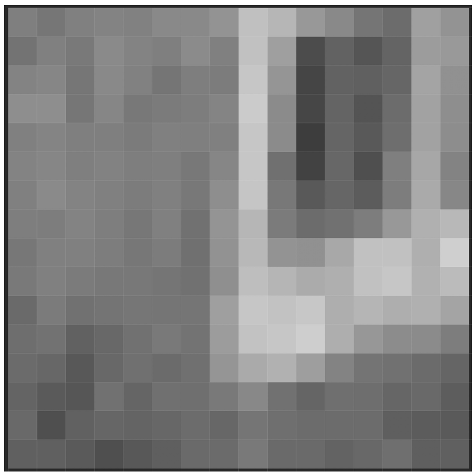
Applications of Information Theory in Image Processing

3. Modeling continuous tone images

Continuous tone images

- *Grayscale*: 2D array of *pixel intensity values* (integers) in a given range $[\,0\ ..\ (\alpha\ -1)]$ (e.g. $\alpha = 256$)

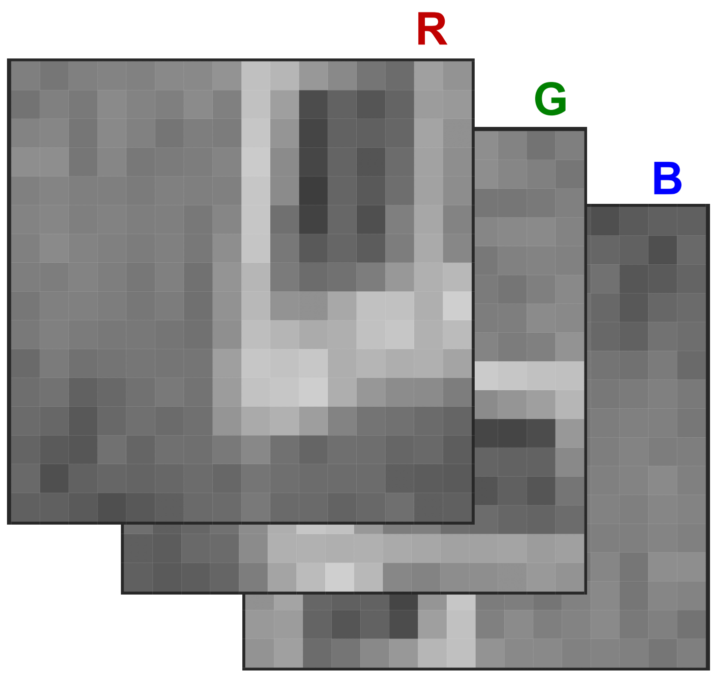


68	5F	6A	6D	6B	73	73	7D	B0	A5	83	73	5E	56	8C	7E
5C	6A	62	74	6D	68	75	69	B2	8B	38	4C	40	4E	87	84
6D	70	5F	73	6B	5E	67	65	B7	80	32	4C	4A	50	90	7B
78	78	5F	70	61	64	66	6E	BD	75	33	4E	3F	56	8E	78
6A	6E	68	68	64	6A	68	69	B8	75	2C	4F	44	57	8E	77
6D	70	68	6C	67	68	61	70	B7	59	30	51	3B	68	94	6D
69	74	6D	6A	65	6A	61	78	B6	61	44	50	46	66	97	71
67	66	6D	67	60	69	5A	7F	A5	64	56	5A	66	83	9E	A7
60	6A	6A	66	60	65	59	7C	A7	7D	7B	95	B2	B2	9D	C2
62	6A	64	61	61	5E	5A	79	AE	A3	98	9D	B2	B7	9F	AA
54	64	5A	5D	5F	5E	5E	8B	B8	B3	B9	9C	A3	9C	9E	90
57	5B	4B	52	5A	62	5C	87	B3	B7	C1	9C	82	76	75	66
55	51	43	52	59	55	59	80	97	9F	89	6D	5D	5A	55	4F
4E	45	41	5A	4F	59	58	62	72	5A	4F	58	57	50	53	47
53	3B	4B	50	4E	51	56	51	5E	58	56	56	56	48	46	45
4A	49	45	3B	43	48	54	55	62	54	54	4D	52	58	48	49

- Notation. $\mathbf{x}$: an image,
  x_i: i-th pixel of $\mathbf{x}$ in some linear order (e.g. raster),
  $x_{i,j}$: pixel in (i, j) location when using 2D indexing.

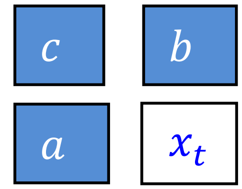
Continuous tone images

- *Color*: a 2D array of *vectors* (e.g. triplets) whose coordinates represent intensity in a given *color space* (e.g., RGB, YUV); similar principles
  - Alternative interpretation: a vector of grayscale images, one per color component (e.g., R, G, B).



  - Other color spaces: YUV, Lab, CMYK, etc.

The Challenge

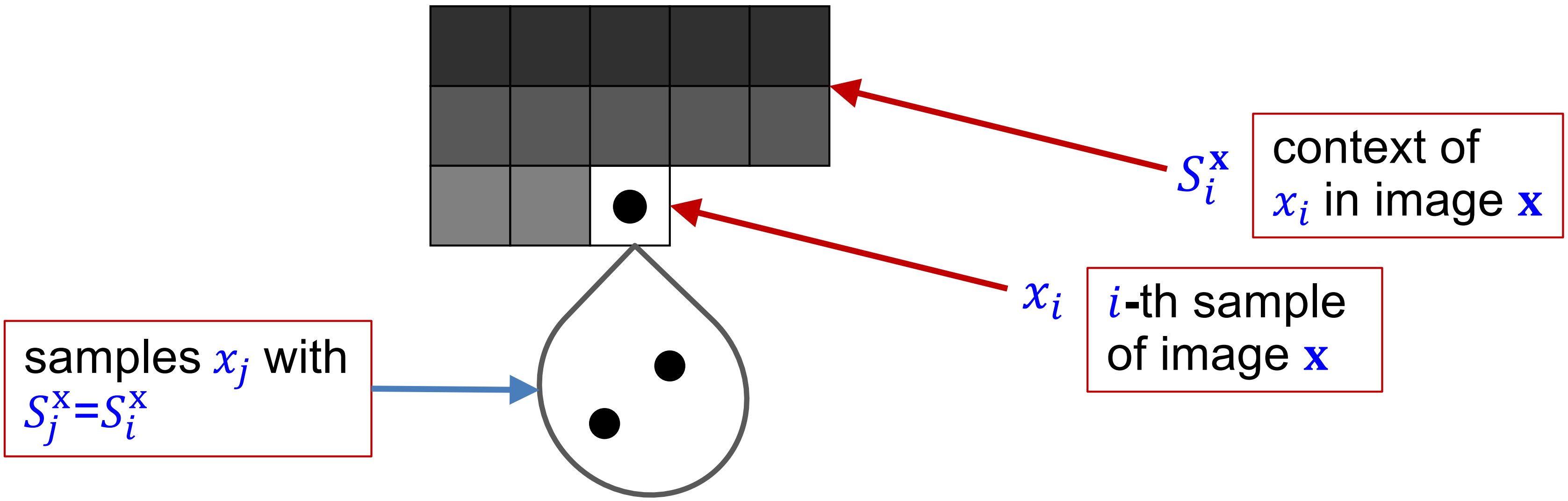


- Even with almost minimal 2D context (8-bit samples, 3 closest samples, grayscale), we have $2^{24} \approx 17 \cdot 10^6$ possible different context patterns
  - gets astronomical very quickly: *big numbers = big problem*

Modeling grayscale images

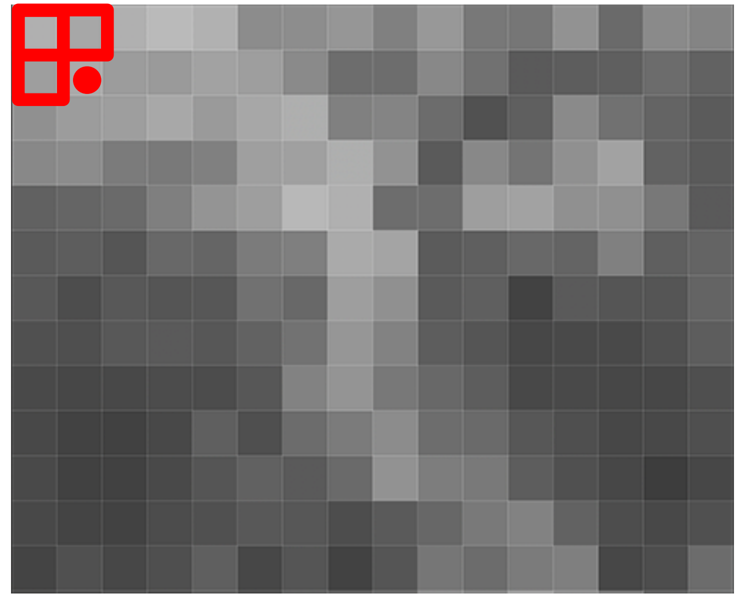
- *Natural images*
  - images with a relatively large dynamic range (e.g. 256 grayscale values): *large alphabets*
  - main assumption: *numerical sample values preserve continuity of brightness in the physical world* (up to quantization)

- *Main challenge*
  - even for small contexts, number of potential context patterns is huge
  - patterns are not likely to repeat exactly $\Rightarrow$ very few samples will occur in the same context $\Rightarrow$ *sparse statistics*



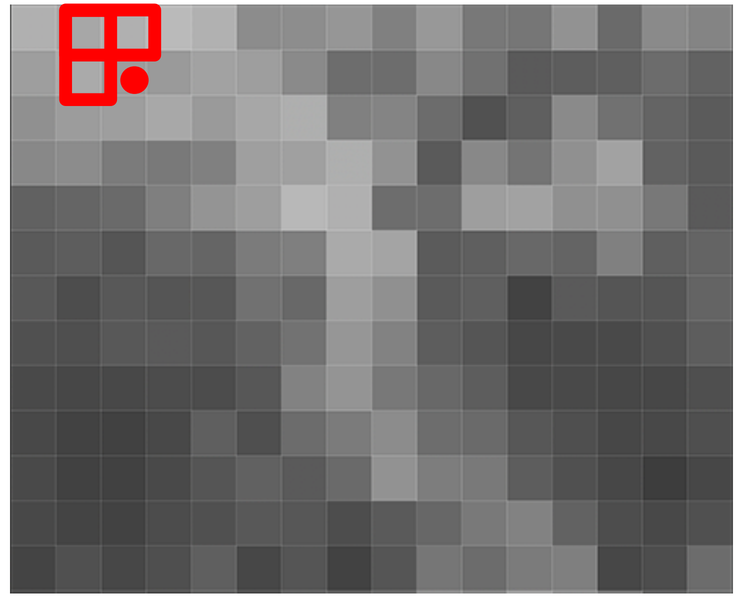
Example: 3-pixel context

720x576 grayscale image
414720 pixels



33	35	37	33	35	2d	2c	2a	2f	2e	2f	30	4f	42	62	77
2f	37	2f	33	30	2e	2c	2e	2f	2c	32	38	46	5a	8a	a2
35	2f	2b	36	31	32	28	2e	30	4c	3e	45	5e	80	a2	c1
43	2b	2f	3e	3c	2d	27	24	28	58	57	41	56	79	ae	ca
51	32	3c	50	4f	30	26	24	27	38	5a	41	32	48	a3	cc
5a	32	39	5e	5d	4c	30	29	2b	45	70	4d	38	42	9d	cc
6b	47	34	5c	62	5a	48	42	42	5a	6c	4a	35	57	b1	cf
79	63	43	3a	54	62	56	56	5b	53	48	39	42	7e	c5	d1
7f	73	58	3b	3c	4f	57	5a	50	4d	37	41	71	ad	cf	d3
72	7d	71	59	3e	39	36	3d	3a	41	4d	6b	a0	c6	d0	ce
70	75	78	71	5f	52	3f	42	4b	58	7f	9e	bc	ca	cb	ca
77	77	7c	84	78	6c	65	66	74	89	a3	ba	c5	c6	ca	cd
6e	77	82	8b	8d	83	85	8a	93	9e	b6	bb	c1	c5	c7	cb

Example: 3-pixel context

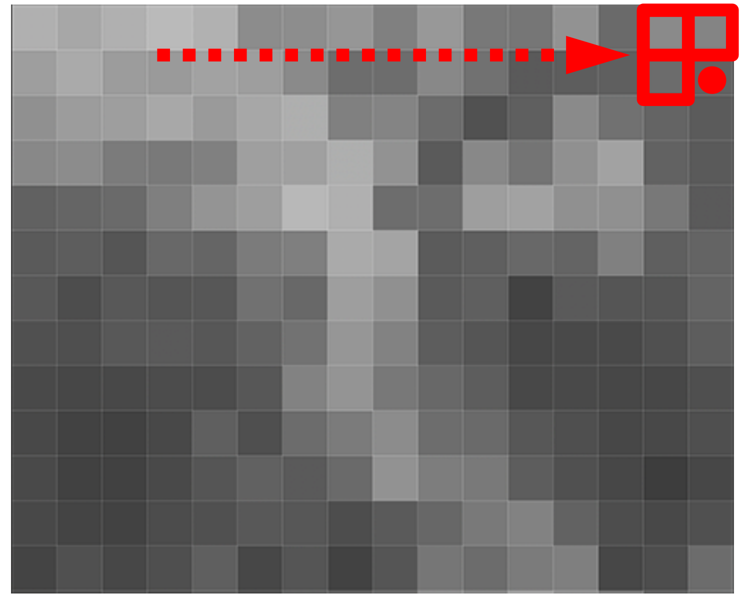


720x576 grayscale image
414720 pixels

33	35	37	33	35	2d	2c	2a	2f	2e	2f	30	4f	42	62	77
2f	37	2f	33	30	2e	2c	2e	2f	2c	32	38	46	5a	8a	a2
35	2f	2b	36	31	32	28	2e	30	4c	3e	45	5e	80	a2	c1
43	2b	2f	3e	3c	2d	27	24	28	58	57	41	56	79	ae	ca
51	32	3c	50	4f	30	26	24	27	38	5a	41	32	48	a3	cc
5a	32	39	5e	5d	4c	30	29	2b	45	70	4d	38	42	9d	cc
6b	47	34	5c	62	5a	48	42	42	5a	6c	4a	35	57	b1	cf
79	63	43	3a	54	62	56	56	5b	53	48	39	42	7e	c5	d1
7f	73	58	3b	3c	4f	57	5a	50	4d	37	41	71	ad	cf	d3
72	7d	71	59	3e	39	36	3d	3a	41	4d	6b	a0	c6	d0	ce
70	75	78	71	5f	52	3f	42	4b	58	7f	9e	bc	ca	cb	ca
77	77	7c	84	78	6c	65	66	74	89	a3	ba	c5	c6	ca	cd
6e	77	82	8b	8d	83	85	8a	93	9e	b6	bb	c1	c5	c7	cb

Example: 3-pixel context

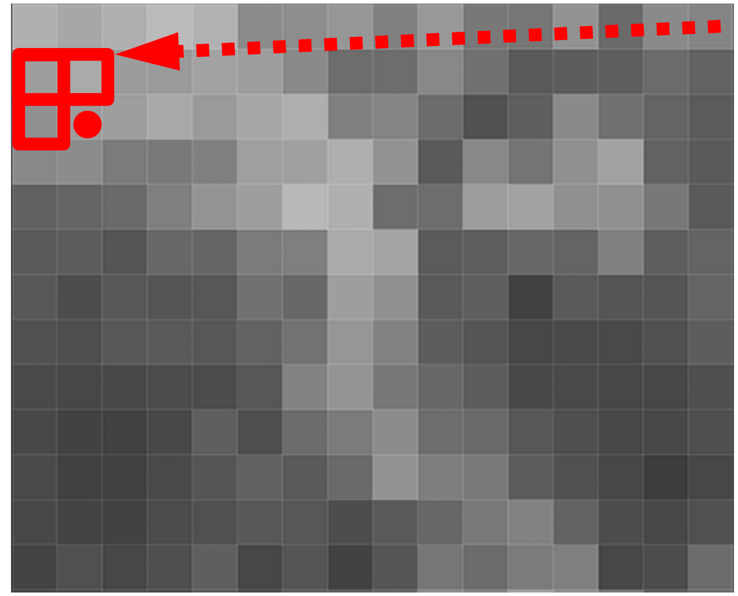
720x576 grayscale image
414720 pixels



33	35	37	33	35	2d	2c	2a	2f	2e	2f	30	4f	42	62	77
2f	37	2f	33	30	2e	2c	2e	2f	2c	32	38	46	5a	8a	a2
35	2f	2b	36	31	32	28	2e	30	4c	3e	45	5e	80	a2	c1
43	2b	2f	3e	3c	2d	27	24	28	58	57	41	56	79	ae	ca
51	32	3c	50	4f	30	26	24	27	38	5a	41	32	48	a3	cc
5a	32	39	5e	5d	4c	30	29	2b	45	70	4d	38	42	9d	cc
6b	47	34	5c	62	5a	48	42	42	5a	6c	4a	35	57	b1	cf
79	63	43	3a	54	62	56	56	5b	53	48	39	42	7e	c5	d1
7f	73	58	3b	3c	4f	57	5a	50	4d	37	41	71	ad	cf	d3
72	7d	71	59	3e	39	36	3d	3a	41	4d	6b	a0	c6	d0	ce
70	75	78	71	5f	52	3f	42	4b	58	7f	9e	bc	ca	cb	ca
77	77	7c	84	78	6c	65	66	74	89	a3	ba	c5	c6	ca	cd
6e	77	82	8b	8d	83	85	8a	93	9e	b6	bb	c1	c5	c7	cb

Example: 3-pixel context

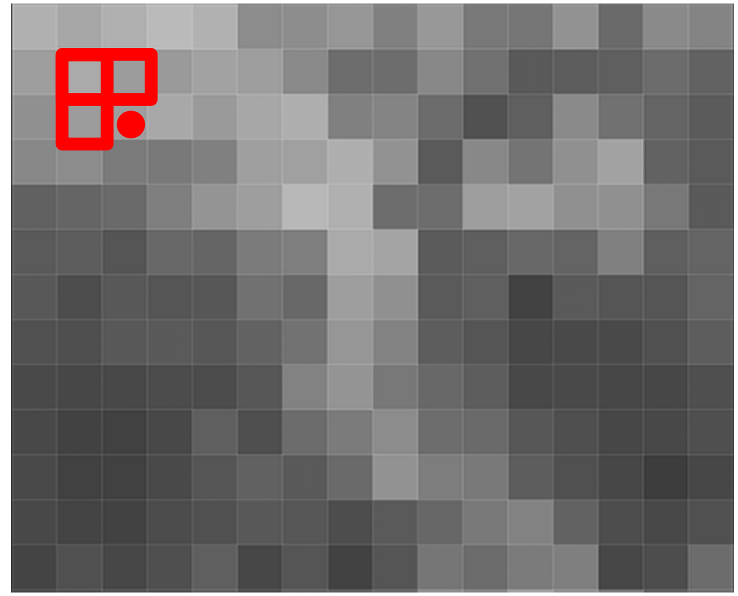
720x576 grayscale image
414720 pixels



33	35	37	33	35	2d	2c	2a	2f	2e	2f	30	45	42	62	77
2f	37	2f	33	30	2e	2c	2e	2f	2c	32	38	46	5a	8a	a2
35	2f	2b	36	31	32	28	2e	30	4c	3e	45	5e	80	a2	c1
43	2b	2f	3e	3c	2d	27	24	28	58	57	41	56	79	ae	ca
51	32	3c	50	4f	30	26	24	27	38	5a	41	32	48	a3	cc
5a	32	39	5e	5d	4c	30	29	2b	45	70	4d	38	42	9d	cc
6b	47	34	5c	62	5a	48	42	42	5a	6c	4a	35	57	b1	cf
79	63	43	3a	54	62	56	56	5b	53	48	39	42	7e	c5	d1
7f	73	58	3b	3c	4f	57	5a	50	4d	37	41	71	ad	cf	d3
72	7d	71	59	3e	39	36	3d	3a	41	4d	6b	a0	c6	d0	ce
70	75	78	71	5f	52	3f	42	4b	58	7f	9e	bc	ca	cb	ca
77	77	7c	84	78	6c	65	66	74	89	a3	ba	c5	c6	ca	cd
6e	77	82	8b	8d	83	85	8a	93	9e	b6	bb	c1	c5	c7	cb

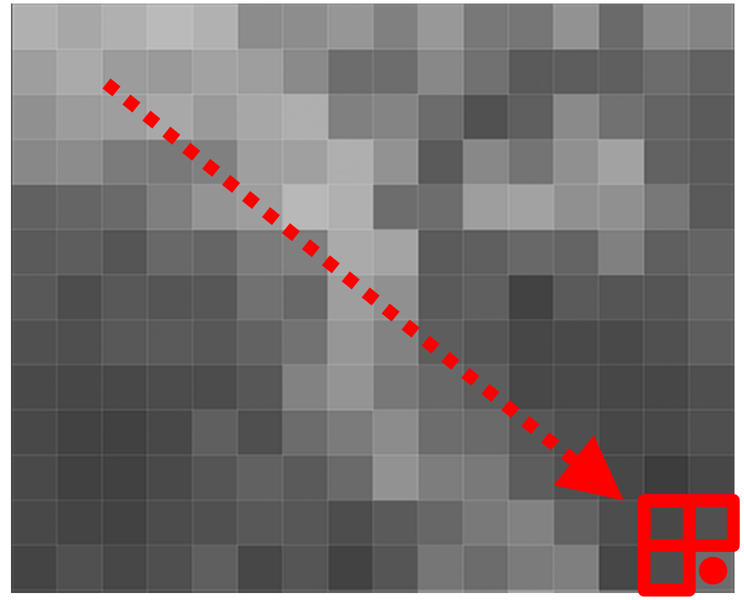
Example: 3-pixel context

720x576 grayscale image
414720 pixels



33	35	37	33	35	2d	2c	2a	2f	2e	2f	30	4f	42	62	77
2f	37	2f	33	30	2e	2c	2e	2f	2c	32	38	46	5a	8a	a2
35	2f	2b	36	31	32	28	2e	30	4c	3e	45	5e	80	a2	c1
43	2b	2f	3e	3c	2d	27	24	28	58	57	41	56	79	ae	ca
51	32	3c	50	4f	30	26	24	27	38	5a	41	32	48	a3	cc
5a	32	39	5e	5d	4c	30	29	2b	45	70	4d	38	42	9d	cc
6b	47	34	5c	62	5a	48	42	42	5a	6c	4a	35	57	b1	cf
79	63	43	3a	54	62	56	56	5b	53	48	39	42	7e	c5	d1
7f	73	58	3b	3c	4f	57	5a	50	4d	37	41	71	ad	cf	d3
72	7d	71	59	3e	39	36	3d	3a	41	4d	6b	a0	c6	d0	ce
70	75	78	71	5f	52	3f	42	4b	58	7f	9e	bc	ca	cb	ca
77	77	7c	84	78	6c	65	66	74	89	a3	ba	c5	c6	ca	cd
6e	77	82	8b	8d	83	85	8a	93	9e	b6	bb	c1	c5	c7	cb

Example: 3-pixel context



33	35	37	33	35	2d	2c	2a	2f	2e	2f	30	4f	42	62	77
2f	37	2f	33	30	2e	2c	2e	2f	2c	32	38	46	5a	8a	a2
35	2f	3b	36	31	32	28	2e	30	4c	3e	45	5e	80	a2	c1
43	2b	2f	3e	3c	2d	27	24	28	58	57	41	56	79	ae	ca
51	32	3c	50	4f	30	26	24	27	38	5a	41	32	48	a3	cc
5a	32	39	5e	5d	4c	30	29	2b	45	70	4d	38	42	9d	cc
6b	47	34	5c	62	5a	48	42	42	5a	6c	4a	35	57	b1	cf
79	63	43	3a	54	62	56	56	5b	53	48	39	42	7e	c5	d1
7f	73	58	3b	3c	4f	57	5a	50	4d	35	41	71	ad	cf	d3
72	7d	71	59	3e	39	36	3d	3a	41	4d	6b	a0	c6	d0	ce
70	75	78	71	5f	52	3f	42	4b	58	7f	9e	[illegible]	[illegible]	cb	ca
77	77	7c	84	78	6c	65	66	74	89	a3	ba	c5	c6	ca	cd
6e	77	82	8b	8d	83	85	8a	93	9e	b6	bb	c1	c5	c7	cb

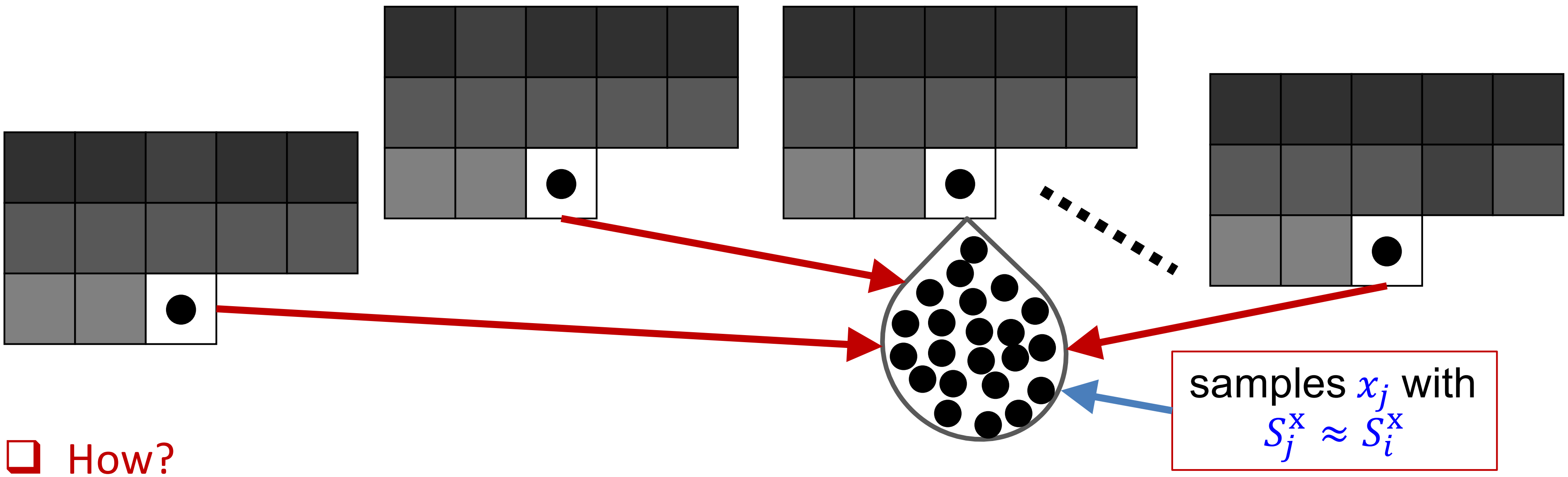
720x576 grayscale image
414720 pixels

Number of distinct contexts: 164303

Compression ratio: ≈ 7.5 bits/pixel
(≈ 4 bpp would be typical for a decent image compressor)

Modeling grayscale images

- *Key to the solution*
  - let contexts share and aggregate their information
  - learn additional information about the distribution of, say, x_i given its context from occurrences of samples x_j, depending on how "close" S_j^{x} is to S_i^{x} in an appropriate sense

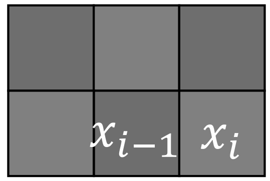


- How?
  - incorporate *a priori knowledge* of the properties of images into the model
  - *do not impose on the modeling unit the task to "learn" what we already know!*

Properties of natural images

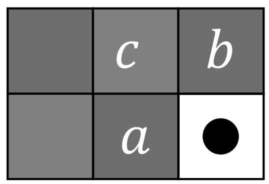
❑ *Smoothness* (1)

- brightness (generally) varies smoothly; values of neighboring samples are likely to be close



$$x_i \approx x_{i-1}$$

$\Rightarrow$ it is often possible to predict quite accurately the value of a sample given its context



$$\hat{x}_i = f(a, b, c, \dots)$$

predicted value

deterministic function (*predictor*)

  - simplest predictor: $\hat{x}_i = x_{i-1}$

- *Prediction error (residual)*: $e_i = x_i - \hat{x}_i$.
  If the predictor is decent, e_i tends to follow a distribution peaked at zero, well modeled by a *discrete Laplacian* or *two-sided geometric distribution (TSGD)*

$$P(e_i) = c \cdot \theta^{|e_i|}, \qquad 0 < \theta < 1$$

Statistics of grayscale images

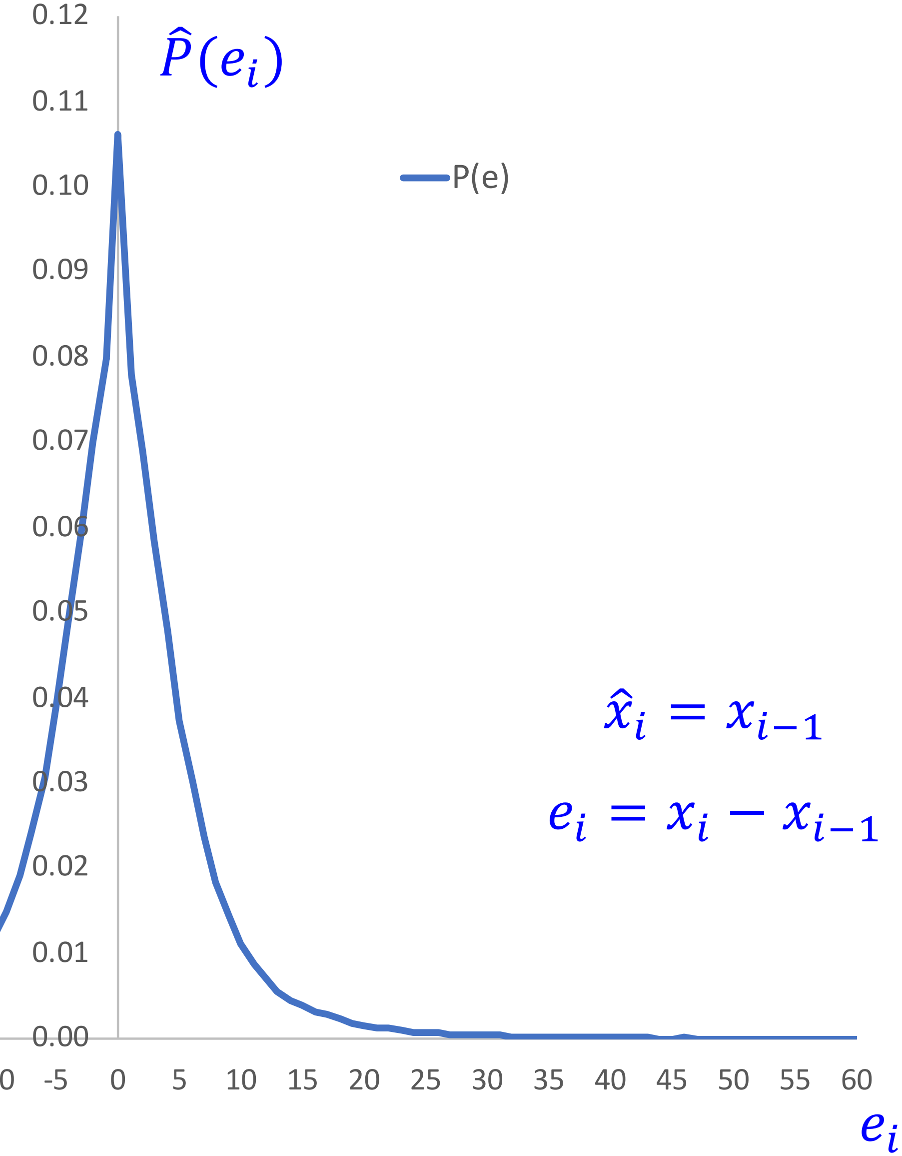
Grayscale image: 720x576 8-bit pixels

Zero order empirical distribution

Empirical $\widehat{H} = 7.53$ bits/pixel

Statistics of prediction errors

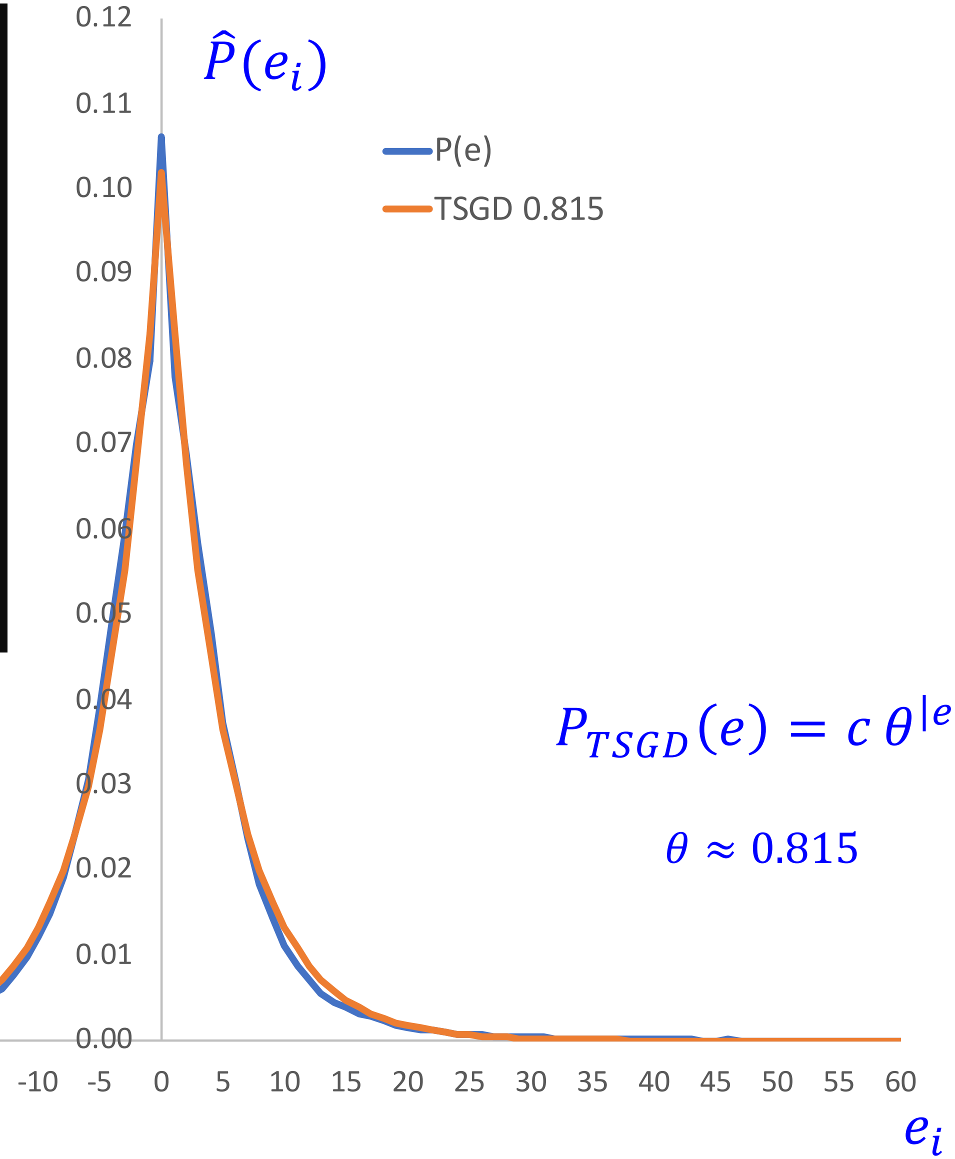
global stats for image
size: 720x576



$$\hat{x}_i = x_{i-1}$$

$$e_i = x_i - x_{i-1}$$

Statistics of prediction errors



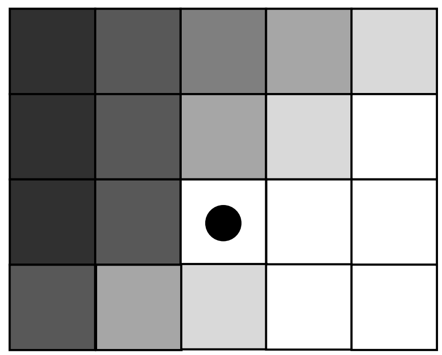
global stats for image size: 720x576

$$P_{TSGD}(e) = c\,\theta^{|e|}$$

$$\theta \approx 0.815$$

Using prediction with context models

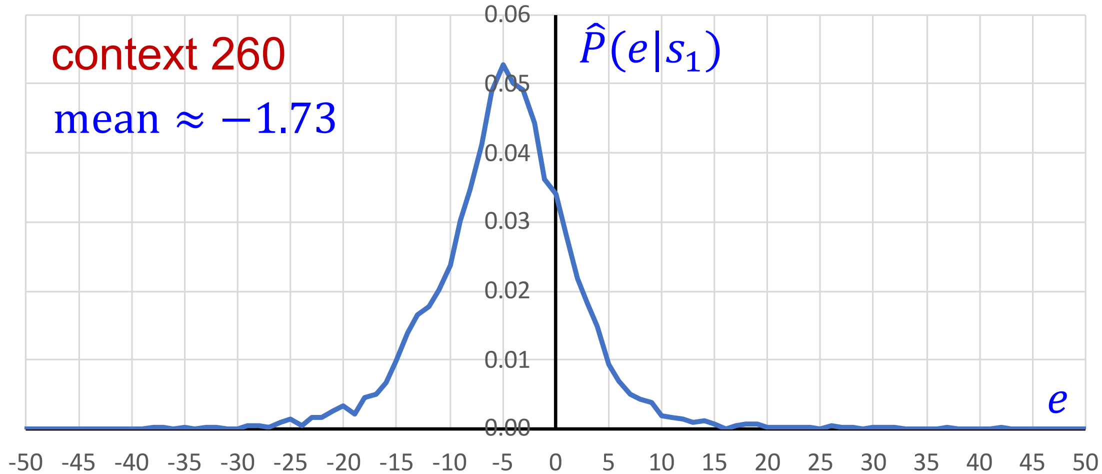
- In the previous example, we looked at statistics of prediction errors accumulated over the whole image.
- What happens if we look at distributions of prediction errors *conditioned on contexts*?

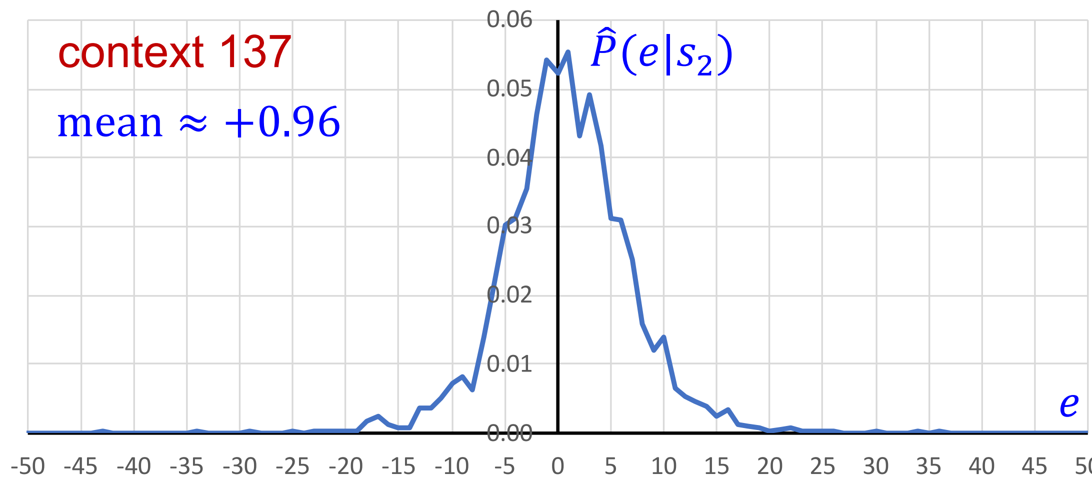


  - If, for example, we are in a region exhibiting a strong gradient in brightness, the predictor may not be able to fully capture the pattern, and it may show *biases* (due to causality, or other predictor limitations).

Bias in context-conditioned distributions

Grayscale image 720x576

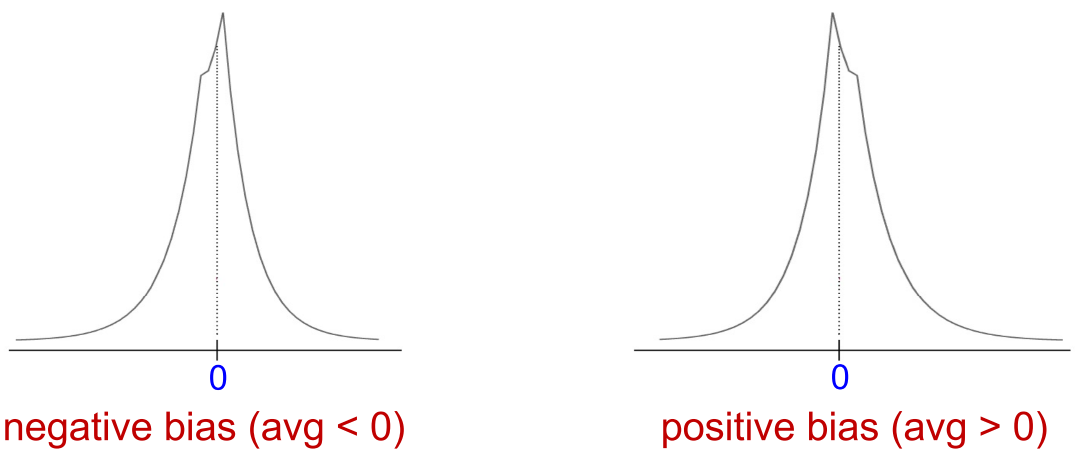




Distributions of prediction errors in two different contexts

Bias in context-conditioned distributions

- The symmetric TSGD is a good approximation for the *global* prediction error distribution
- However, context-conditioned prediction error distributions sometimes exhibit *context-dependent biases*



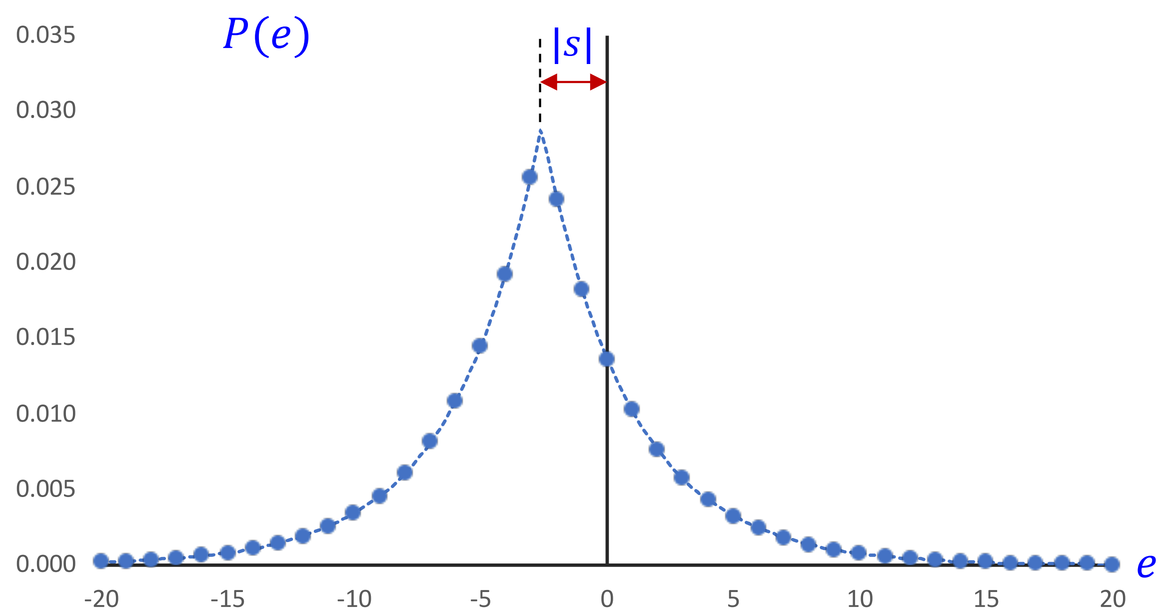
Bias in context-conditioned distributions

- *More accurate TSGD model for context-conditioned prediction errors*:

$$P_{\theta,s}(e) = c_{\theta,s}\theta^{|e-s|}, \qquad 0 < \theta < 1, \qquad s \in \mathbb{R},$$

normalization

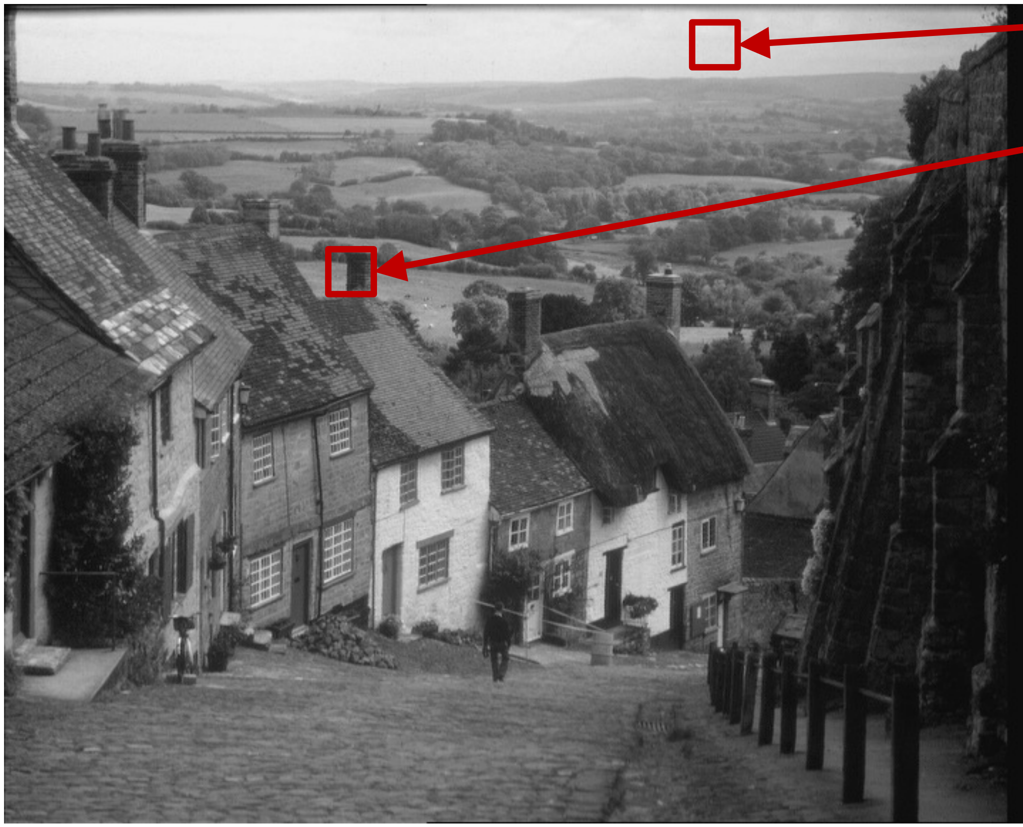
- two-parameter model : θ (*decay*) and s (*bias*), offers *model cost* advantages



Being careful with smoothness assumptions

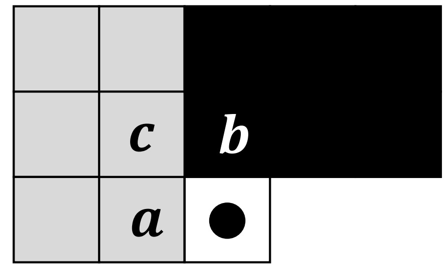
great here

Being careful with smoothness assumptions



great here

but not here



a linear predictor like

$$\hat{x} = \frac{a+b+c}{3}$$

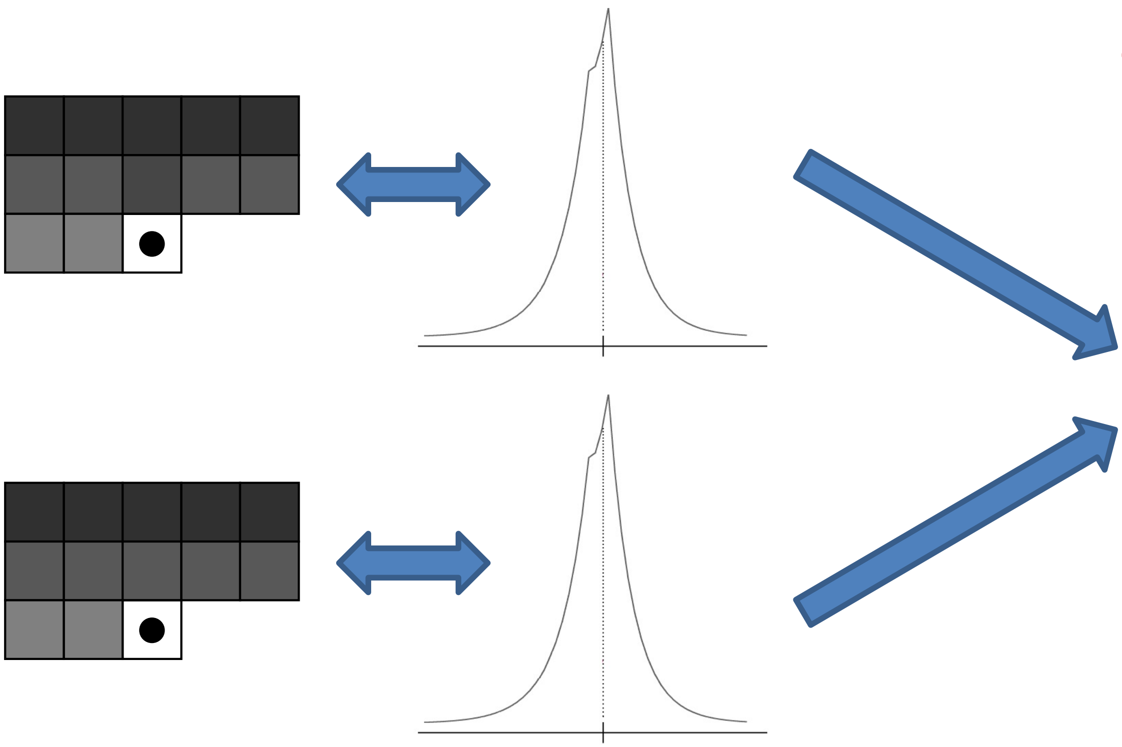
would not be a good idea here!

if possible, avoid interpolating over sharp edges

Properties of grayscale images (cont.)

- *Smoothness* (2)
  - contexts that are close (e.g., as vectors, in Euclidean distance) will tend to produce similar conditional probability distributions
  - in a sense, this generalizes the first smoothness property: we postulate not only that *pixel values* vary smoothly, but that *conditional distributions vary smoothly with the context*

context quantization

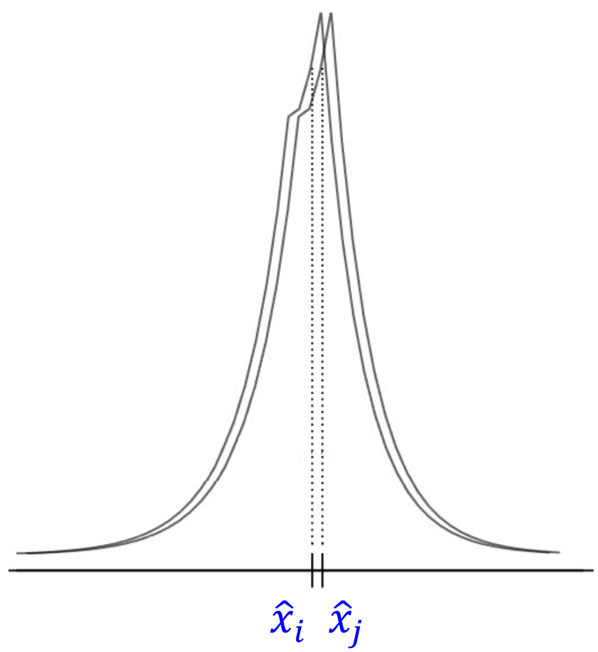


merge counts

potential problem: similar shapes, slightly misaligned centers: *blurring*

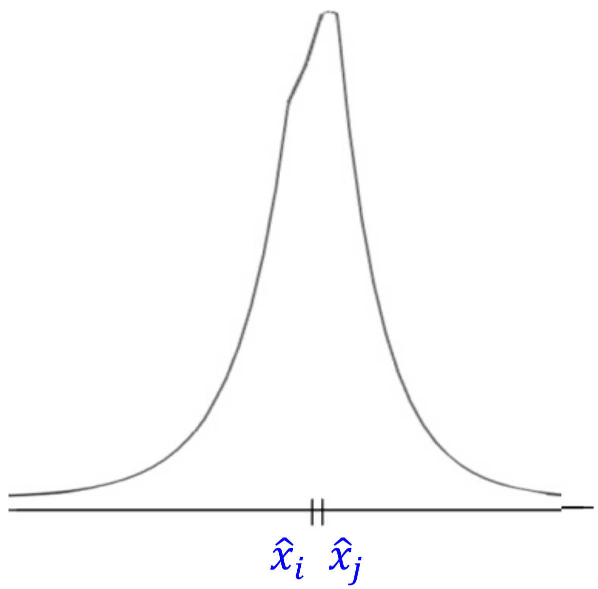
Properties of grayscale images (cont.)

- *DC Invariance* (1)
  - merged distributions might be slightly misaligned
  - *solution:* align distributions by *centering at the predicted value* for the context (equivalently: gather *statistics of prediction errors*)



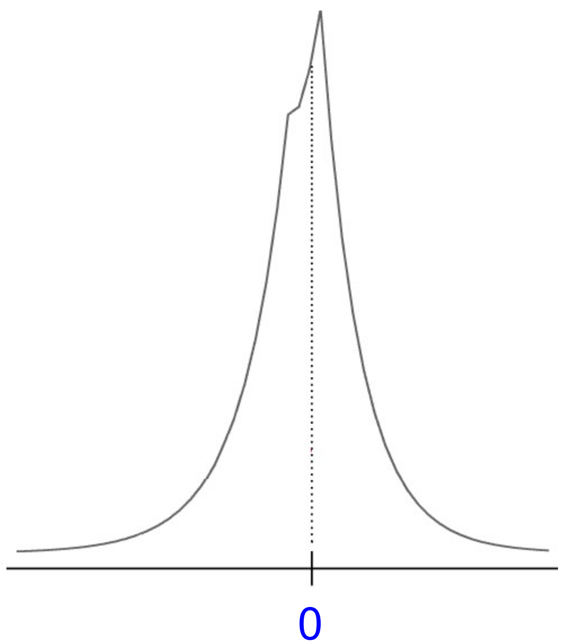
Properties of grayscale images (cont.)

- *DC Invariance* (1)
  - merged distributions might be slightly misaligned
  - *solution*: align distributions by *centering at the predicted value* for the context (equivalently: gather *statistics of prediction errors*)



Properties of grayscale images (cont.)

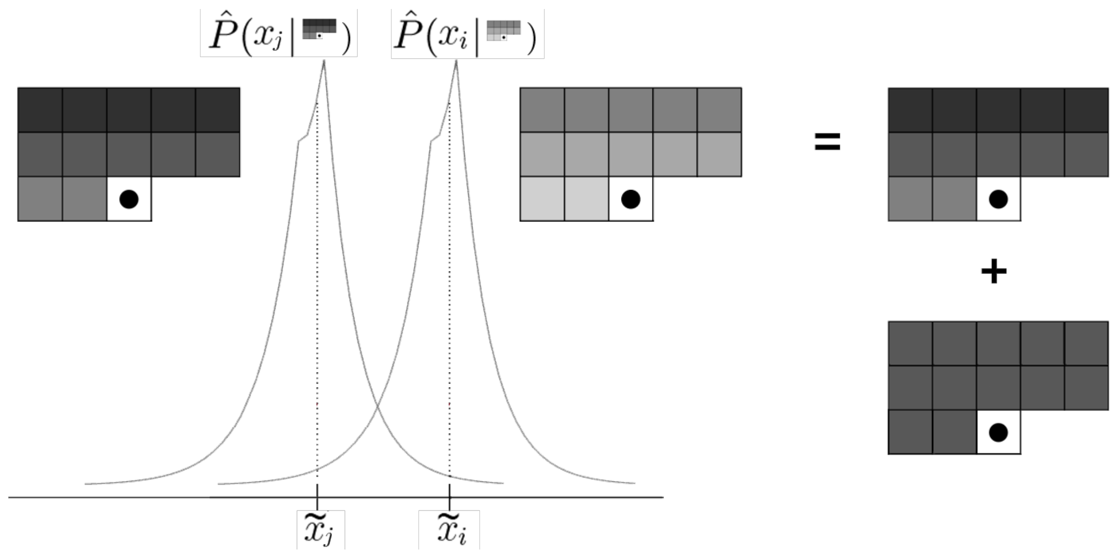
- *DC Invariance* (1)
  - merged distributions might be slightly misaligned
  - *solution:* align distributions by *centering at the predicted value* for the context (equivalently: gather *statistics of prediction errors*)



  - generally, "sharper" distribution ⇒ lower entropy

Properties of grayscale images (cont.)

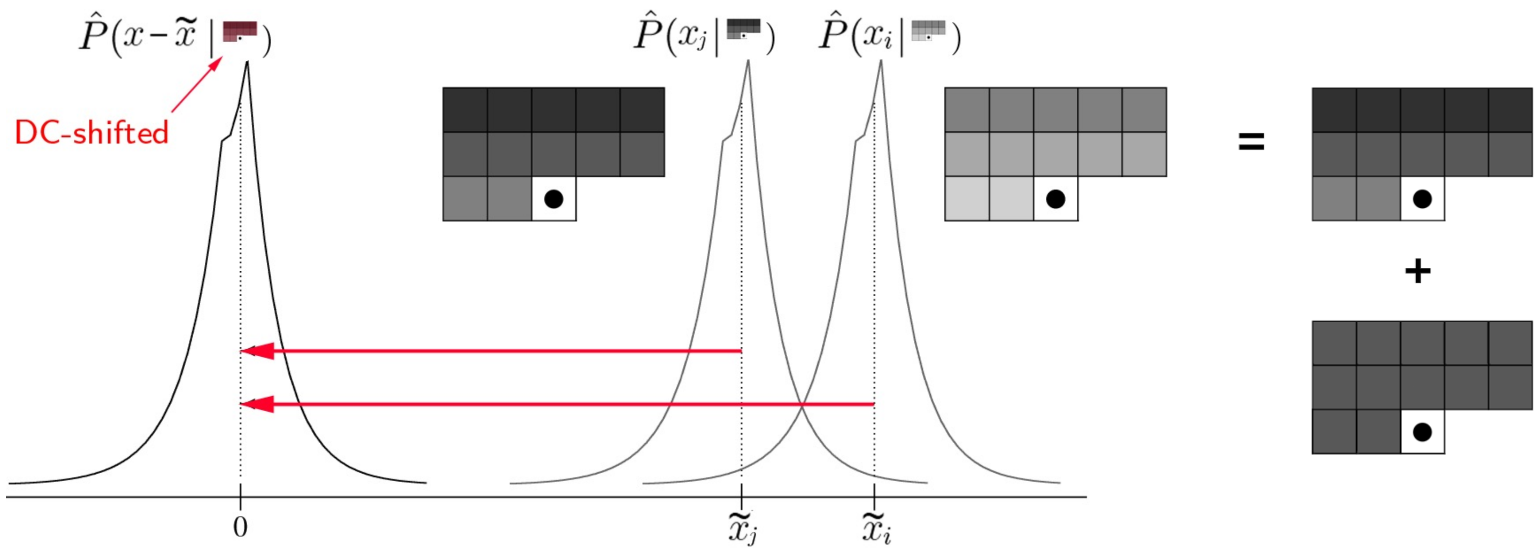
- *DC Invariance* (2)
  - similar patterns can occur at different brightness levels $\Rightarrow$ contexts that differ by a constant *brightness shift* tend to produce similar distributions, up to a corresponding *support shift*



Properties of grayscale images (cont.)

DC Invariance (2)

- similar patterns can occur at different brightness levels $\Longrightarrow$ contexts that differ by a constant *brightness shift* tend to produce similar distributions, up to a corresponding *support shift*

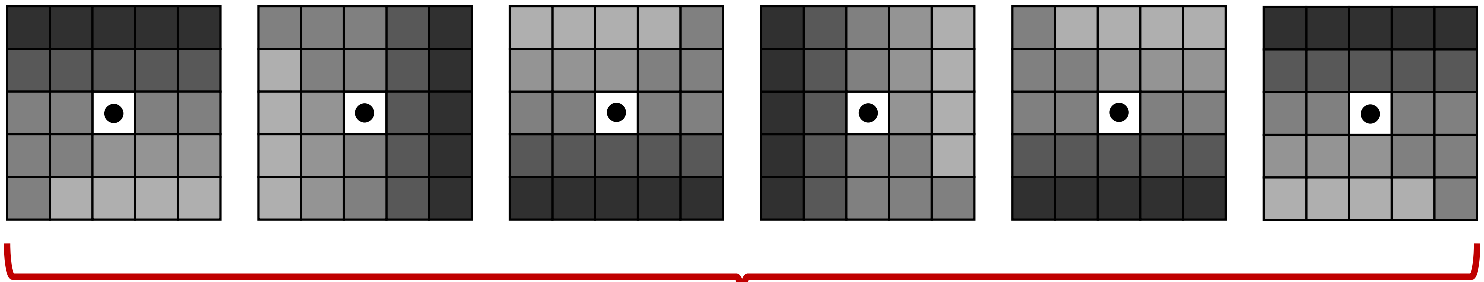


- $\boldsymbol{S}, \boldsymbol{S}'$ contexts, $\boldsymbol{S}' = \boldsymbol{S} + (u, u, \ldots, u) \triangleq \boldsymbol{S} + u\mathbf{1} \Rightarrow P(x|\boldsymbol{S}) \approx P(x+u|\boldsymbol{S}')$
- *natural implementation: differential representation* of the context. For example, build contexts out of *gradients* (differences) rather than absolute pixel values. In this case, prediction *must* be used to re-center distributions.

Properties of grayscale images (cont.)

- *Symmetries*
  - Patterns often repeat in different orientations, and statistics are not very sensitive to left/right, up/down, or black/white reflections $\Rightarrow$ *merge contexts that are equivalent up to such symmetries*

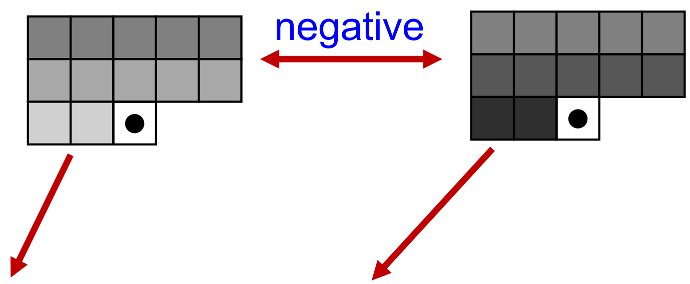
Spatial symmetries (less directly relevant for causal contexts)



similar conditional distributions

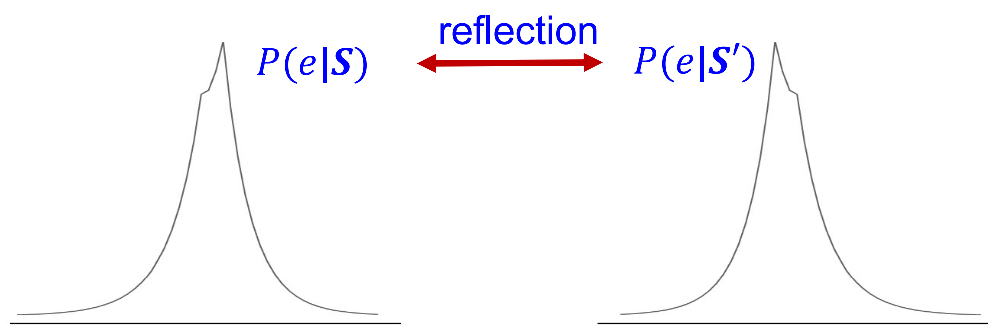
Properties of grayscale images (cont.)

- *Symmetries*
  - Black/white symmetry: An image and its negative have the same statistics, with gray levels reversed.



$$\boldsymbol{S}' = (\alpha-1)\mathbf{1} - \boldsymbol{S} \quad \Rightarrow \quad P(x|\boldsymbol{S}') \approx P(\alpha-1-x|\boldsymbol{S}) \qquad (\alpha-1 = \text{white})$$

for prediction errors: $P(e|S') \approx P(-e|S)$



  - bring context to *canonical form*, e.g., upper left corner $< \alpha/2$
  - choose sign for accumulating statistics, remember to flip if necessary when statistics are used: *are the counts for* x *or for* $\alpha-1-x$ *?*

Plan for a continuous tone image compressor

- Use a predictor, preferably one that avoids interpolating over sharp edges.
- Collect statistics of *prediction errors*.
- Use contexts, but group the into *context classes*, collecting one set of statistics per class, and using the classes as *coding states*.
  - Build the classes by grouping together contexts that are likely to produce similar distributions, up to controllable transformations, such as DC shift, symmetry, etc.
- Make the contexts *differential*, i.e. form them out of *gradients* , rather than absolute pixel values.
  - This will automatically group together contexts that differ only by constant brightness shifts.
- Take advantage of the fact that prediction error distributions are well modeled by a two-parameter TSGD.
- Use a coding strategy matched to the TSGD model.