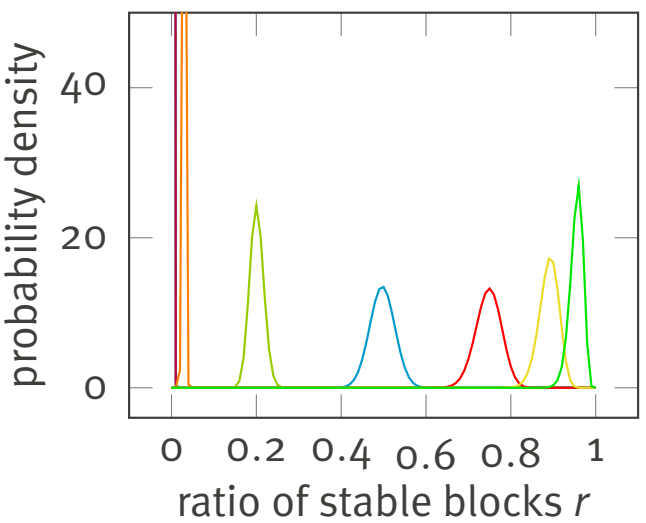
Multiple Recompressions

- Track ratio over multiple recompressions: $r_0, r_1, \ldots, r_n$
- Option 1: Machine Learning
  - use as features to train machine learning algorithm
- Option 2: Maximum-Likelihood

Empirical Distributions



Fitted Beta Distributions



Outline

Experimental Setup

- 1600 images, up to 8 compressions
- results reported as confusion matrices
  - true detection rate
  - off-by-two ratio

Detector output $\hat{t}$	Number of compressions t								
	0	1	2	3	4	5	6	7	8
0	■		■	■	■	■	■	■	■
1		■		■	■	■	■	■	■
2	■		■		■	■	■	■	■
3	■	■		■		■	■	■	■
4	■	■	■		■		■	■	■
5	■	■	■	■		■		■	■
6	■	■	■	■	■		■		■
7	■	■	■	■	■	■		■	
8	■	■	■	■	■	■	■		■
9	■	■	■	■	■	■	■	■	
$\geq$ 10	■	■	■	■	■	■	■	■	■

Detection Accuracy

	subsampling			Random Forest		Maximum-likelihood	
Q	rate	algorithm	DCT	true detection	off-by-two	true detection	off-by-two
100	4:4:4	–	'slow'	98.61%	10.00%	88.36%	7.64%
100	4:4:4	–	'float'	96.35%	9.89%	85.88%	20.45%
100	4:2:0	simple	'slow'	98.75%	4.44%	85.42%	6.10%
100	4:2:0	fancy (ignore)	'slow'	94.12%	15.84%	41.81%	44.84%
100	4:2:0	fancy (prevent)	'slow'	90.06%	23.04%	34.71%	60.46%
100	4:2:0	DCT scaling	'slow'	98.72%	2.17%	93.86%	4.07%
99	4:4:4	–	'slow'	98.83%	2.38%	90.21%	6.52%
95	4:4:4	–	'slow'	88.62%	32.60%	64.54%	42.11%
90	4:4:4	–	'slow'	69.31%	57.15%	29.19%	66.91%
75	4:4:4	–	'slow'	49.06%	72.25%	23.67%	79.86%
50	4:4:4	–	'slow'	35.69%	82.83%	33.39%	54.86%

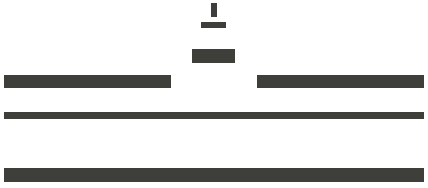
Confusion Matrix

Detector output $\hat{t}$	Number of compressions t								
	0	1	2	3	4	5	6	7	8
0	800	1	0	0	0	0	0	0	0
1	0	799	2	0	0	0	0	0	0
2	0	0	798	2	0	0	0	0	0
3	0	0	0	798	2	1	0	0	0
4	0	0	0	0	798	4	1	0	0
5	0	0	0	0	0	795	7	2	0
6	0	0	0	0	0	0	790	14	2
7	0	0	0	0	0	0	2	776	30
8	0	0	0	0	0	0	0	8	746
9	0	0	0	0	0	0	0	0	18
≥ 10	0	0	0	0	0	0	0	0	4

Outline

Conclusion

- Block Convergence: use ratio of converged blocks to estimate number of JPEG compressions
- Extension to Color
  - observe block convergence for macro-blocks
  - adapt to sub- and upsampling algorithm
  - track ratio over multiple recompressions
  - use maximum-likelihood or machine learning
- Outlook
  - Drawback: assumes repeated compression with the same settings
  - apply to other lossy compression formats



Westfälische
Wilhelms-Universität
Münster

Forensics of High-Quality JPEG Images with Color Subsampling

Thank you for your attention!

19th November 2015

References

Dang-Nguyen, Duc-Tien, Cecilia Pasquini, Valentina Conotter and Giulia Boato (2015). 'RAISE: A Raw Images Dataset for Digital Image Forensics'. In: *Proceedings of the 6th ACM Multimedia Systems Conference*. MMSys '15. Portland, Oregon: ACM, pp. 219–224. URL: `http://mmlab.science.unitn.it/RAISE/` (visited on 17/11/2015).

Independent JPEG Group (2015). *libjpeg*. URL: `http://www.ijg.org/` (visited on 17/11/2015).

Lai, ShiYue and Rainer Böhme (2013). 'Block convergence in repeated transform coding: JPEG-100 forensics, carbon dating, and tamper detection'. In: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3028–3032.