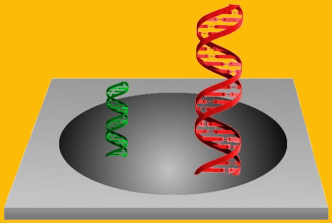
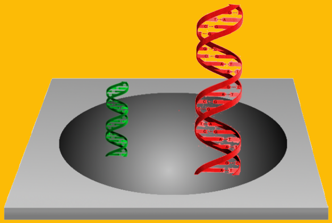
DNA Microarray Image Compression



Miguel Hernández Cabronero

June 19, 2015

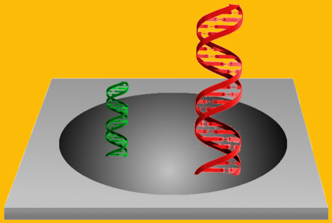
DNA Microarray Image Compression



Miguel Hernández Cabronero

June 19, 2015

DNA Microarray Image Compression



Miguel Hernández Cabronero

June 19, 2015

Supervisor:
Dr. Joan Serra-Sagristà
Co-supervisor:
Dr. Michael W. Marcellin

Committee:
Dr. Manuel Perez Malumbres
Dr. Cristina Fernández Córdoba
Dr. António José Ribeiro Neves

Universitat Autònoma de Barcelona

About this thesis...

Financial support:

- Spanish Government
  - FPU 2010
  - EST13/00180
  - TIN2012-38102-C03-00
- Catalan Government
  - 2014 SGR 691
  - SGR2009-1224
- Santander Bank

About this thesis...

Financial support:

- Spanish Government
  - FPU 2010
  - EST13/00180
  - TIN2012-38102-C03-00
- Catalan Government
  - 2014 SGR 691
  - SGR2009-1224
- Santander Bank

Development:

2010 BSc Computer Science
BScMathematics

2011 MSc Computer Science

2011 Thesis beginning

2014 Stay in Tucson, AZ, USA

2015 Stay in Aveiro, Portugal

Contents

1. Introduction: DNA Microarrays

Contents

1. Introduction: DNA Microarrays
2. Lossless compression
3. Lossy compression

Contents

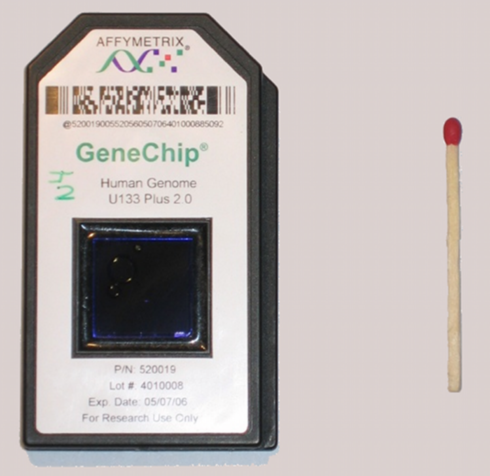
1. Introduction: DNA Microarrays
2. Lossless compression
3. Lossy compression
4. Conclusions

Contents

1. Introduction: DNA Microarrays
2. Lossless compression
3. Lossy compression
4. Conclusions

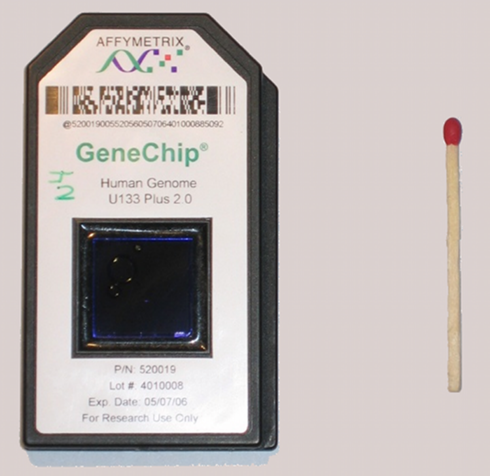
DNA Microarrays

DNA Microarray (match for scale)



DNA Microarrays

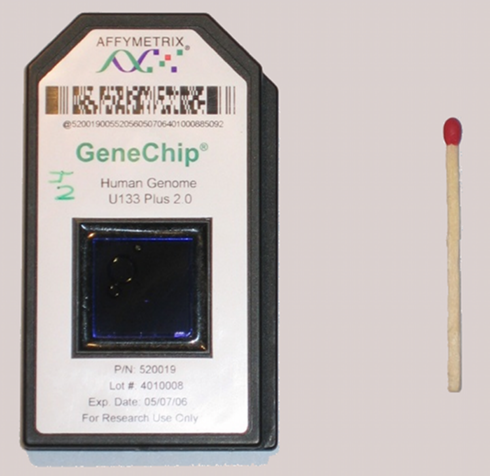
DNA Microarray (match for scale)



- Quantify genetic expression

DNA Microarrays

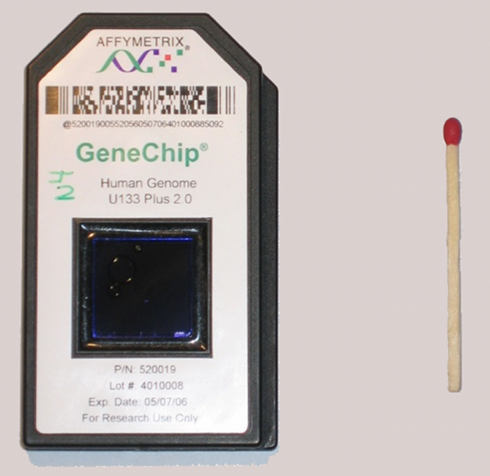
DNA Microarray (match for scale)



- Quantify genetic expression
- Analyze genetic function and regulation

DNA Microarrays

DNA Microarray (match for scale)



- Quantify genetic expression
- Analyze genetic function and regulation

Some applications:

$\Rightarrow$ Understanding Cancer, HIV, Malaria...

$\Rightarrow$ Drug development

$\Rightarrow$ Evolutionary biology

DNA Microarrays

History:

1975 Southern Blot:
genetic sequence detection

1982 Augentlich:
378 in parallel

1987 Augentlich:
4000 human genes

DNA Microarrays

History:

1975 Southern Blot:
genetic sequence detection

1982 Augentlich:
378 in parallel

1987 Augentlich:
4000 human genes

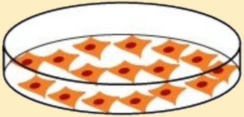
1995 Schena:
Miniaturized chips

1997 Lashkari:
Whole eukaryiotic genome
(∼6000 genes)

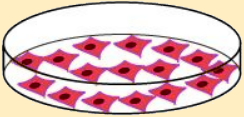
2005 Affymetrix:
Whole human genome
(∼20000 genes)

DNA Microarrays

Biological samples



Tumoral tissue



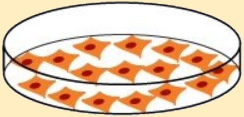
Healthy tissue

Underlying principle:

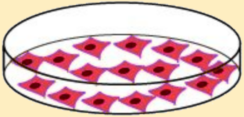
- Two samples

DNA Microarrays

Biological samples



Tumoral tissue



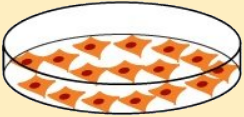
Healthy tissue

Underlying principle:

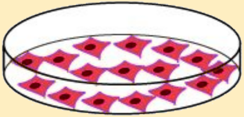
- Two samples
- Same genes

DNA Microarrays

Biological samples



Tumoral tissue



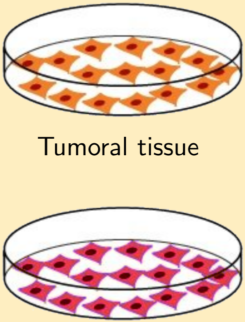
Healthy tissue

Underlying principle:

- Two samples
- Same genes
- Different expression intensity

DNA Microarrays

Biological samples



Tumoral tissue

Healthy tissue

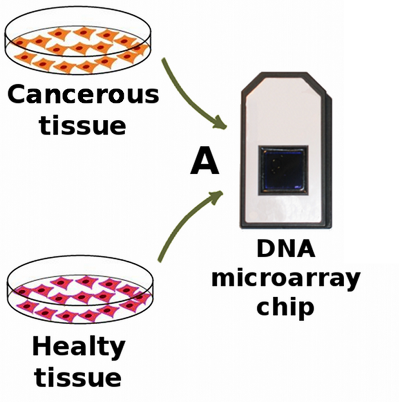
Underlying principle:

- Two samples
- Same genes
- Different expression intensity

Compare expression $\Rightarrow$ understand function

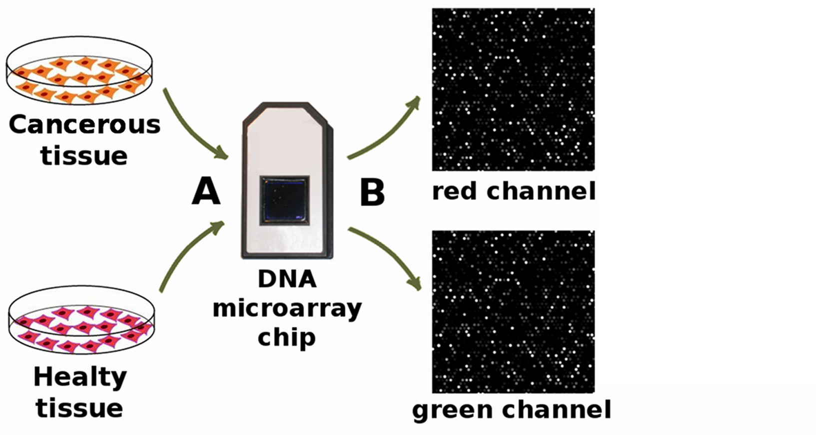
DNA Microarray Experiments

How to obtain the genetic data:



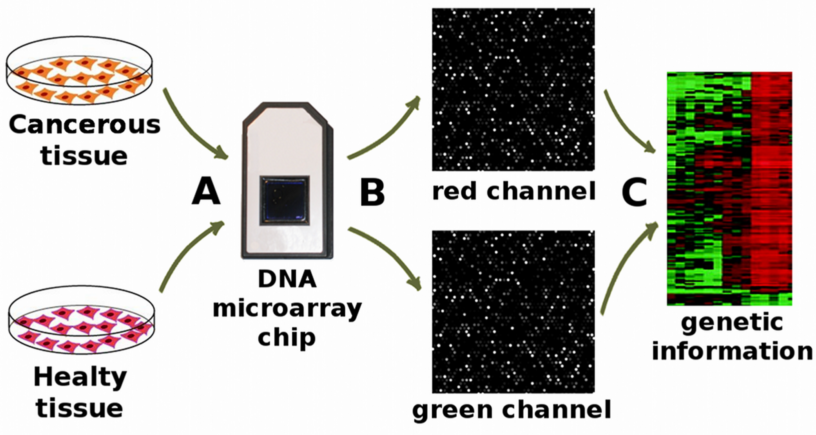
DNA Microarray Experiments

How to obtain the genetic data:



DNA Microarray Experiments

How to obtain the genetic data:



DNA Microarray Experiments

Example of genetic data output:

Gene	Expression intensity Sample 1 μ_1	Expression intensity Sample 2 μ_1	Intensity ratio μ_1/μ_2
Gene 1	245	249	0.98
Gene 2	54	13	4.15
Gene 3	200	1530	0.13
⋮	⋮	⋮	⋮
Gene N	100	103	0.97

DNA Microarray Experiments

Example of genetic data output:

Gene	Expression intensity Sample 1 μ_1	Expression intensity Sample 2 μ_1	Intensity ratio μ_1/μ_2
Gene 1	245	249	**0.98**
Gene 2	54	13	4.15
Gene 3	200	1530	0.13
⋮	⋮	⋮	⋮
Gene N	100	103	0.97

DNA Microarray Experiments

Example of genetic data output:

Gene	Expression intensity Sample 1	Sample 2	Intensity ratio
	μ_1	μ_1	μ_1/μ_2
Gene 1	245	249	0.98
Gene 2	54	13	**4.15**
Gene 3	200	1530	0.13
⋮	⋮	⋮	⋮
Gene N	100	103	0.97

DNA Microarray Experiments

Example of genetic data output:

Gene	Expression intensity Sample 1	Expression intensity Sample 2	Intensity ratio
	μ_1	μ_1	μ_1/μ_2
Gene 1	245	249	0.98
Gene 2	54	13	4.15
Gene 3	200	1530	**0.13**
⋮	⋮	⋮	⋮
Gene N	100	103	0.97

Why Compression?

Why compress DNA microarray images?

Why Compression?

New image analysis methods

Why Compression?

New image analysis methods

- Athanasiadis, 2015
- Fouad, 2015
- Li, 2014
- El-Gawady, 2014

Why Compression?

New image analysis methods $\Rightarrow$ better genetic data

- Athanasiadis, 2015
- Fouad, 2015
- Li, 2014
- El-Gawady, 2014

Why Compression?

Want to re-analyze

Why Compression?

Want to re-analyze

but

Cannot repeat
experiments

Why Compression?

Want to re-analyze

but

Cannot repeat experiments

$\Longrightarrow$ Must keep images

Why Compression?

Want to re-analyze

but $\Longrightarrow$ Must keep images $\Longrightarrow$ Need efficient storage and transmission

Cannot repeat experiments

Why Compression?

Want to re-analyze

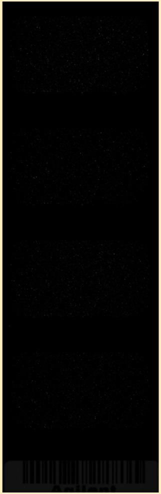
but $\Longrightarrow$ Must keep images $\Longrightarrow$ Need efficient storage and transmission

Cannot repeat experiments

Image compression: natural approach

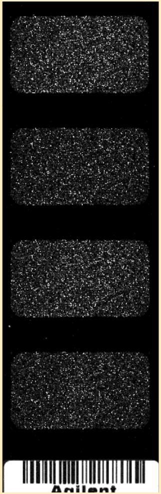
DNA Microarray Images

DNA microarray image



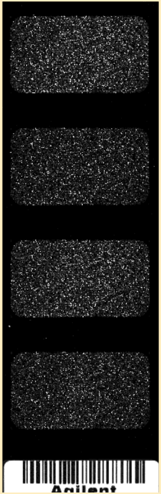
DNA Microarray Images

DNA microarray image (brighter)



DNA Microarray Images

DNA microarray image (brighter)

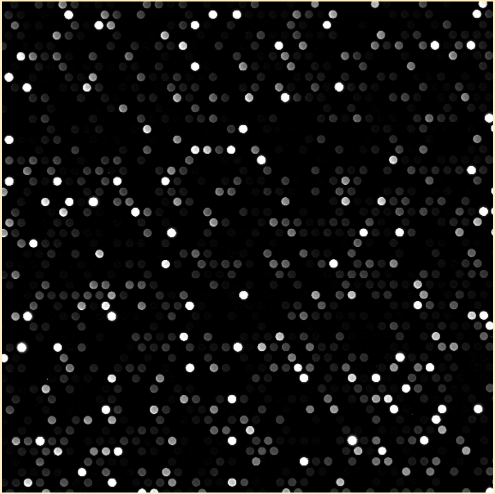


Key properties

- Large dimensions (4400×13800)

DNA Microarray Images

DNA microarray image (crop)

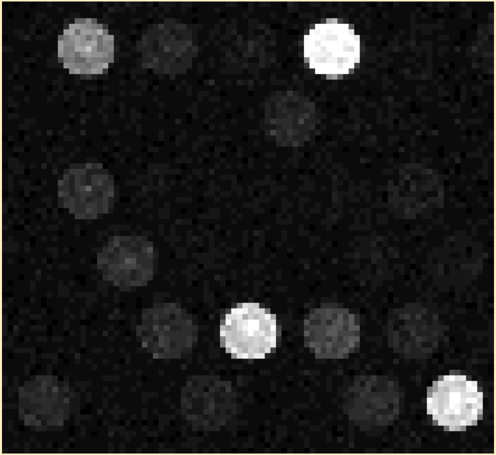


Key properties

- Large dimensions (4400×13800)
- Abrupt changes

DNA Microarray Images

DNA microarray image (detail)



Key properties

- Large dimensions (4400×13800)
- Abrupt changes

DNA Microarray Images

Natural image *vs* Microarray image

Natural images (8 bpp):
$2^8 = 256$ possible intensities

Microarray images (16 bpp):
$2^{16} = 65535$ possible intensities

Key properties

- Large dimensions (4400×13800)
- Abrupt changes
- 16 bits per pixel

DNA Microarray Images

Range of intensities

Natural images (8 bpp):
$2^8 = 256$ possible intensities

Microarray images (16 bpp):
$2^{16} = 65535$ possible intensities

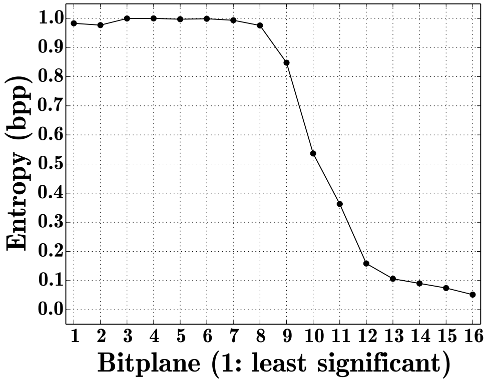
but only 30-80% are used

Key properties

- Large dimensions (4400×13800)
- Abrupt changes
- 16 bits per pixel
- Unused intensities: 20-70%

DNA Microarray Images

Range of intensities

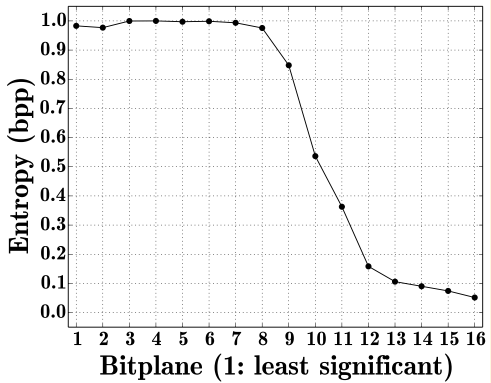


Key properties

- Large dimensions (4400×13800)
- Abrupt changes
- 16 bits per pixel
- Unused intensities: 20-70%
- Noisy bitplanes: up to 9

DNA Microarray Images

Bitplane entropy



Key properties

- Large dimensions (4400×13800)
- Abrupt changes
- 16 bits per pixel
- Unused intensities: 20-70%
- Noisy bitplanes: up to 9

Need adapted compressors

Thesis goals

Thesis **main goal**: efficient compression

Thesis goals

Thesis **main goal**: efficient compression

Lossless compression

Lossy compression

Thesis goals

Thesis **main goal**: efficient compression

Lossless compression

✓: Perfect data fidelity

Lossy compression

Thesis goals

Thesis **main goal**: efficient compression

Lossless compression

- ✓: Perfect data fidelity
- ×: Small compression ratios
- ×: Not standard

Lossy compression

Thesis goals

Thesis **main goal**: efficient compression

Lossless compression

- ✓: Perfect data fidelity
- ×: Small compression ratios
- ×: Not standard

Lossy compression

- ✓: Arbitrary compression ratios

Thesis goals

Thesis **main goal**: efficient compression

Lossless compression

- ✓: Perfect data fidelity
- ×: Small compression ratios
- ×: Not standard

Lossy compression

- ✓: Arbitrary compression ratios
- ×: Analysis result distortion

Thesis goals

Thesis **main goal**: efficient compression

Lossless compression

- ✓: Perfect data fidelity
- ×: Small compression ratios
- ×: Not standard

Lossy compression

- ✓: Arbitrary compression ratios
- ×: Analysis result distortion

Thesis **contributions**: solutions to these problems

Thesis goals

Thesis **main goal**: efficient compression

Lossless compression

- ✓: Perfect data fidelity
- ×: Small compression ratios
- ×: Not standard

Lossy compression

- ✓: Arbitrary compression ratios
- ×: Analysis result distortion

Thesis **contributions**: solutions to these problems

⇒ Improve standards

Thesis goals

Thesis **main goal**: efficient compression

Lossless compression

✓: Perfect data fidelity

×: Small compression ratios

×: Not standard

Lossy compression

✓: Arbitrary compression ratios

×: Analysis result distortion

Thesis **contributions**: solutions to these problems

⇒ Improve standards

⇒ Assess distortion

Thesis goals

Thesis **main goal**: efficient compression

Lossless compression

- ✓: Perfect data fidelity
- ×: Small compression ratios
- ×: Not standard

Lossy compression

- ✓: Arbitrary compression ratios
- ×: Analysis result distortion

Thesis **contributions**: solutions to these problems

- ⇒ Improve standards

- ⇒ Assess distortion
- ⇒ Lossy coder with acceptable distortion

Benchmark Corpora

Image Sets

Name	Year	Images	Size
Yeast	1998	109	1024×1024
ApoA1	2001	32	1044×1041
ISREC	2001	14	1000×1000
Stanford	2001	20	2000×2000
MicroZip	2004	3	1800×1900
Omnibus	2006	25	12200×4320
Arizona	20011	6	4400×13800
IBB	2013	44	2019×6235

Benchmark Corpora

Image Sets

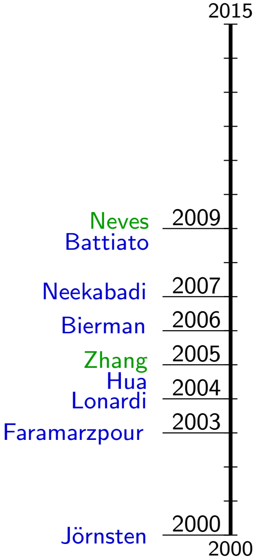
Name	Year	Images	Size
Yeast	1998	109	1024×1024
ApoA1	2001	32	1044×1041
ISREC	2001	14	1000×1000
Stanford	2001	20	2000×2000
MicroZip	2004	3	1800×1900
Omnibus	2006	25	12200×4320
Arizona	20011	6	4400×13800
IBB	2013	44	2019×6235

Representative of many scanner types

Contents

1. Introduction: DNA Microarrays
2. Lossless compression
3. Lossy compression
4. Conclusions

State of the art



Most algorithms:

- Segmentation-based coding
- Context-base coding

State of the art

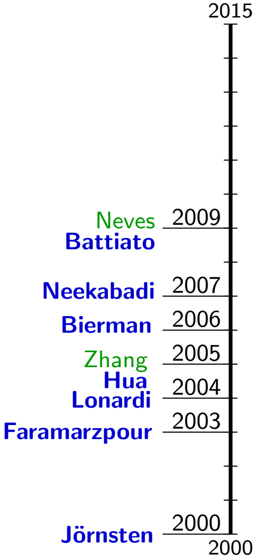
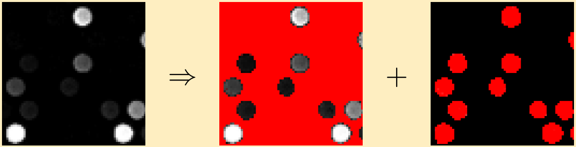


Image segmentation



State of the art

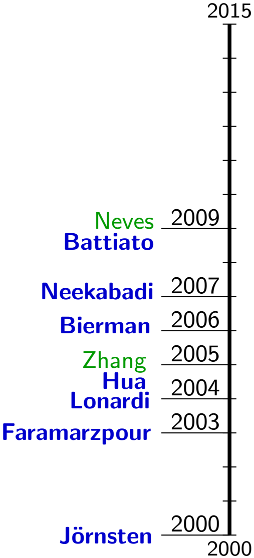
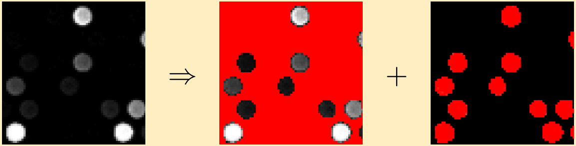
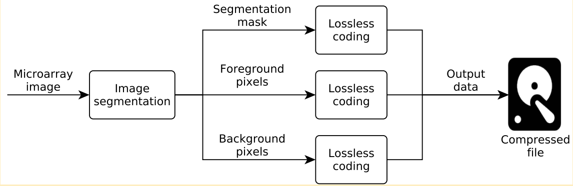


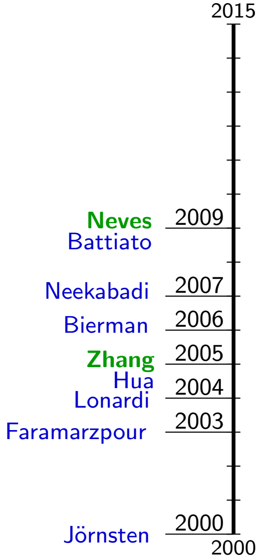
Image segmentation



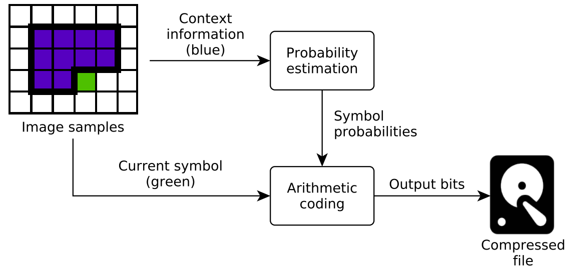
Segmentation-based coding



State of the art



Context-based coding



Standard *vs* microarray-specific coders

Algorithm	Yeast	ApoA1	ISREC	Stanford	MicroZip	Omnibus	AZ	IBB
Generic algorithms and standard image compressors								
Bzip2	6.075	11.067	10.921	7.867	9.394	7.523	8.944	9.081
CALIC	8.502	**10.515**	**10.615**	7.592	9.582	**6.929**	8.767	9.327
JBIG2	6.888	10.852	10.925	7.776	9.747	7.198	8.858	9.344
JPEG-LS	8.580	10.608	11.145	**7.571**	9.441	6.952	**8.646**	9.904
HEVC/H.265	10.660	14.482	14.876	8.897	11.179	8.350	10.592	12.262
JPEG2000	6.829	10.999	10.888	7.969	9.467	8.121	7.549	9.064
Microarray-specific techniques								
Jörnsten	8.556	—	—	—	—	—	—	—
Faramarzpour	9.091	—	—	—	—	—	—	—
Hua	6.985	—	—	—	—	—	—	—
Lonardi	—	—	—	—	9.843	—	—	—
Zhang	6.601	—	—	—	9.587	—	—	—
Neekabadi	—	10.250	10.202	—	8.856	—	—	—
Battiato	—	**9.520**	**9.490**	—	**8.369**	—	—	—
Neves	**5.521**	10.223	10.199	**7.335**	8.667	**7.743**	**8.275**	**8.039**

Standard *vs* microarray-specific coders

Algorithm	Yeast	ApoA1	ISREC	Stanford	MicroZip	Omnibus	AZ	IBB
Generic algorithms and standard image compressors								
Bzip2	6.075	11.067	10.921	7.867	9.394	7.523	8.944	9.081
CALIC	8.502	**10.515**	**10.615**	7.592	9.582	**6.929**	8.767	9.327
JBIG2	6.888	10.852	10.925	7.776	9.747	7.198	8.858	9.344
JPEG-LS	8.580	10.608	11.145	**7.571**	9.441	6.952	**8.646**	9.904
HEVC/H.265	10.660	14.482	14.876	8.897	11.179	8.350	10.592	12.262
JPEG2000	6.829	10.999	10.888	7.969	9.467	8.121	7.549	9.064
Microarray-specific techniques								
Jörnsten	8.556	—	—	—	—	—	—	—
Faramarzpour	9.091	—	—	—	—	—	—	—
Hua	6.985	—	—	—	—	—	—	—
Lonardi	—	—	—	—	9.843	—	—	—
Zhang	6.601	—	—	—	9.587	—	—	—
Neekabadi	—	10.250	10.202	—	8.856	—	—	—
Battiato	—	**9.520**	**9.490**	—	**8.369**	—	—	—
Neves	**5.521**	10.223	10.199	**7.335**	8.667	**7.743**	**8.275**	**8.039**

Considered the best-performing methods

Standard *vs* microarray-specific coders

Compression results (bpp)

Algorithm	IBB	MicroZip	Arizona
Neves			
JPEG2000			
JPEG-LS			

[a]

Standard *vs* microarray-specific coders

Compression results (bpp)

Algorithm	IBB	MicroZip	Arizona
Neves	8.039		
JPEG2000	9.182		
JPEG-LS	9.904		

[a]

Standard *vs* microarray-specific coders

Compression results (bpp)

Algorithm	IBB	MicroZip	Arizona
Neves	8.039	8.667	
JPEG2000	9.182	9.467	
JPEG-LS	9.904	9.441	

[a]

Standard *vs* microarray-specific coders

Compression results (bpp)

Algorithm	IBB	MicroZip	Arizona
Neves	8.039	8.667	8.275
JPEG2000	9.182	9.467	9.064[a]
JPEG-LS	9.904	9.441	8.646

[a] Page 44, Table 3.1

Standard *vs* microarray-specific coders

Compression results (bpp)

Algorithm	IBB	MicroZip	Arizona
Neves	8.039	8.667	8.275
JPEG2000	9.182	9.467	9.064[a]
JPEG-LS	9.904	9.441	8.646

[a] Page 44, Table 3.1

- Microarray-specific: up to 1 bpp better

Standard *vs* microarray-specific coders

Compression results (bpp)

Algorithm	IBB	MicroZip	Arizona
Neves	8.039	8.667	8.275
JPEG2000	9.182	9.467	9.064[a]
JPEG-LS	9.904	9.441	8.646

[a] Page 44, Table 3.1

- Microarray-specific: up to 1 bpp better
- Standard: DICOM compatible (clinic scenarios)

Standard *vs* microarray-specific coders

Compression results (bpp)

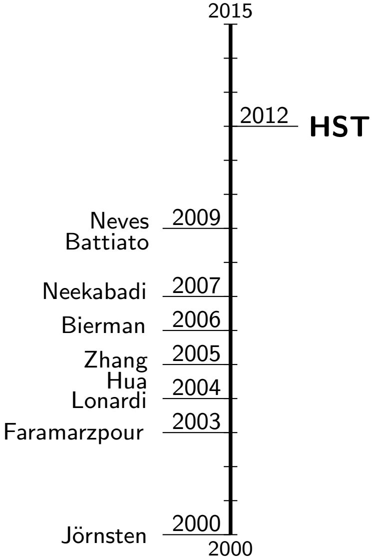
Algorithm	IBB	MicroZip	Arizona
Neves	8.039	8.667	8.275
JPEG2000	9.182	9.467	9.064[a]
JPEG-LS	9.904	9.441	8.646

[a] Page 44, Table 3.1

- Microarray-specific: up to 1 bpp better
- Standard: DICOM compatible (clinic scenarios)

Our approach: improve standards

Contribution 1: Histogram Swap Transform



"DNA microarray image coding,"

Miguel Hernández-Cabronero,
Juan Muñoz-Gómez, Ian Blanes,
Joan Serra-Sagristà,
Michael W. Marcellin

IEEE Data Compression Conference,
DCC, 2012

Contribution 1: Histogram Swap Transform

Why JPEG2000?

Contribution 1: Histogram Swap Transform

Why JPEG2000?

- DICOM compatible

Contribution 1: Histogram Swap Transform

Why JPEG2000?

- DICOM compatible
- Acceptable base performance

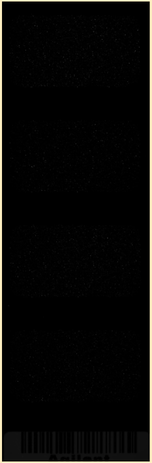
Contribution 1: Histogram Swap Transform

Why JPEG2000?

- DICOM compatible
- Acceptable base performance
- Features and Flexibility

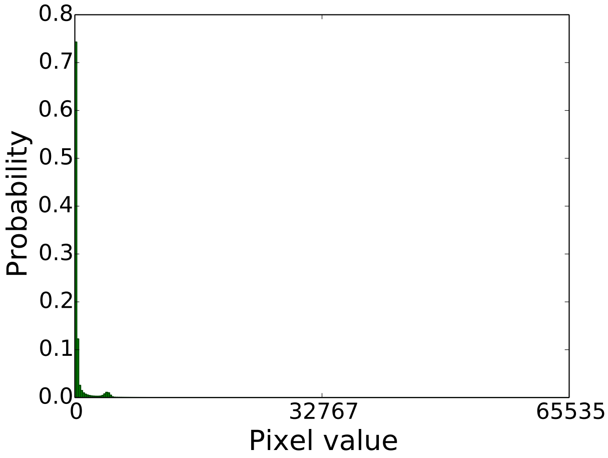
Contribution 1: Histogram Swap Transform

DNA microarray image (original)



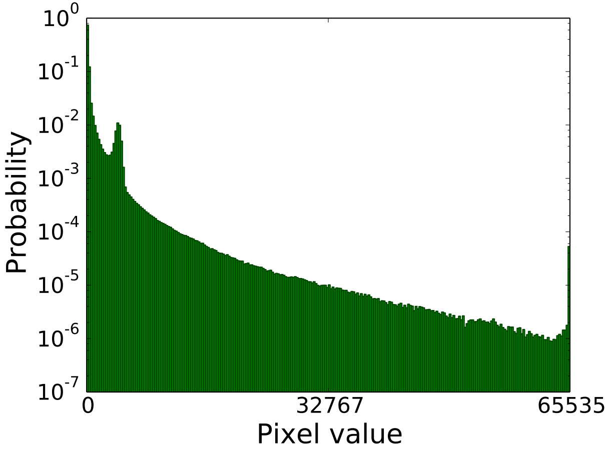
Contribution 1: Histogram Swap Transform

DNA microarray image (original)



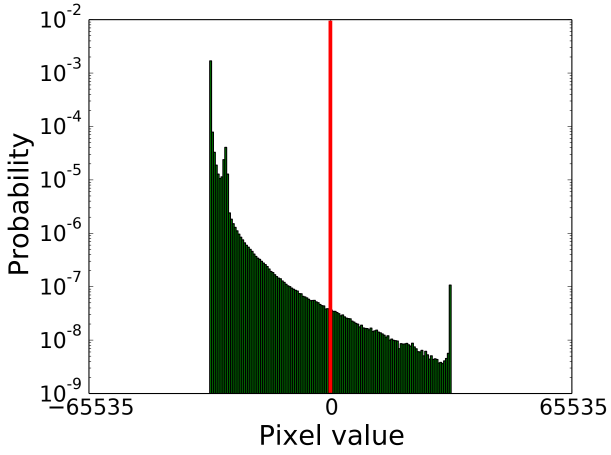
Contribution 1: Histogram Swap Transform

DNA microarray image (original)



Contribution 1: Histogram Swap Transform

DNA microarray image **(level offset)**

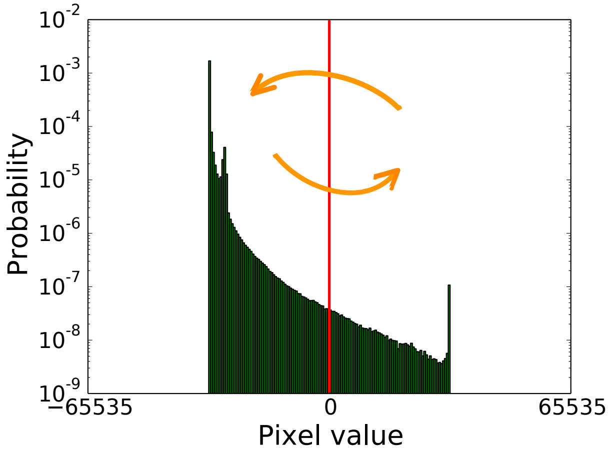


- Asymmetric

$\Rightarrow$ Poor performance

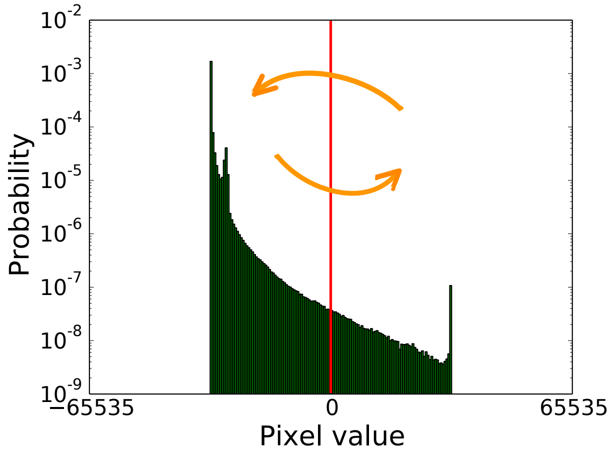
Contribution 1: Histogram Swap Transform

Histogram Swap Transform (before)



Contribution 1: Histogram Swap Transform

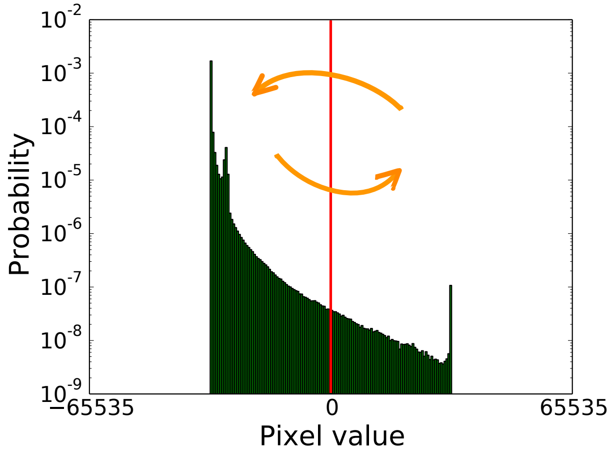
Histogram Swap Transform (before)



- Reversible

Contribution 1: Histogram Swap Transform

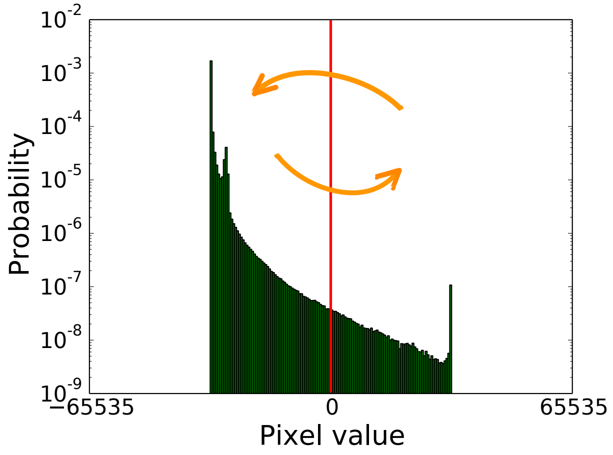
Histogram Swap Transform (before)



- Reversible
- Fast

Contribution 1: Histogram Swap Transform

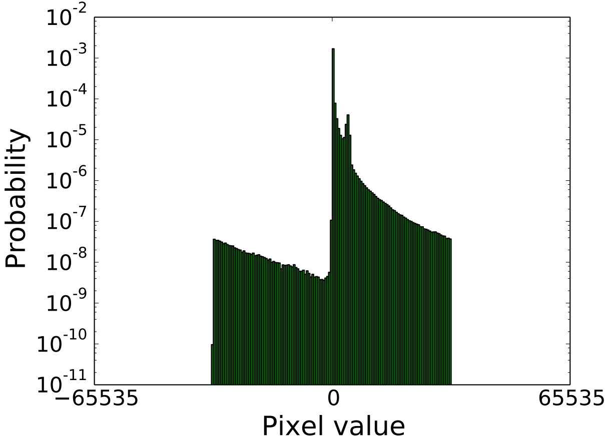
Histogram Swap Transform (before)



- Reversible
- Fast
- JPEG2000 Compatible

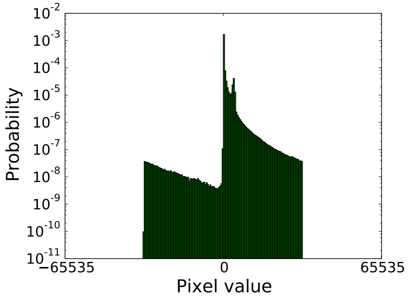
Contribution 1: Histogram Swap Transform

Histogram Swap Transform (after)

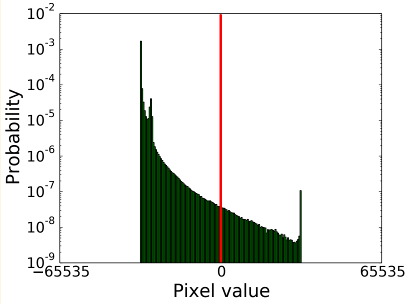


Contribution 1: Histogram Swap Transform

Before



After



Contribution 1: Histogram Swap Transform

JPEG2000 compression results (bpp)

Corpus	Without HST	With HST	Difference	
ApoA1	10.999	10.786	1.97%	-
ISREC	10.888	10.624	2.48%	
Arizona	9.064	8.795	3.06%	
MicroZip	9.467	9.157	3.39%	
Stanford	7.969	7.685	3.70%	
Omnibus	7.549	7.103	6.28%	
IBB	9.182	8.392	9.41%	
Yeast	6.829	5.911	15.53%	+

Contribution 1: Histogram Swap Transform

JPEG2000 compression results (bpp)

Corpus	Without HST	With HST	Difference	
ApoA1	10.999	10.786	1.97%	-
ISREC	10.888	10.624	2.48%	
Arizona	9.064	8.795	3.06%	
MicroZip	9.467	9.157	3.39%	
Stanford	7.969	7.685	3.70%	
Omnibus	7.549	7.103	6.28%	
IBB	9.182	8.392	9.41%	
Yeast	6.829	5.911	15.53%	+

- Improvements for **all** datasets (2%-15%)

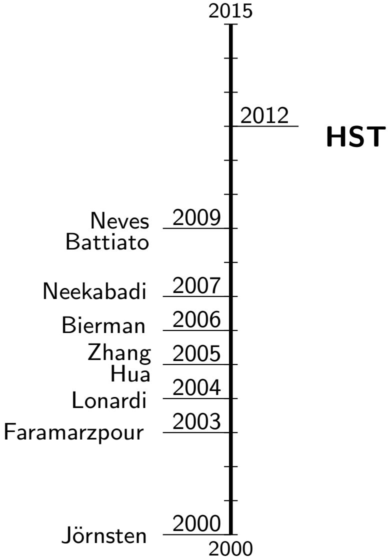
Contribution 1: Histogram Swap Transform

JPEG2000 compression results (bpp)

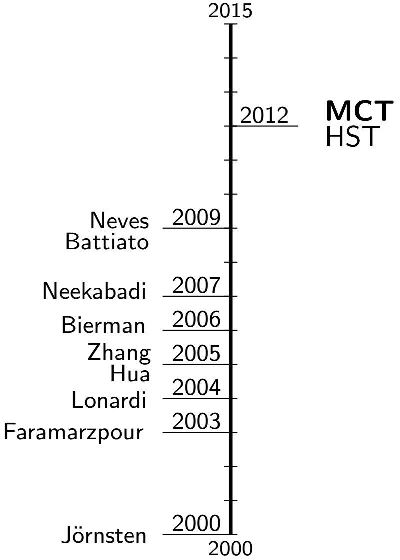
Corpus	Without HST	With HST	Difference	
ApoA1	10.999	10.786	1.97%	-
ISREC	10.888	10.624	2.48%	
Arizona	9.064	8.795	3.06%	
MicroZip	9.467	9.157	3.39%	
Stanford	7.969	7.685	3.70%	
Omnibus	7.549	7.103	6.28%	
IBB	9.182	8.392	9.41%	
Yeast	6.829	5.911	15.53%	+

- Improvements for **all** datasets (2%-15%)
- **Can improve further?**

Contribution 1: Histogram Swap Transform



Contribution 2: Multicomponent compression



- "Multicomponent compression of DNA microarray images,"

  Miguel Hernández-Cabronero *et al.*

  CEDI Workshop on Multimedia Data Coding and Transmission, WMDCT, 2012

- "Compression of DNA Microarray Images," Miguel

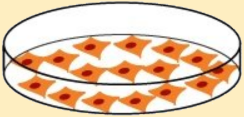
  Hernández-Cabronero, Victor Sanchez, Michael W. Marcellin, Joan Serra-Sagristà

  In Book "Microarray Image and Data

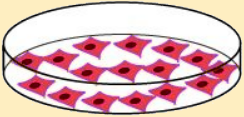
  Analysis: Theory and Practice"

Contribution 2: Multicomponent compression

Biological samples



Tumoral tissue



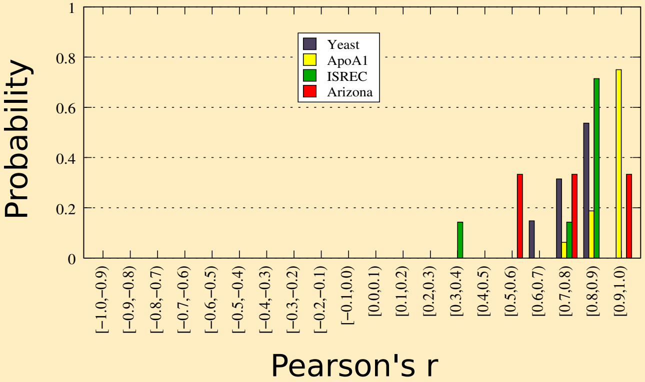
Healthy tissue

Questions:

- Similar images?
- Compress together?

Contribution 2: Multicomponent compression

Correlation (Reg/green pairs only)



Correlation betweeen 0.75-0.92

Contribution 2: Multicomponent transform

JPEG2000 compression results (bpp)

Corpus	No transform	5/3 DWT	RKLT	DPCM	RHaar
Yeast	6.829	6.786	9.279	6.439	6.790
ApoA1	11.524	11.217	10.956	11.289	11.218
ISREC	10.888	11.451	11.468	11.203	11.452
Arizona	9.548	9.649	9.439	9.386	9.649
IBB	9.182	9.948	10.269	9.602	9.948

Contribution 2: Multicomponent transform

JPEG2000 compression results (bpp)

Corpus	No transform	5/3 DWT	RKLT	DPCM	RHaar
Yeast	6.829	**6.786**	9.279	**6.439**	**6.790**
ApoA1	11.524	**11.217**	**10.956**	**11.289**	**11.218**
ISREC	10.888	11.451	11.468	11.203	11.452
Arizona	9.548	9.649	**9.439**	**9.386**	9.649
IBB	9.182	9.948	10.269	9.602	9.948

Results:

- Some improvements
- No "golden" transform

Lossless compression conclusions

Compression results (bpp)

Corpus	JPEG-LS	JPEG2000	HST	Neves
Yeast	8.580	6.829	**5.911**	5.521
ApoA1	10.608	10.999	**10.786**	10.223
ISREC	11.145	10.888	**10.624**	10.199
Stanford	7.571	7.969	**7.685**	7.335
MicroZip	9.441	9.467	**9.157**	8.667
Omnibus	6.952	8.121	**7.103**	7.743
Arizona	8.646	9.064	**8.795**	8.275
IBB	9.904	9.064	**8.392**	8.039

Lossless compression conclusions

Compression results (bpp)

Corpus	JPEG-LS	JPEG2000	HST	Neves
Yeast	8.580	6.829	**5.911**	5.521
ApoA1	10.608	10.999	**10.786**	10.223
ISREC	11.145	10.888	**10.624**	10.199
Stanford	7.571	7.969	**7.685**	7.335
MicroZip	9.441	9.467	**9.157**	8.667
Omnibus	6.952	8.121	**7.103**	7.743
Arizona	8.646	9.064	**8.795**	8.275
IBB	9.904	9.064	**8.392**	8.039
Average	9.106	9.001	**8.557**	8.250

Lossless compression conclusions

Compression results (bpp)

Corpus	JPEG-LS	JPEG2000	HST	Neves
Yeast	8.580	6.829	**5.911**	5.521
ApoA1	10.608	10.999	**10.786**	10.223
ISREC	11.145	10.888	**10.624**	10.199
Stanford	7.571	7.969	**7.685**	7.335
MicroZip	9.441	9.467	**9.157**	8.667
Omnibus	6.952	8.121	**7.103**	7.743
Arizona	8.646	9.064	**8.795**	8.275
IBB	9.904	9.064	**8.392**	8.039
Average	9.106	9.001	**8.557**	8.250

- JPEG2000: Best overall standard

Lossless compression conclusions

Compression results (bpp)

Corpus	JPEG-LS	JPEG2000	HST	Neves
Yeast	8.580	6.829	**5.911**	5.521
ApoA1	10.608	10.999	**10.786**	10.223
ISREC	11.145	10.888	**10.624**	10.199
Stanford	7.571	7.969	**7.685**	7.335
MicroZip	9.441	9.467	**9.157**	8.667
Omnibus	6.952	8.121	**7.103**	7.743
Arizona	8.646	9.064	**8.795**	8.275
IBB	9.904	9.064	**8.392**	8.039
Average	9.106	9.001	**8.557**	8.250

- JPEG2000: Best overall standard
- HST+MCT or HST+DWT: Worse performance

Contents

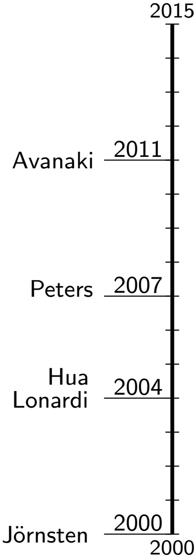
1 Introduction: DNA Microarrays

2 Lossless compression

3 Lossy compression

4 Conclusions

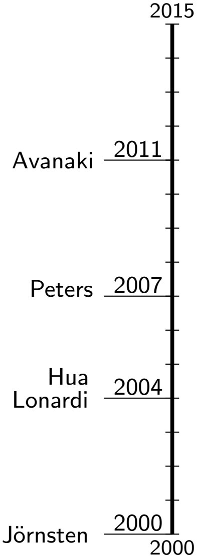
State of the art



Lossless compression

- Perfect fidelity
- Small compression ratios

State of the art



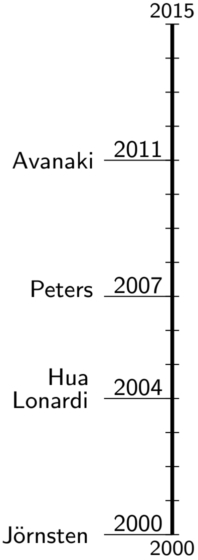
Lossless compression

- Perfect fidelity
- Small compression ratios

Lossy compression?

- Arbitrary compression ratios

State of the art



Lossless compression

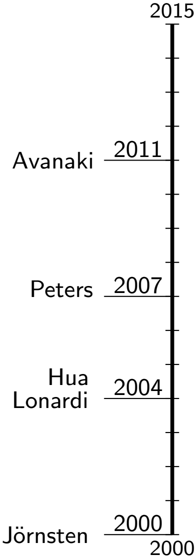
- Perfect fidelity
- Small compression ratios

Lossy compression?

- Arbitrary compression ratios

but image distortion

State of the art



Lossless compression

- Perfect fidelity
- Small compression ratios

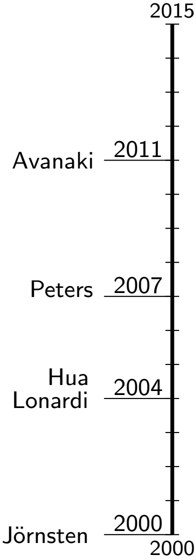
Lossy compression?

- Arbitrary compression ratios

but image distortion

$\Rightarrow$ analysis distortion

State of the art



Lossless compression

- Perfect fidelity
- Small compression ratios

Lossy compression?

- Arbitrary compression ratios

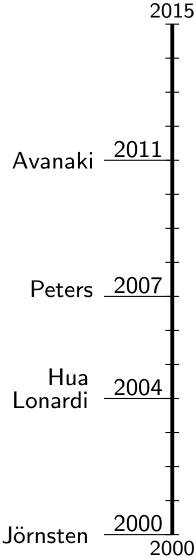
but image distortion

$\Rightarrow$ analysis distortion

Our contributions:

- Analysis distortion assessment

State of the art



Lossless compression

- Perfect fidelity
- Small compression ratios

Lossy compression?

- Arbitrary compression ratios

but image distortion

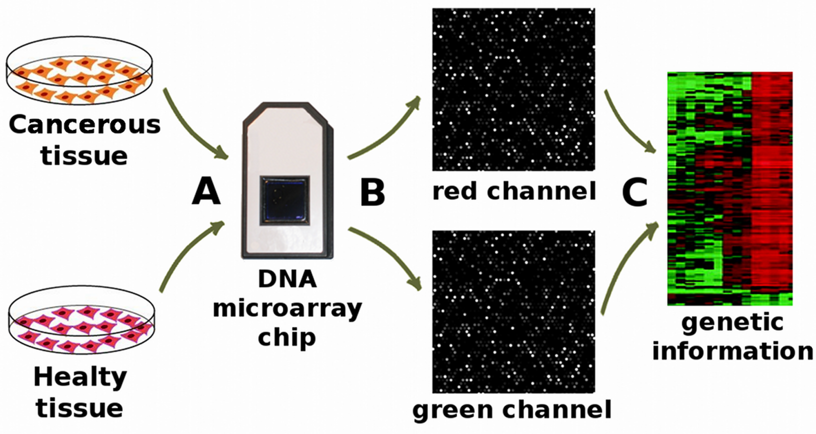
$\Rightarrow$ analysis distortion

Our contributions:

- Analysis distortion assessment
- Analysis-driven compression

Analysis of DNA Microarray Images

Analysis process



Analysis of DNA Microarray Images

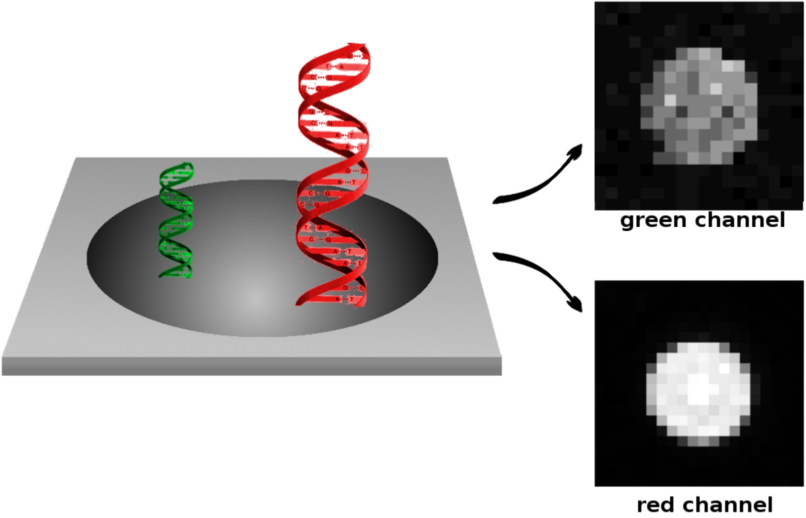
Analysis process

Gene	Expression intensity Sample 1 μ_1	Expression intensity Sample 2 μ_2	Intensity ratio μ_1/μ_2
Gene 1	245	249	**0.98**
Gene 2	54	13	4.15
Gene 3	200	1530	0.13
⋮	⋮	⋮	⋮
Gene N	100	103	0.97

How are they obtained?

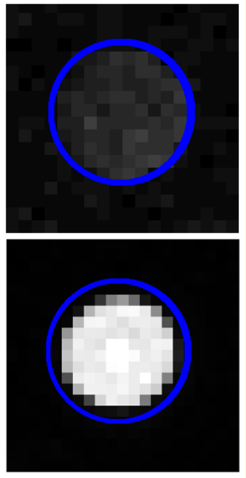
Analysis of DNA Microarray Images

Spot Information



Analysis of DNA Microarray Images

Spot detection

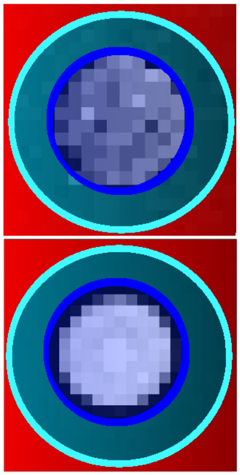


How to obtain the genetic data:

1. Detect the spots

Analysis of DNA Microarray Images

Spot segmentation

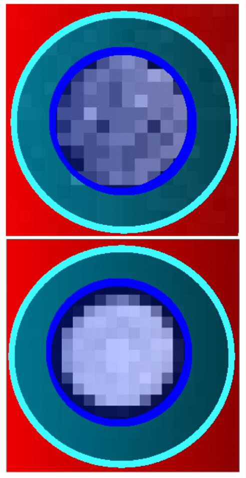


How to obtain the genetic data:

1. Detect the spots
2. Segment
   - Spot
   - Local background
   - Background

Analysis of DNA Microarray Images

Spot segmentation



How to obtain the genetic data:

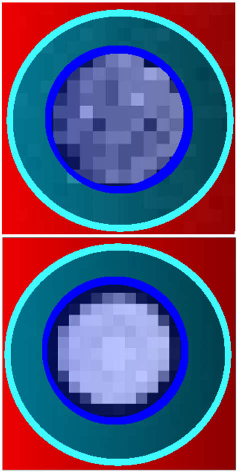
1. Detect the spots
2. Segment
   - Spot
   - Local background
   - Background
3. Calculate the
   Corrected Ratio of Means (CRM):

$$\mathrm{CRM} = \frac{\mu_{\text{spot}}^{\text{green}}}{\mu_{\text{spot}}^{\text{red}}}$$

μ: average intensity

Analysis of DNA Microarray Images

Spot segmentation



How to obtain the genetic data:

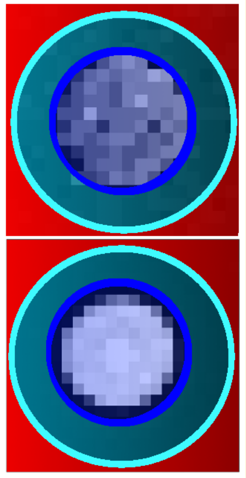
1. Detect the spots
2. Segment
   - Spot
   - Local background
   - Background
3. Calculate the Corrected Ratio of Means (CRM):

$$\mathrm{CRM} = \frac{\mu_{\text{spot}}^{\text{green}} - \mu_{\text{localBG}}^{\text{green}}}{\mu_{\text{spot}}^{\text{red}} - \mu_{\text{localBG}}^{\text{red}}}$$

μ: average intensity

Analysis of DNA Microarray Images

Spot segmentation



How to obtain the genetic data:

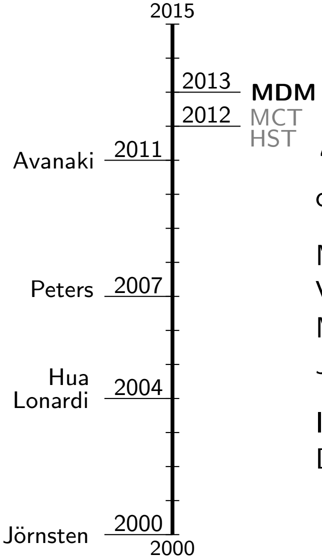
1. Detect the spots
2. Segment
   - Spot
   - Local background
   - Background
3. Calculate the Corrected Ratio of Means (CRM):

$$\mathrm{CRM} = \frac{\mu_{\mathrm{spot}}^{\mathrm{green}} - \mu_{\mathrm{localBG}}^{\mathrm{green}}}{\mu_{\mathrm{spot}}^{\mathrm{red}} - \mu_{\mathrm{localBG}}^{\mathrm{red}}}$$

μ: average intensity

4. normalize

Contribution 3: Microarray Distortion Metric



"A distortion metric for the lossy compression of DNA microarray images"

Miguel Hernández-Cabronero,
Victor Sanchez,
Michael W. Marcellin,
Joan Serra-Sagristà

IEEE Data Compression Conference, DCC, 2013

Contribution 3: Microarray Distortion Metric

How to assess loss importance:

Contribution 3: Microarray Distortion Metric

How to assess loss importance:

- PSNR, MSE

Contribution 3: Microarray Distortion Metric

How to assess loss importance:

- PSNR, MSE $\Rightarrow$ not enough

Contribution 3: Microarray Distortion Metric

How to assess loss importance:

- PSNR, MSE $\Rightarrow$ not enough
- Compare analysis results

Contribution 3: Microarray Distortion Metric

How to assess loss importance:

- PSNR, MSE $\Rightarrow$ not enough
- Compare analysis results (slow but accurate)

  2003 Jörnsten
  2004 Hua
  2009 Xu

Contribution 3: Microarray Distortion Metric

How to assess loss importance:

- PSNR, MSE $\Rightarrow$ not enough
- Compare analysis results (slow but accurate)
  2003 Jörnsten
  2004 Hua
  2009 Xu
- **Assess loss without analyzing?**

Contribution 3: Microarray Distortion Metric

Calculation

$$MDM = 10 \log_{10} \frac{\text{max\_val}^2}{ME}$$

Contribution 3: Microarray Distortion Metric

Calculation

$$MDM = 10 \log_{10} \frac{\text{max\_val}^2}{ME}$$

$$ME \sim \text{max\_val}^p, \quad p \in [1, 2]$$

- ME: Microarray Error

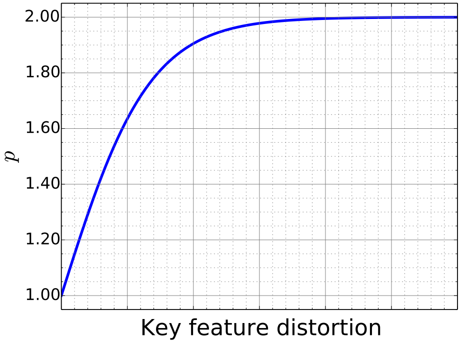
Contribution 3: Microarray Distortion Metric

Calculation

$$MDM = 10 \log_{10} \frac{\text{max\_val}^2}{ME}$$

$$ME \sim \text{max\_val}^p, \quad p \in [1, 2]$$

- ME: Microarray Error
- p: sensitive error exponent



Contribution 3: Microarray Distortion Metric

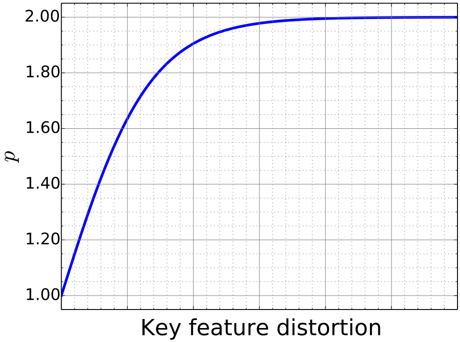
Calculation

$$MDM = 10 \log_{10} \frac{\text{max\_val}^2}{ME}$$

$$ME \sim \text{max\_val}^p, \quad p \in [1, 2]$$

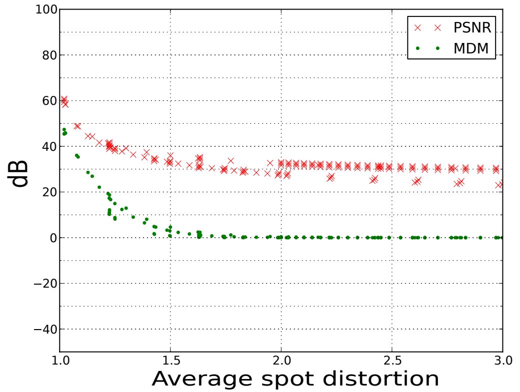
$$p = \text{logistic(feat. distortion)}$$

- ME: Microarray Error
- p: sensitive error exponent



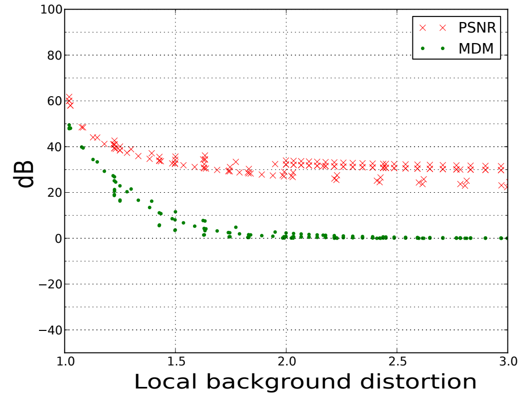
Contribution 3: Microarray Distortion Metric

Results (spots)



Contribution 3: Microarray Distortion Metric

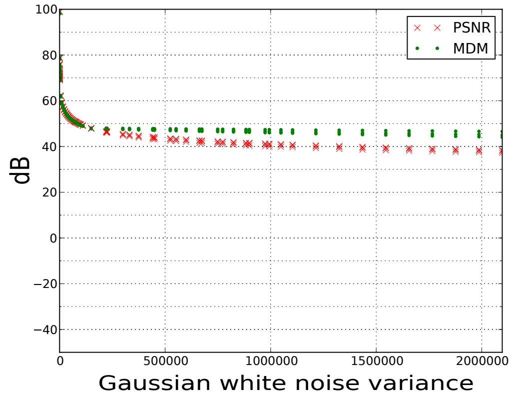
Results (local background)



✓ Important changes

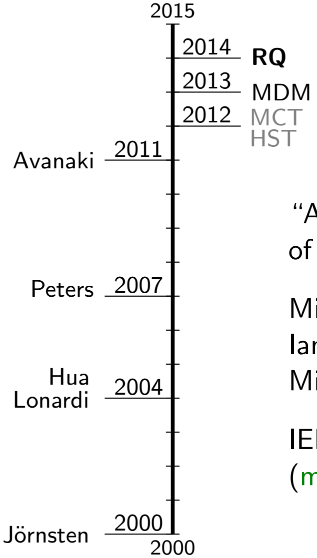
Contribution 3: Microarray Distortion Metric

Results (Gaussian white noise)



- ✓ Important changes
- ✓ Unimportant changes

Contribution 4: Relative Quantization



"Analysis-driven lossy compression of DNA microarray images"

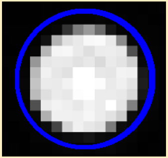
Miguel Hernández-Cabronero
Ian Blanes, Armando J. Pinho
Michael W. Marcellin, Joan Serra-Sagristà

IEEE Transactions on Medical Imaging, 2015 (major revision)

Contribution 4: Relative Quantization

Relative error: important for

Segmentation

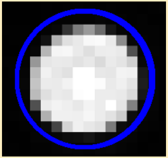


(contrast)

Contribution 4: Relative Quantization

Relative error: important for

Segmentation



(contrast)

CRM calculation

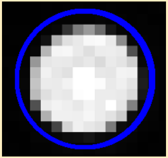
$$\mathrm{CRM} = \frac{\mu^{\mathrm{green}}_{\mathrm{spot}} - \mu^{\mathrm{green}}_{\mathrm{localBG}}}{\mu^{\mathrm{red}}_{\mathrm{spot}} - \mu^{\mathrm{red}}_{\mathrm{localBG}}}$$

(error calculation)

Contribution 4: Relative Quantization

Relative error: important for

Segmentation



(contrast)

CRM calculation

$$\mathrm{CRM} = \frac{\mu^{\mathrm{green}}_{\mathrm{spot}} - \mu^{\mathrm{green}}_{\mathrm{localBG}}}{\mu^{\mathrm{red}}_{\mathrm{spot}} - \mu^{\mathrm{red}}_{\mathrm{localBG}}}$$

(error calculation)

Our approach:

$$\text{Control pixel relative error} \Rightarrow \begin{cases} \text{Preserve edge detection} \\ \text{Limit CRM relative error} \end{cases}$$

Contribution 4: Relative Quantization

Keep only k bits from each pixel

MSB	0	0	0
	0	0	0
	⋯	⋯	⋯
	1	0	0
	0	0	0
	1	1	0
	0	1	0
	0	0	0
	1	1	0
	0	1	0
	1	1	0
LSB	1	1	1

Contribution 4: Relative Quantization

Keep only k bits from each pixel

MSB	0	0	0
	0	0	0
	...	...	...
	$\boxed{1}$	0	0
	$\boxed{0}$	0	0
$k = 2$	1	$\boxed{1}$	0
	0	$\boxed{1}$	0
	0	0	0
	1	1	0
	0	1	0
	1	1	0
LSB	1	1	$\boxed{1}$

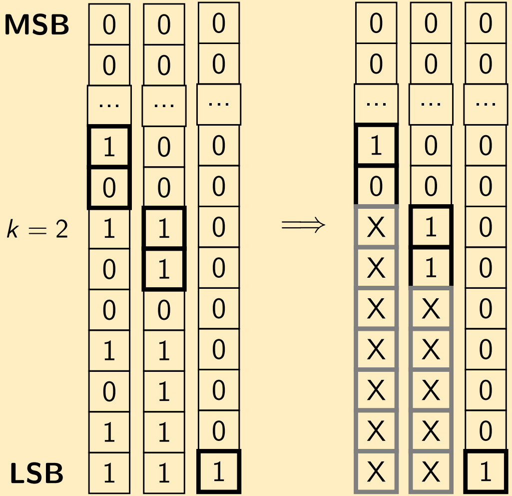
Contribution 4: Relative Quantization

Keep only k bits from each pixel

MSB	0	0	0		0	0	0
	0	0	0		0	0	0
	…	…	…		…	…	…
	1	0	0		1	0	0
	0	0	0		0	0	0
$k = 2$	1	1	0	$\Longrightarrow$	X	1	0
	0	1	0		X	1	0
	0	0	0		X	X	0
	1	1	0		X	X	0
	0	1	0		X	X	0
	1	1	0		X	X	0
LSB	1	1	1		X	X	1

Contribution 4: Relative Quantization

Keep only k bits from each pixel



✓ Controlled maximum relative error

k	maximum rel error (2^{-k})
1	0.500
2	0.250
3	0.125
4	0.062
5	0.031
6	0.016
7	0.008

Contribution 4: Relative Quantization

Why call it Relative *Quantization*?

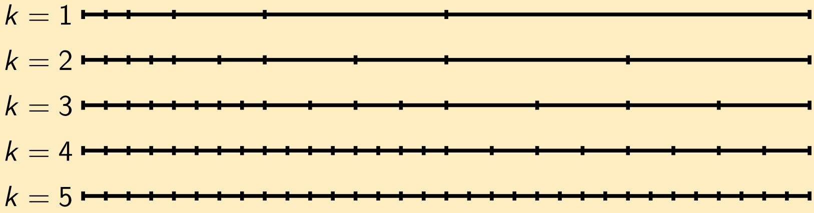
- Same value after discarding $\Rightarrow$ same quantization index

Contribution 4: Relative Quantization

	0	0	0	0	0	0	0		1
	0	0	0	0	0	0	0		1
	⋯	⋯	⋯	⋯	⋯	⋯	⋯		⋯
	0	0	0	0	0	0	0		X
Pixel values	0	0	0	0	0	0	0	⋯	X
	0	0	0	0	0	0	0		X
	0	0	0	0	0	0	0		X
	0	0	0	0	0	0	0		X
	0	0	0	0	0	0	1		X
$(k = 2)$	0	0	0	0	1	1	0		X
	0	0	1	1	0	1	X		X
	0	1	0	1	X	X	X		X
Index	0	1	2	3	4	5	6	⋯	31

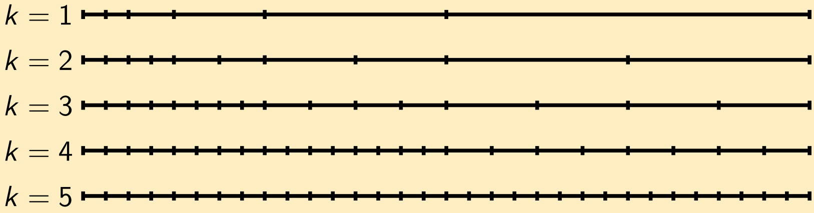
Contribution 4: Relative Quantization

Quantization Interval (5 bitplanes)



Contribution 4: Relative Quantization

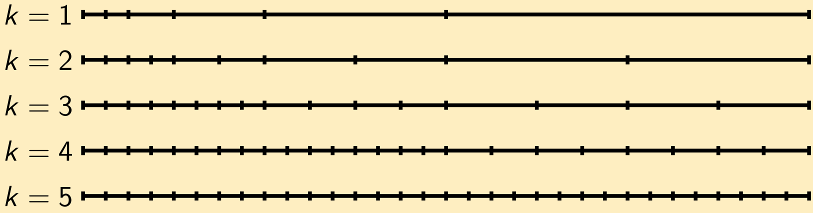
Quantization Interval (5 bitplanes)



- Effect: $\downarrow$number of symbols $\Rightarrow$ $\downarrow$compressed size

Contribution 4: Relative Quantization

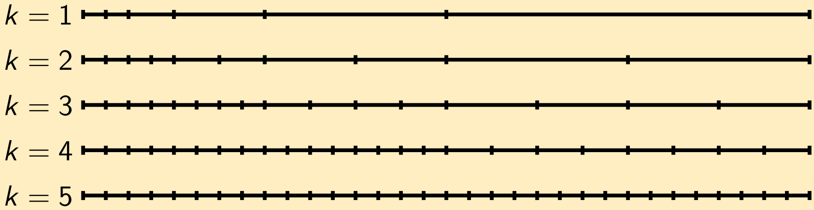
Quantization Interval (5 bitplanes)



- Effect: ↓number of symbols ⇒ ↓compressed size
- Parameterizable: $k \in \{1, 2, \mathbf{3}, \mathbf{4}, \mathbf{5}, 6, \ldots, 16\}$
  ($\downarrow k \Rightarrow$ more aggressive)

Contribution 4: Relative Quantization

Quantization Interval (5 bitplanes)



- Effect: $\downarrow$number of symbols $\Rightarrow$ $\downarrow$compressed size
- Parameterizable: $k \in \{1, 2, \mathbf{3}, \mathbf{4}, \mathbf{5}, 6, \ldots, 16\}$
  ($\downarrow k \Rightarrow$ more aggressive)
- Can use any (lossless) coder

Contribution 4: Relative Quantization

Relative Quantization Coder:

1. Quantize images
2. Compress quantization indices

Contribution 4: Relative Quantization

Relative Quantization Coder:

1. Quantize images
2. Compress quantization indices

Compression results (bpp, avg. all sets)

Coder	Relative Quantization k							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**Original**
JPEG2000	2.01	2.75	3.60	4.53	5.51	6.48	7.39	9.76
JPEG-LS	1.87	2.65	3.54	4.47	5.38	6.28	7.13	9.35
Neves	1.51	2.22	3.07	4.01	4.96	5.86	6.60	8.32

Contribution 4: Relative Quantization

Relative Quantization Coder:

1. Quantize images
2. Compress quantization indices

Compression results (bpp, avg. all sets)

Coder	**Relative Quantization** k							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**Original**
JPEG2000	2.01	2.75	3.60	4.53	5.51	6.48	7.39	9.76
JPEG-LS	1.87	2.65	3.54	4.47	5.38	6.28	7.13	9.35
Neves	1.51	2.22	3.07	4.01	4.96	5.86	6.60	8.32

- Good compression performance

Contribution 4: Relative Quantization

Relative Quantization Coder:

1. Quantize images
2. Compress quantization indices

Compression results (bpp, avg. all sets)

Coder	**Relative Quantization** k							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**Original**
JPEG2000	2.01	2.75	3.60	4.53	5.51	6.48	7.39	9.76
JPEG-LS	1.87	2.65	3.54	4.47	5.38	6.28	7.13	9.35
Neves	1.51	2.22	3.07	4.01	4.96	5.86	6.60	8.32

- Good compression performance (*e.g.*, 1.5 bpp vs 8.3 bpp)

Contribution 4: Relative Quantization

Relative Quantization Coder:

1. Quantize images
2. Compress quantization indices

Compression results (bpp, avg. all sets)

Coder	**Relative Quantization k**							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**Original**
JPEG2000	2.01	2.75	3.60	4.53	5.51	6.48	7.39	9.76
JPEG-LS	1.87	2.65	3.54	4.47	5.38	6.28	7.13	9.35
Neves	1.51	2.22	3.07	4.01	4.96	5.86	6.60	8.32

- Good compression performance (*e.g.*, 1.5 bpp vs 8.3 bpp)
- **Acceptable analysis distortion?**

Contribution 4: Relative Quantization

CRM distortion

$$\mathrm{CRM} = \frac{\mu^{\mathrm{green}}_{\mathrm{spot}} - \mu^{\mathrm{green}}_{\mathrm{localBG}}}{\mu^{\mathrm{red}}_{\mathrm{spot}} - \mu^{\mathrm{red}}_{\mathrm{localBG}}}$$

Analysis-based distortion metrics:

- **ARE$_{\mathbf{CRM}}$**
- FWDOC

Contribution 4: Relative Quantization

CRM distortion

$$\mathrm{CRM} = \frac{\mu_{\mathrm{spot}}^{\mathrm{green}} - \mu_{\mathrm{localBG}}^{\mathrm{green}}}{\mu_{\mathrm{spot}}^{\mathrm{red}} - \mu_{\mathrm{localBG}}^{\mathrm{red}}}$$

CRM Average Relative Error:

$$\mathrm{ARE}_{\mathrm{CRM}} = \frac{1}{n} \sum_{i=1}^{n} \frac{|\mathrm{CRM}_i - \widehat{\mathrm{CRM}}_i|}{\delta + |\mathrm{CRM}_i|}$$

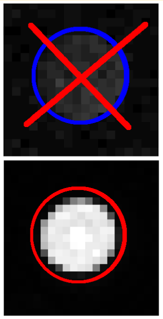
Analysis-based distortion metrics:

- **$\mathrm{ARE}_{\mathrm{CRM}}$**
- FWDOC

Contribution 4: Relative Quantization

How to run an spot:

Wrong detection



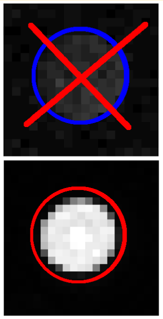
Analysis-based distortion metrics:

- ARE_{CRM}
- FWDOC

Contribution 4: Relative Quantization

How to ruin an spot:

Wrong detection



Wrong classification

A $\mathrm{CRM} < 0.5$

B $0.5 \leq \mathrm{CRM} \leq 2$

C $\mathrm{CRM} > 2$

(should not alter classification)

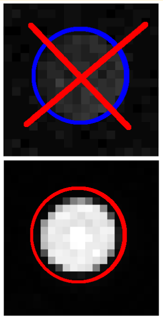
Analysis-based distortion metrics:

- $\mathrm{ARE}_{\mathrm{CRM}}$
- FWDOC

Contribution 4: Relative Quantization

How to ruin an spot:

Wrong detection



Wrong classification

- A $\mathrm{CRM} < 0.5$
- B $0.5 \leq \mathrm{CRM} \leq 2$
- C $\mathrm{CRM} > 2$

(should not alter classification)

Analysis-based distortion metrics:

- $\mathrm{ARE}_{\mathrm{CRM}}$
- **FWDOC**

Fraction of spots
Wrongly Detected Or Classified (FWDOC)

Contribution 4: Relative Quantization

Impact on the image analysis

	CRM error	Detect.+Class. error

Contribution 4: Relative Quantization

Impact on the image analysis

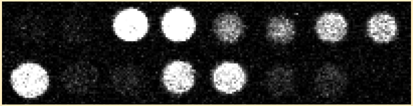
	CRM error	Detect.+Class. error
$k = 1$	0.562	0.148
$k = 2$	0.124	0.100
$k = 3$	0.121	0.064
$k = 4$	0.078	0.044
$k = 5$	0.064	0.030
$k = 6$	0.039	0.019
$k = 7$	0.028	0.014

Acceptability?

Contribution 4: Relative Quantization

Impact on the image analysis

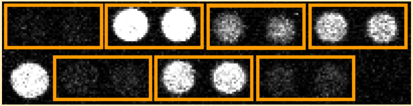
	CRM error	Detect.+Class. error
$k=1$	0.562	0.148
$k=2$	0.124	0.100
$k=3$	0.121	0.064
$k=4$	0.078	0.044
$k=5$	0.064	0.030
$k=6$	0.039	0.019
$k=7$	0.028	0.014



Contribution 4: Relative Quantization

Impact on the image analysis

	CRM error	**Detect.+Class. error**
$k=1$	0.562	0.148
$k=2$	0.124	0.100
$k=3$	0.121	0.064
$k=4$	0.078	0.044
$k=5$	0.064	0.030
$k=6$	0.039	0.019
$k=7$	0.028	0.014
Replicated	0.254	0.212



Contribution 4: Relative Quantization

Impact on the image analysis

	CRM error	**Detect.+Class. error**
$k = 1$	0.562	0.148
$k = 2$	0.124	0.100
$k = 3$	0.121	0.064
$k = 4$	0.078	0.044
$k = 5$	0.064	0.030
$k = 6$	0.039	0.019
$k = 7$	0.028	0.014
Replicated	0.254	0.212

- ✓ Acceptable
- ✓ Good compression performance

Contribution 4: Relative Quantization

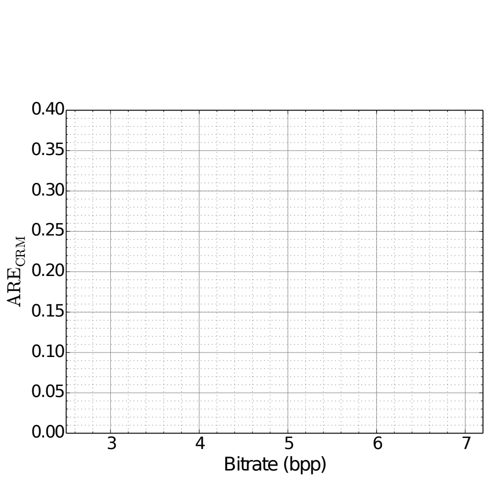
Impact on the image analysis

	CRM error	Detect.+Class. error
$k = 1$	0.562	0.148
$k = 2$	0.124	0.100
$k = 3$	0.121	0.064
$k = 4$	0.078	0.044
$k = 5$	0.064	0.030
$k = 6$	0.039	0.019
$k = 7$	0.028	0.014
Replicated	0.254	0.212

- ✓ Acceptable
- ✓ Good compression performance
- ? Comparison to other coders?

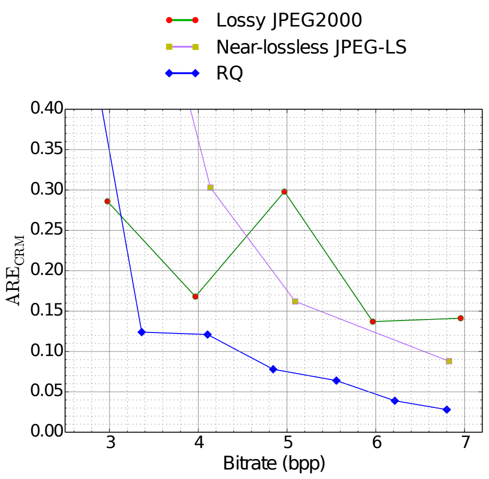
Contribution 4: Relative Quantization

RQ vs Others (CRM)



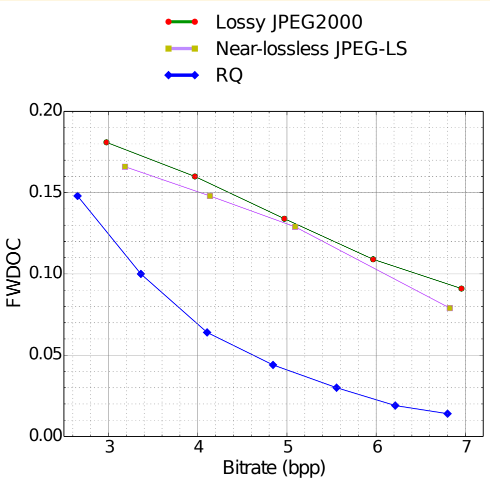
Contribution 4: Relative Quantization

RQ vs Others (CRM)

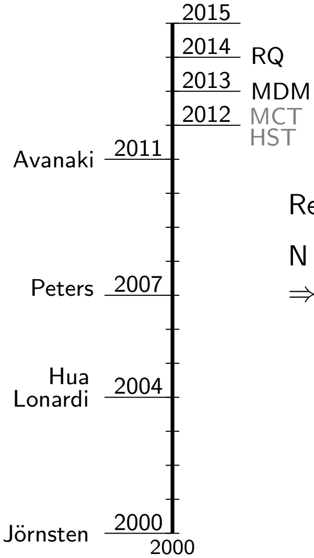


Contribution 4: Relative Quantization

RQ vs Others (detect. or class.)



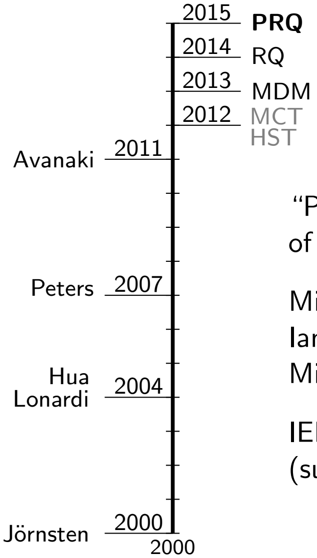
Contribution 4: Relative Quantization



Relative Quantization Coder:

N decompression qualities
$\Rightarrow$ N compressed versions (bad!)

Contribution 5: Progressive RQ

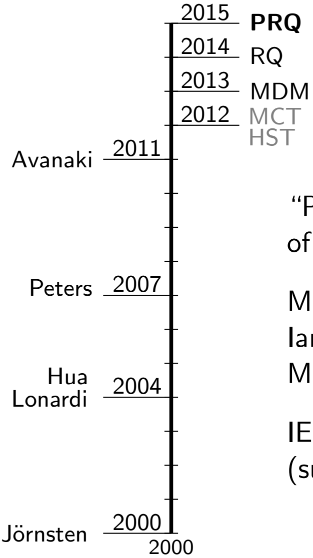


"Progressive Lossy-to-lossless compression of DNA Microarray Images"

Miguel Hernández-Cabronero
Ian Blanes, Armando J. Pinho
Michael W. Marcellin, Joan Serra-Sagristà

IEEE Signal Processing Letters, 2015 (submitted)

Contribution 5: Progressive RQ



(Last contribution)

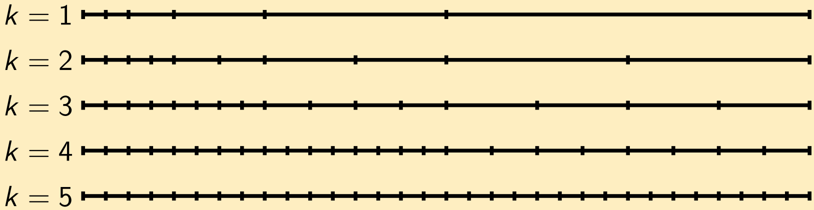
"Progressive Lossy-to-lossless compression of DNA Microarray Images"

Miguel Hernández-Cabronero
Ian Blanes, Armando J. Pinho
Michael W. Marcellin, Joan Serra-Sagristà

IEEE Signal Processing Letters, 2015 (submitted)

Contribution 5: Progressive Relative Quantization

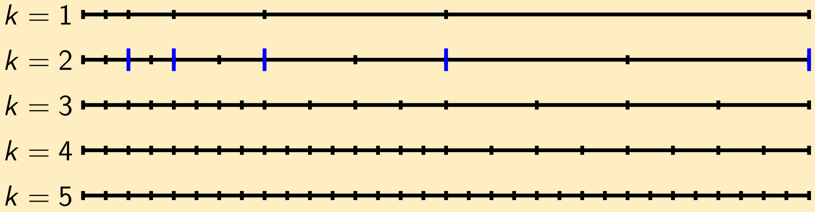
Progressivity (5 bitplanes)



Obs 1: Can refine RQ($k = 1$) into RQ($k = 2$)

Contribution 5: Progressive Relative Quantization

Progressivity (5 bitplanes)



Obs 1: Can refine RQ($k = 1$) into RQ($k = 2$)

Contribution 5: Progressive Relative Quantization

Progressivity (5 bitplanes)

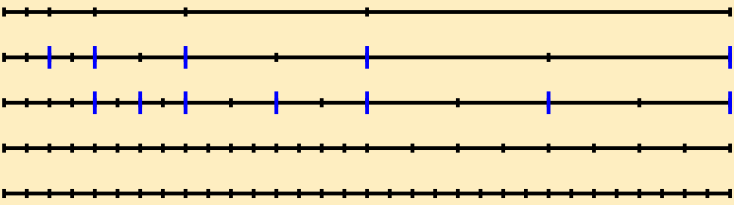
$k = 1$

$k = 2$

$k = 3$

$k = 4$

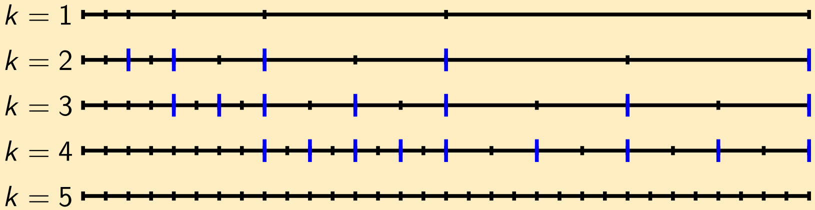
$k = 5$



Obs 1: Can refine RQ($k = 1$) into RQ($k = 2$)

Contribution 5: Progressive Relative Quantization

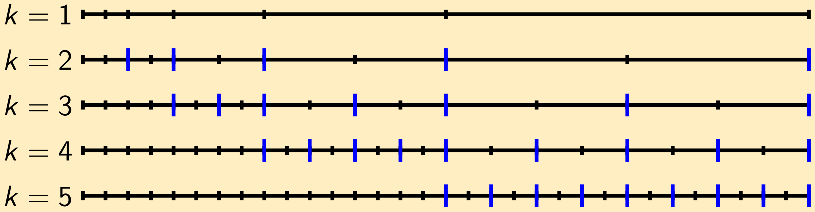
Progressivity (5 bitplanes)



Obs 1: Can refine RQ($k = 1$) into RQ($k = 2$)

Contribution 5: Progressive Relative Quantization

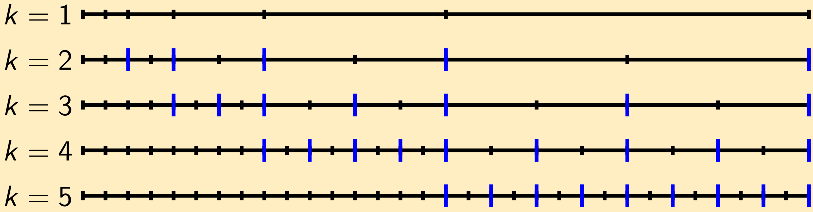
Progressivity (5 bitplanes)



Obs 1: Can refine RQ($k = 1$) into RQ($k = 2$)

Contribution 5: Progressive Relative Quantization

Progressivity (5 bitplanes)



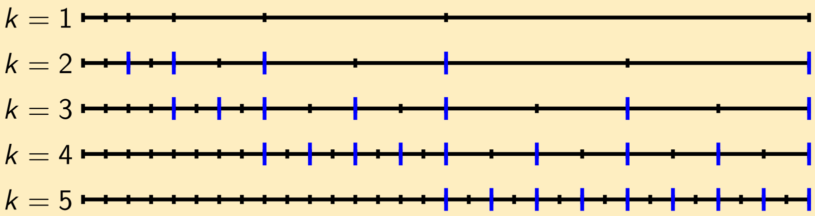
Obs 1: Can refine k into $k+1$ ($\Delta_{k \to k+1}$)

Progressive RQ (PRQ) representation

$$\mathrm{PRQ} = \mathrm{RQ}(k=1), \Delta_{1\to 2}, \ldots, \Delta_{15 \to 16}$$

Contribution 5: Progressive Relative Quantization

Progressivity (5 bitplanes)



Obs 1: Can refine k into $k+1$ ($\Delta_{k \to k+1}$)

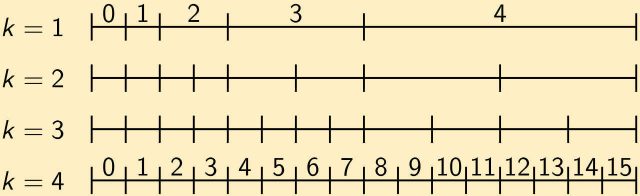
Progressive RQ (PRQ) representation

$$\mathrm{PRQ} = \mathrm{RQ}(k=1), \Delta_{1\to2}, \ldots, \Delta_{15\to16}$$

Obs 2: Can reconstruct as RQ(1), RQ(2), ... or RQ(16) = lossless

Contribution 5: Progressive Relative Quantization

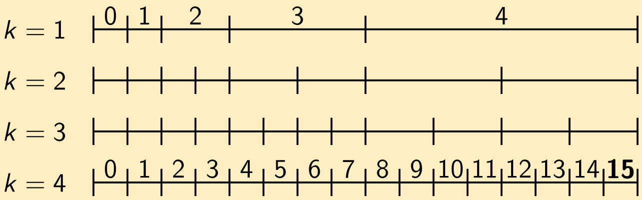
Quantization Indices (4 bitplanes)



(Here, $k = 4 \Rightarrow$ quantization index = pixel intensity)

Contribution 5: Progressive Relative Quantization

Quantization Indices (4 bitplanes)

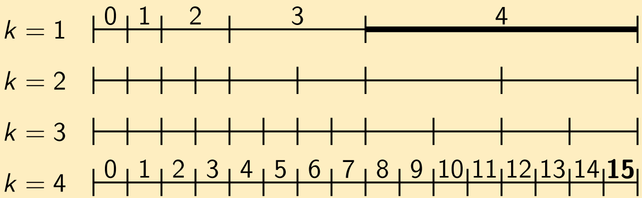


(Here, $k = 4 \Rightarrow$ quantization index = pixel intensity)

Ex 1 PRQ(**15**) = 4, right, right, right

Contribution 5: Progressive Relative Quantization

Quantization Indices (4 bitplanes)

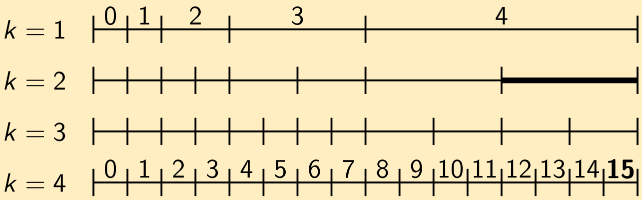


(Here, $k=4 \Rightarrow$ quantization index = pixel intensity)

Ex 1 PRQ(15) = **4**, right, right, right

Contribution 5: Progressive Relative Quantization

Quantization Indices (4 bitplanes)

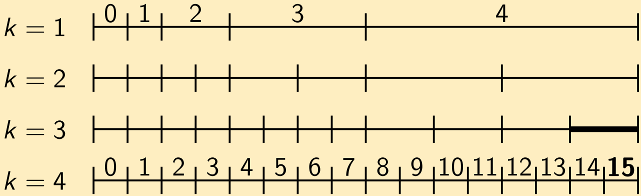


(Here, $k = 4 \Rightarrow$ quantization index = pixel intensity)

Ex 1 PRQ(15) = 4, **right**, right, right

Contribution 5: Progressive Relative Quantization

Quantization Indices (4 bitplanes)

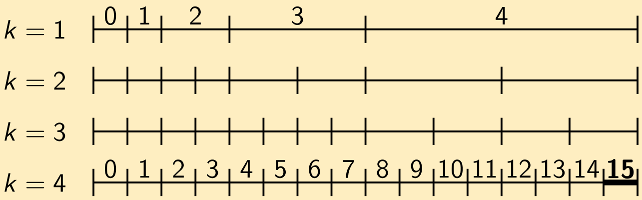


(Here, $k = 4 \Rightarrow$ quantization index = pixel intensity)

Ex 1 PRQ(15) = 4, right, **right**, right

Contribution 5: Progressive Relative Quantization

Quantization Indices (4 bitplanes)

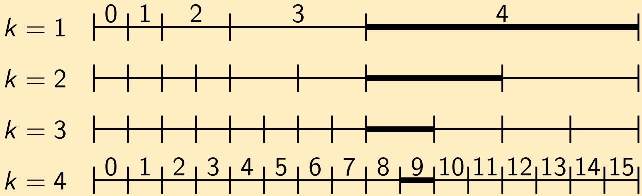


(Here, $k = 4 \Rightarrow$ quantization index = pixel intensity)

Ex 1 PRQ(15) = 4, right, right, **right**

Contribution 5: Progressive Relative Quantization

Quantization Indices (4 bitplanes)



(Here, $k = 4 \Rightarrow$ quantization index $=$ pixel intensity)

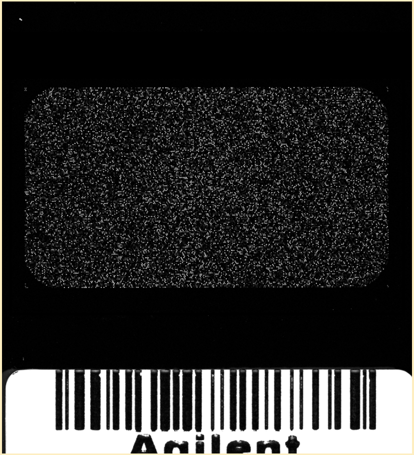
Ex 1 PRQ(15) $=$ 4, right, right, right

Ex 2 PRQ(9) $=$ 4, left, left, right

Contribution 5: Progressive Relative Quantization

How to code the PRQ representation?

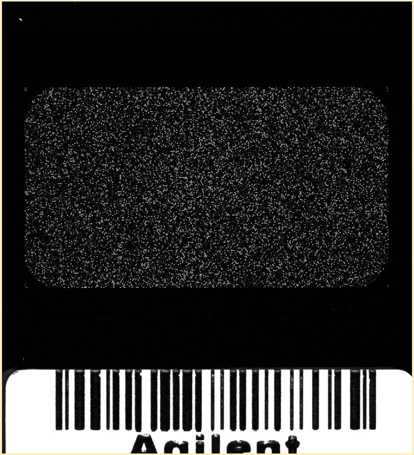
DNA Microarray Image



Contribution 5: Progressive Relative Quantization

How to code the PRQ representation?

DNA Microarray Image



PRQ

All image:

1. RQ($k = 1$)
2. $\Delta_{1\to 2}$
3. $\dots$
4. $\Delta_{15\to 16}$

Contribution 5: Progressive Relative Quantization

How to code the PRQ representation?

DNA Microarray Image

PRQ

All image:

1. RQ$(k = 1)$
2. $\Delta_{1\to 2}$
3. $\cdots$
4. $\Delta_{15\to 16}$

Contribution 5: Progressive Relative Quantization

How to code the PRQ representation?

DNA Microarray Image

PRQ

All image:

1. RQ($k = 1$)
2. $\Delta_{1\to 2}$
3. $\cdots$
4. $\Delta_{15\to 16}$

PRQ-ROI

First ROI:

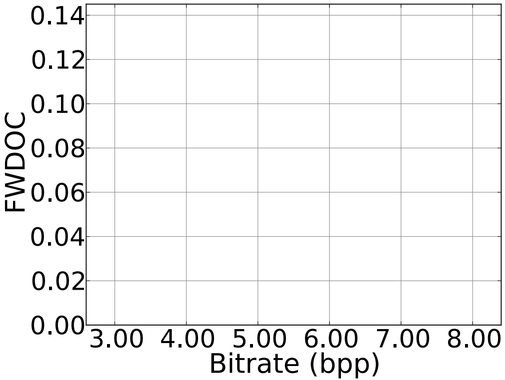
1. RQ($k = 1$)
2. $\Delta_{1\to 2}$
3. $\cdots$
4. $\Delta_{15\to 16}$

Then BG:

1. RQ($k = 1$)
2. $\Delta_{1\to 2}$
3. $\cdots$
4. $\Delta_{15\to 16}$

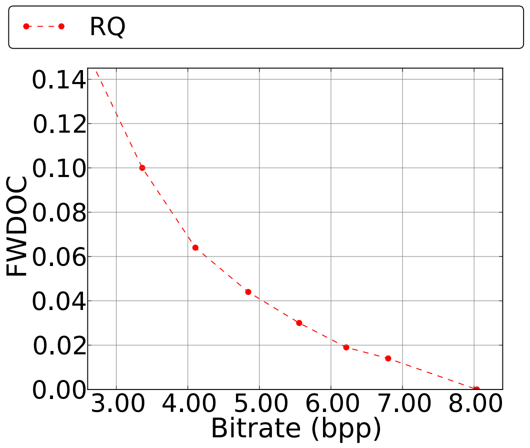
Contribution 5: Progressive Relative Quantizer

Rate-Distortion (Detection+Classification)



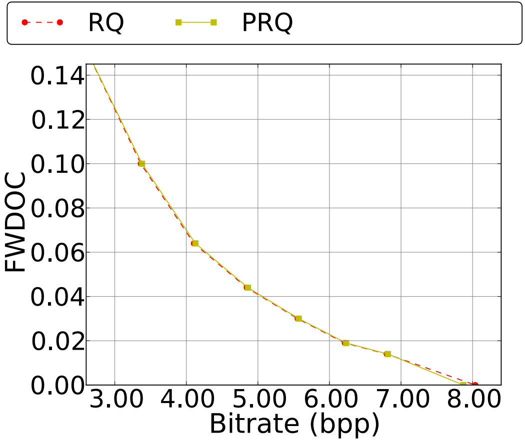
Contribution 5: Progressive Relative Quantizer

Rate-Distortion (Detection+Classification)



Contribution 5: Progressive Relative Quantizer

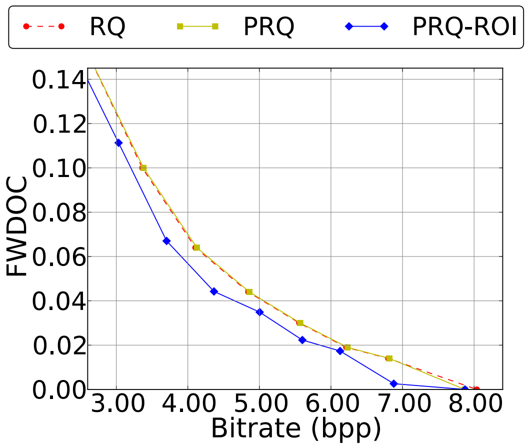
Rate-Distortion (Detection+Classification)



✓ PRQ $\sim$ RQ

Contribution 5: Progressive Relative Quantizer

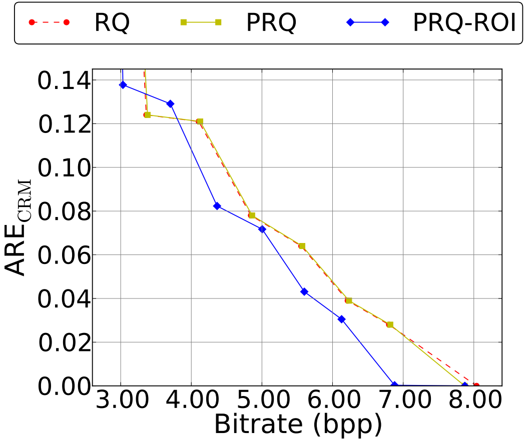
Rate-Distortion (Detection+Classification)



- ✓ PRQ $\sim$ RQ
- ✓ PRQ-ROI better than RQ,PRQ

Contribution 5: Progressive Relative Quantizer

Rate-Distortion (Detection+Classification)

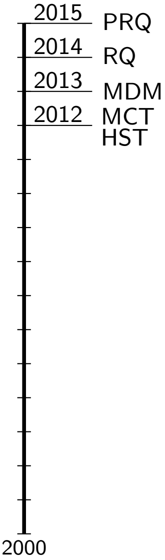


- ✓ PRQ $\sim$ RQ
- ✓ PRQ-ROI better than RQ,PRQ
- ✓ ARE_{CRM} and FWDOC

Contents

1. Introduction: DNA Microarrays
2. Lossless compression
3. Lossy compression
4. Conclusions

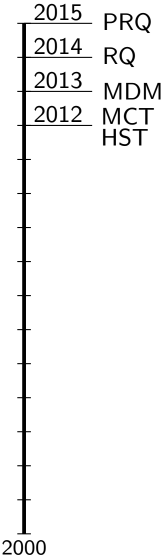
Problems and Goals



Problems

Lossless	**Lossy**
✓: Perfect data fidelity	✓: Arbitrary compression ratios
×: Small compression ratios	×: Analysis result distortion
×: Not standard	

Problems and Goals



Problems

Lossless

- ✓: Perfect data fidelity
- ×: Small compression ratios
- ×: Not standard

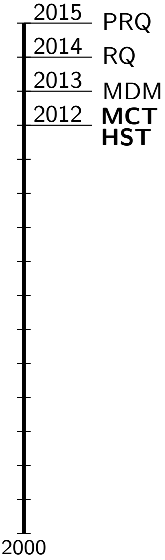
Lossy

- ✓: Arbitrary compression ratios
- ×: Analysis result distortion

Main goals

- ⇒ Improve standards
- ⇒ Assess distortion
- ⇒ Lossy coder with acceptable distortion

Problems and Goals



Problems

Lossless

- ✓: Perfect data fidelity
- ×: Small compression ratios
- ×: Not standard

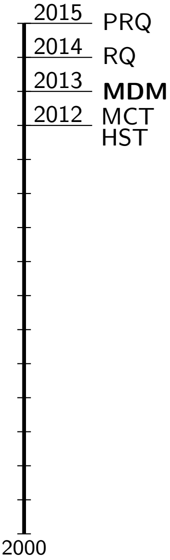
Lossy

- ✓: Arbitrary compression ratios
- ×: Analysis result distortion

Main goals

- ⇒ **Improve standards**
- ⇒ Assess distortion
- ⇒ Lossy coder with acceptable distortion

Problems and Goals



Problems

Lossless

- ✓: Perfect data fidelity
- ×: Small compression ratios
- ×: Not standard

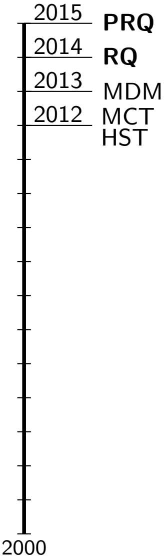
Lossy

- ✓: Arbitrary compression ratios
- ×: Analysis result distortion

Main goals

- ⇒ Improve standards
- ⇒ **Assess distortion**
- ⇒ Lossy coder with acceptable distortion

Problems and Goals



Problems

Lossless

- ✓: Perfect data fidelity
- ×: Small compression ratios
- ×: Not standard

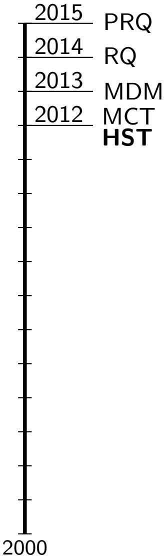
Lossy

- ✓: Arbitrary compression ratios
- ×: Analysis result distortion

Main goals

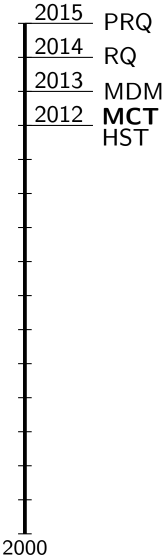
- ⇒ Improve standards
- ⇒ Assess distortion
- ⇒ **Lossy coder with acceptable distortion**

Contributions



- **Histogram Swap Transform**
  - Improvements 1.97%-15.53%
  - JPEG2000 close to microarray-specific

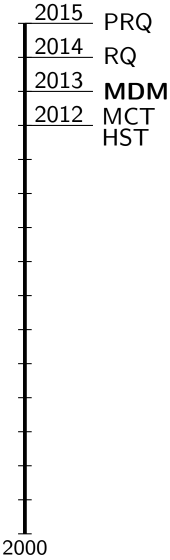
Contributions



- Histogram Swap Transform
  - Improvements 1.97%-15.53%
  - JPEG2000 close to microarray-specific

- **Multicomponent compression**
  - Best grouping: red/green pairs
  - No “golden” transform

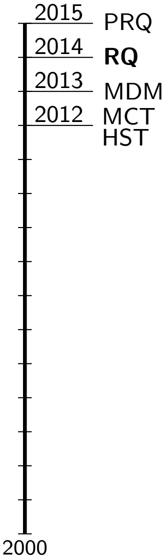
Contributions



- Histogram Swap Transform
  - Improvements 1.97%-15.53%
  - JPEG2000 close to microarray-specific
- Multicomponent compression
  - Best grouping: red/green pairs
  - No "golden" transform

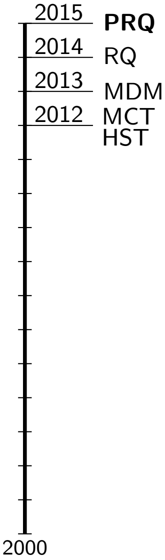
- **Microarray Distortion Metric**
  - Predict analysis distortion
  - Important vs unimportant changes

Contributions



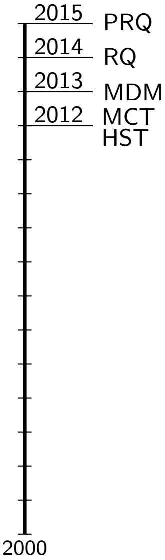
- Histogram Swap Transform
  - Improvements 1.97%-15.53%
  - JPEG2000 close to microarray-specific
- Multicomponent compression
  - Best grouping: red/green pairs
  - No "golden" transform
- Microarray Distortion Metric
  - Predict analysis distortion
  - Important vs unimportant changes
- **Relative Quantization**
  - Acceptable distortion
  - Good compression (1.5 bpp vs 8.3 bpp)

Contributions



- Histogram Swap Transform
  - Improvements 1.97%-15.53%
  - JPEG2000 close to microarray-specific
- Multicomponent compression
  - Best grouping: red/green pairs
  - No "golden" transform
- Microarray Distortion Metric
  - Predict analysis distortion
  - Important vs unimportant changes
- Relative Quantization
  - Acceptable distortion
  - Good compression (1.5 bpp vs 8.3 bpp)
- **Progressive RQ**
  - Lossy to lossless
  - Improved rate-distortion

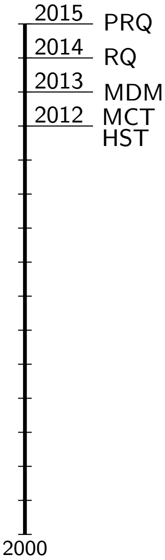
Publications



Conferences

- 1× Data Compression, Communication and Processing, CCP – 2011
- 1× Workshop on Multimedia Data Coding and Transmission, WMDCT – 2012
- **2**× Data Compression Conference, DCC – 2012, 2013

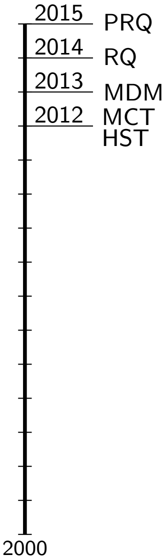
Publications



Conferences

- 1× Data Compression, Communication and Processing, CCP – 2011
- 1× Workshop on Multimedia Data Coding and Transmission, WMDCT – 2012
- **2**× Data Compression Conference, DCC – 2012, 2013 **(CORE A*)**

Publications



Conferences

1× Data Compression, Communication and Processing, CCP – 2011

1× Workshop on Multimedia Data Coding and Transmission, WMDCT – 2012

2× Data Compression Conference, DCC – 2012, 2013 **(CORE A*)**

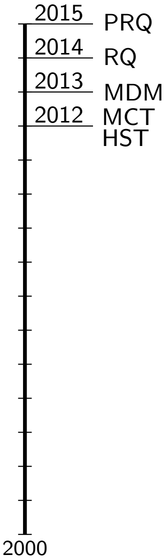
Journals

1× MDPI Algorithms – 2012

1× IEEE Transactions on Medical Imaging – 2015 (major revision)

1× IEEE Signal Processing Letters – 2015 (submitted)

Publications



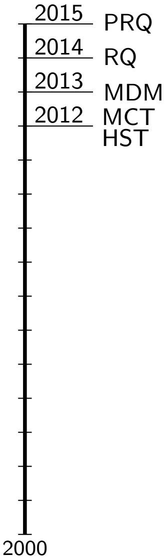
Conferences

- 1× Data Compression, Communication and Processing, CCP – 2011
- 1× Workshop on Multimedia Data Coding and Transmission, WMDCT – 2012
- **2**× Data Compression Conference, DCC – 2012, 2013 **(CORE A*)**

Journals

- 1× MDPI Algorithms – 2012
- 1× IEEE Transactions on Medical Imaging – 2015 (major revision) **(Q1)**
- 1× IEEE Signal Processing Letters – 2015 (submitted)

Publications



Conferences

- 1× Data Compression, Communication and Processing, CCP – 2011
- 1× Workshop on Multimedia Data Coding and Transmission, WMDCT – 2012
- **2**× Data Compression Conference, DCC – 2012, 2013 **(CORE A*)**

Journals

- 1× MDPI Algorithms – 2012
- 1× IEEE Transactions on Medical Imaging – 2015 (major revision) **(Q1)**
- 1× IEEE Signal Processing Letters – 2015 (submitted)

Book chapters

- 1× CRC Press – 2014

Reflections

Lossless compression

- Performance $\sim$ limit
- Standard - Specific: 0.5 bpp

$\Rightarrow$ Difficult to improve

Reflections

Lossless compression

- Performance $\sim$ limit
- Standard - Specific: 0.5 bpp

$\Rightarrow$ Difficult to improve

Lossy compression

- Image changes
  vs
  analysis distortion
- Margin to
  Improve rate-distortion

Reflections

Lossless compression

- Performance $\sim$ limit
- Standard - Specific: 0.5 bpp

$\Rightarrow$ Difficult to improve

Lossy compression

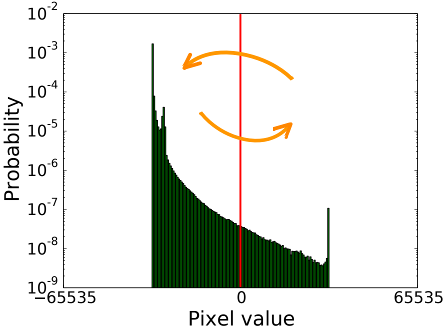
- Image changes
  vs
  analysis distortion
- Margin to
  Improve rate-distortion

Additional slides

- Image Analysis References
- Compression References
- HST Definition
- HST + DWT?
- MDM Definition
- RQ Definition
- RQ vs PRQ vs PRQ-ROI compression

HST definition

Left part	Right part
1^{st} bit was 0	1^{st} bit was 1
$\Downarrow$	$\Downarrow$
add 8000H	**subtract** 8000H



Histogram Swap Transform

Why not HST + DWT?

Histogram Swap Transform

Why not HST + DWT?

- Range expansion

Histogram Swap Transform

Why not HST + DWT?

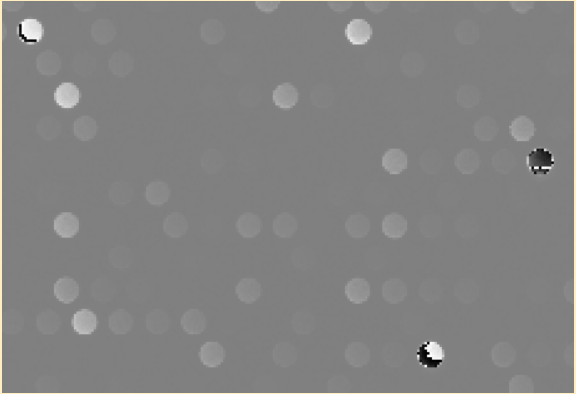
- Range expansion
- Higher % intensities used

Histogram Swap Transform

Why not HST + DWT?

- Range expansion
- Higher % intensities used
- Abrupt changes

Detail after HST



MDM definition

$$\mathrm{MDM} \quad = \quad 10 \log_{10} \frac{(\mathrm{max\_val})^2}{\mathrm{ME}}$$

$$\mathrm{ME} \quad = \quad (\mathrm{max\_val})^p - \mathrm{max\_val} + \min(\mathrm{max\_val}, \mathrm{MSE}_{\mathrm{image}}).$$

$$p \quad = \quad 2/(1 + exp(-\alpha(r_{\mathrm{spot}} + r_{\mathrm{localBG}} + r_{\mathrm{global}} - 3))),$$

$$\begin{aligned}
r_{\mathrm{spot}} &= \max(\mathrm{max_spot_ratio}, 1/\mathrm{min_spot_ratio}), \\
r_{\mathrm{localBG}} &= \max(\mathrm{max_localBG_ratio}, 1/\mathrm{min_localBG_ratio}), \\
r_{\mathrm{global}} &= \max(\mathrm{global_intensity_ratio}, 1/\mathrm{global_intensity_ratio}).
\end{aligned}$$

RQ Definition

- 2^k intervals of size 1
- 2^{k-1} intervals of size 2
- 2^{k-1} intervals of size 2^2
- $\cdots$
- 2^{k-1} intervals of size 2^{B-k}

Image Analysis References

- **Athanasiadis, 2015**
  Complementary DNA microarray image processing based on the fuzzy gaussian mixture model
- **Fouad, 2015**
  Automaticand Accurate Segmentation of Gridded cDNA Microarray Images Using Different Methods
- **Li, 2014**
  Improvements on segment based contours method for DNA microarray image segmentation
- **El-Gawady, 2014**
  Segmentation of Complementary DNA Microarray Images using Marker-Controlled Watershed Technique

Image Compression References

2003 **R. Jornsten et al.**
Microarray image compression: SLOCO and the effect of information loss,
Signal Processing, vol. 83, no. 4, pp. 859–869

2003 **Faramarzpour, N. et al.**
Lossless DNA microarray image compression
Signals, Systems and Computers

2004 **Hua, J. et al.**
Microarray BASICA: Background Adjustment, Segmentation, Image Compression and Analysis of Microarray Images.
EURASIP Journal on Advances in Signal Processing

2004 **Lonardi & Luo**
Gridding and compression of microarray images.
IEEE Computational Systems Bioinformatics Conference, CSB

2006 **Pinho et al.**
On the use of standards for microarray lossless image compression. IEEE Transactions on Biomedical Engineering

2006 **Bierman, R. et al.**
MACE: Lossless Compression and Analysis of Microarray Images,
ACM Symposium on Applied Computing, 23–27

2006 **Xu, Q. et al.**
Accuracy of differential expression detection with compressed microarray images
Signal Image and Video Processing

2007 **Neekabadi, a. et al.**
Lossless Microarray Image Compression using Region Based Predictors.
IEEE International Conference on Image Processing

2007 **Peters, T. J. et al.**
Microarray image compression using a variation of singular value decomposition.
IEEE Engineering in Medicine and Biology Society Conference

2009 **Neves, A. et al.**
Lossless compression of microarray images using image-dependent finite-context models.
IEEE Transactions on Medical Imaging

2009 **Battiato, S. et al.**
A bio-inspired CNN with re-indexing engine for lossless DNA microarray compression and segmentation.
IEEE International Conference on Image Processing (ICIP)

2011 **Avanaki et al.**
Compression of cDNA Microarray Images based on Pure-Fractal and Wavelet- Fractal Techniques
Journal on Graphics, Vision and Image Processing

2014 **Rueda, L. (editor)**
Microarray Image and Data Analysis: Theory and Practice.
CRC Press

RQ vs PRQ vs PRQ-ROI

Compression results (IBB, bpp)

Algorithm	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 6$	$k = 7$	Original
RQ	2.653	3.363	4.105	4.844	5.556	6.214	6.800	8.039
PRQ	2.652	3.375	4.119	4.857	5.566	6.230	6.815	8.001
PRQ-ROI	2.380	3.032	3.703	4.365	5.004	5.600	6.128	6.878

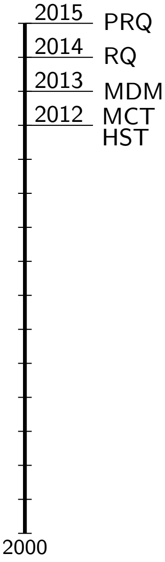
Analysis results (IBB, ARE_{CRM})

Algorithm	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 6$	$k = 7$	Original
RQ	0.526	0.136	0.123	0.073	0.076	0.041	0.029	0
PRQ-ROI	1.356	0.146	0.135	0.076	0.080	0.044	0.032	0

Additional slides

- Image Analysis References
- Compression References
- HST Definition
- HST + DWT?
- MDM Definition
- RQ Definition
- RQ vs PRQ vs PRQ-ROI compression

Contribution summary



- **Histogram Swap Transform**
  - Improvements 1.97%-15.53%
  - JPEG2000 close to microarray-specific
- **Multicomponent compression**
  - Best grouping: red/green pairs
  - No “golden” transform
- **Microarray Distortion Metric**
  - Predict analysis distortion
  - Important vs unimportant changes
- **Relative Quantization**
  - Acceptable distortion
  - Good compression
- **Progressive RQ**
  - Lossy to lossless
  - Improved rate-distortion