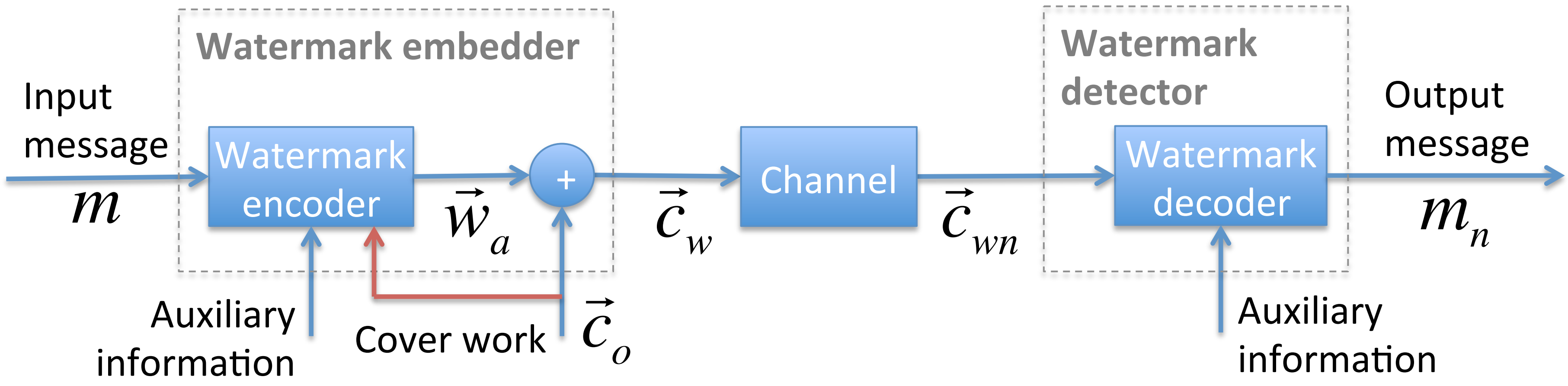
Image Data Compression

Perceptually adaptive watermarks

Reminder: watermarking with side information



Practical dirty-paper watermarking with forward error correction:

- Original multi-bit message encoded with error-correcting (e.g. trellis) code
- Dirty-paper code: several codewords (sub-codebook) per each message, encoder chooses most suitable encoding for each message given a cover work (to minimize distortion)
- Naïve performance estimate: $$R_1 \le I(X;Y) - I(S;X) = \frac{1}{2}\log\left(\frac{P_a + P_e + P_s}{P_a} \cdot \frac{P_e}{P_s + P_e}\right)$$
  (P_a, P_e, P_s – channel noise, added pattern, cover work source power)
- **Ideal Costa Scheme (ICS):** distortion-compensating quantization (QIM): $$R \le \frac{1}{2}\log\left(1 + \frac{P_e}{P_a}\right)$$
  Independent of cover work statistics, same as Shannon's AWGN channel limit!
- Scalar quantization: **LSB** embedding, **Scalar Costa Scheme**, rate always worse than **ICS**
- Fast multi-dimensional quantization: lattice codes; may approach **ICS** rate limit
- In practice: E_8 lattice, 8-dimensional quantization, 4-bit message embedding

Watermarking "perceptibility"

Watermarking is supposed to be "imperceptible"

- What is "perceptible", depends on media space and applications
- Below we consider only human perception of static images (i.e. HVS-based metric)

Fidelity: similarity of images before and after WM embedding

- Primary measure of concern for most applications
- Channel distortions may (or may not) hide, or *mask* watermarking artifacts

Before WM embedding:
low quality

No visual artifacts:
high fidelity

After WM embedding
low quality

Watermarking quality

Quality: absolute measure of appeal (in particular, lack of obvious processing artifacts)

- Quality may even improve after WM embedding!
- Observer may not notice difference, especially if he/she has no access to original work
- Important parameter: quality **change** due to watermarking process
- High fidelity necessarily implies little change in quality!

Before WM embedding: **high quality**

Considerable visual artifacts, pixel value differences: **low fidelity**

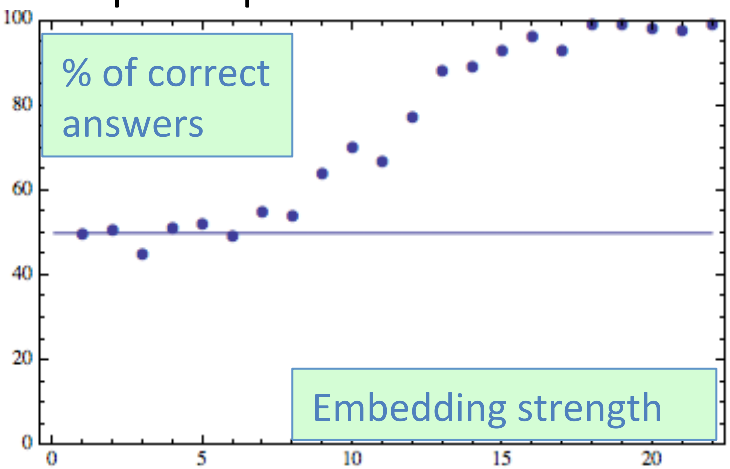
After WM embedding: **high quality**

- Humans have different perceptual acuteness (e.g. “golden eyes”, “golden ears”)
- **Just noticeable difference (JND)** unit: distortion noticeable in 50% of trials
  (while 1 JND is well-defined, multiple JNDs must be further specified!)
- Statistical approach: **two-alternative forced choice** (2AFC):

vs

?

- Sample response statistics:



- ITU-R Rec. 500 quality and impairment scale:

	Quality	Impairment
5	Excellent	Imperceptible
4	Good	Perceptible, but not annoying
3	Fair	Slightly annoying
2	Poor	Annoying
1	Bad	Very annoying

Used in TV quality studies

Human tests: accurate, but expensive and not easily reproducible!

Automated evaluation: perceptual model

Need some substitute for human tests: function **D(c_w, c_0)**, monotonically related to human test outcomes (no need in absolute calibration, only relative results).

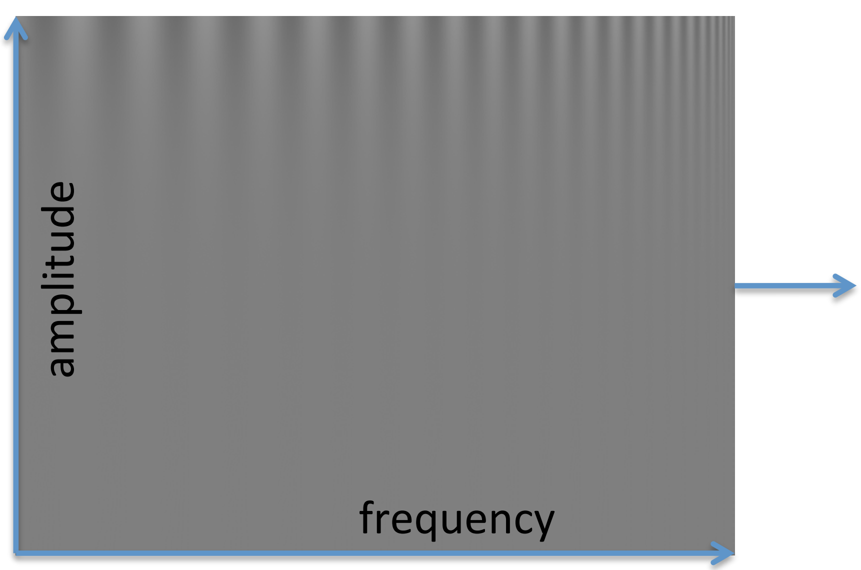
Well-known function – MSE: $D_{mse}(\vec{c}_w, \vec{c}_0) = \frac{1}{N}\sum_i (c_w[i] - c_0[i])^2$

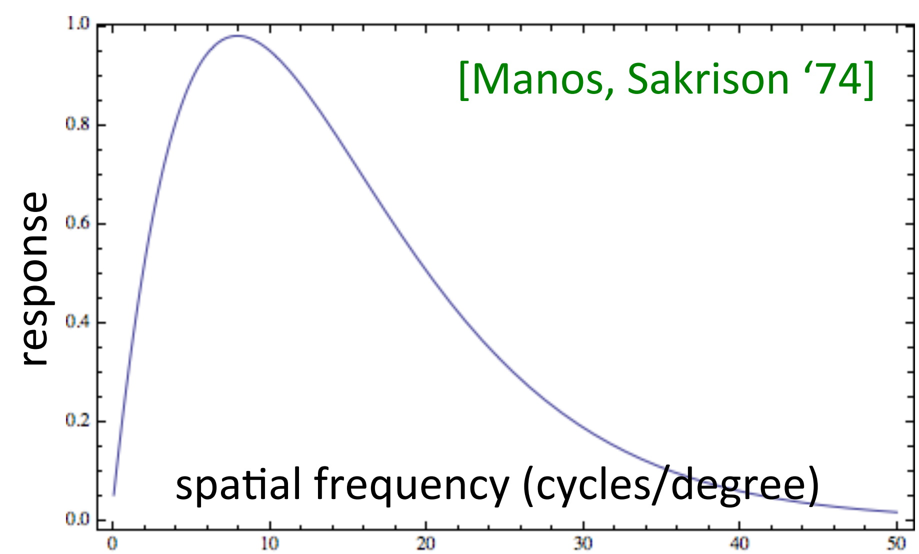
[Girod '93]: MSE is poor estimate of true fidelity!

Human visual system: complex (and unknown) structure, multiple phenomena.
Practical systems mostly account for: sensitivity, masking, pooling.

Sensitivity:

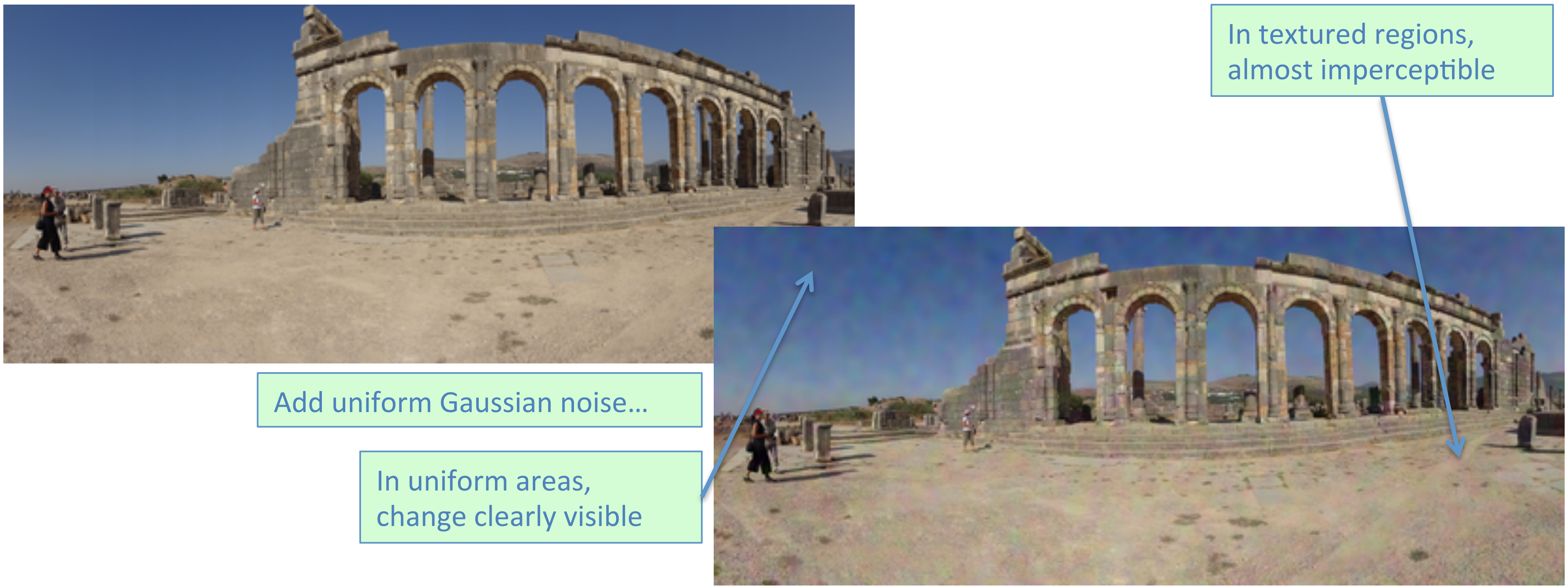
- Spectral - color vision (density of cones/rods, context- and time-dependent sensitivity, …)
- Temporal – motion flicker (exposure and relaxation times, processing speed, …)
- Spatial – Modulation Transfer function (MTF) / Contrast Sensitivity function (CSF):





Masking and pooling in human visual system

Perception sensitivity may be affected by context:



- Frequency masking: presence of one frequency masks the presence of another
- Brightness masking: local brightness masks contrast change

Pooling: given changes in elements of image, how is the overall metric affected?

Common approach: L_p-norm:

$$D(\vec{c}_w, \vec{c}_0) = \left[\sum_i |\Delta[i]|^p \right]^{1/p}$$

E.g. $p = 2$, $\Delta = c_o[i] - c_w[i]$ – is MSE,
$p = \infty$ would be max difference norm,
$p = 1$ – sum of absolute differences

Watson's perceptual model [Watson '93]

Goal: **estimate difference between two images as number of JNDs**

- Far better than MSE in estimating noise effects on fidelity
- Still, relies on pixel alignment, underestimates effect of certain artifacts

Originally developed to fit JPEG standard, but can be used independently

Notation:

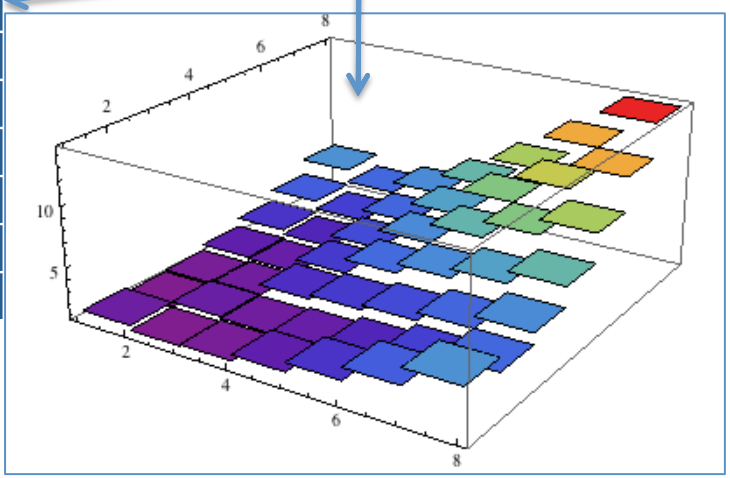
- Images divided into 8x8 blocks (M blocks in total)
- c[i, j, k] = pixel value at position (i, j) in block k, i.e. i,j = 0,…,7
- Compute DCT of each block: C[i, j, k] = DCT coefficient (i, j) in block k
- C[0, 0, k] = DC component in the block k

Sensitivity table:

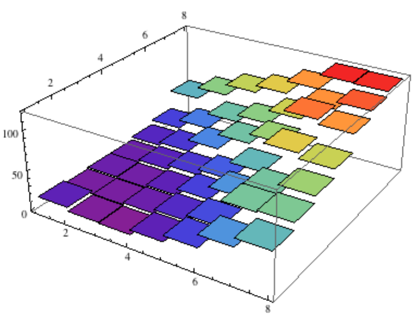
- t[i, j] ≈ min discernible amplitude change of C[i, j, k] (without masking effects!)

1.40	1.01	1.16	1.66	2.4	3.43	4.79	6.56
1.01	1.45	1.32	1.52	2.0	2.71	3.67	4.93
1.16	1.32	2.24	2.59	2.98	3.64	4.6	5.88
1.66	1.52	2.59	3.77	4.55	5.3	6.28	7.6
2.4	2.0	2.98	4.55	6.15	7.46	8.71	10.17
3.43	2.71	3.64	5.3	7.46	9.62	11.58	13.51
4.79	3.67	4.6	6.28	8.71	11.58	14.5	17.29
6.56	4.93	5.88	7.6	10.17	13.51	17.29	21.15

t[i, j]



Compare to JPEG default luminance quantization table:



Watson's perceptual model [Watson '93]

Luminance masking: adjust sensitivity table according to average block brightness

Computed for original work c_o

$$t_L[i,j,k] = t[i,j] \cdot \left(\frac{C_o[0,0,k]}{\langle C_o[0,0] \rangle} \right)^{a_T}, \quad \langle C_o[0,0] \rangle = \frac{1}{M} \sum_k C_o[0,0,k]$$

Watson: $a_T = 0.649$

Contrast masking: reduce visibility of change at some frequency due to its energy

$$s[i,j,k] = \max \left\{ \begin{array}{c} t_L[i,j,k] \\ t_L[i,j,k]^{1-w} \cdot |C_o[i,j,k]|^w \end{array} \right\}$$

"Slacks"

Watson: $w = 0.7$

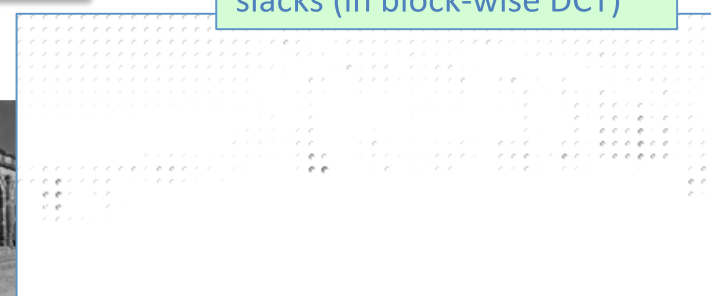
Pooling: combine together relative differences in different frequencies

$$D_{wat}(\vec{c}_w, \vec{c}_0) = \left[\sum_i |\Delta[i]|^p \right]^{1/p}, \quad \Delta[i] = \frac{C_w[i,j,k] - C_o[i,j,k]}{s[i,j,k]}$$

Watson: $p = 4$

c_o

c_w



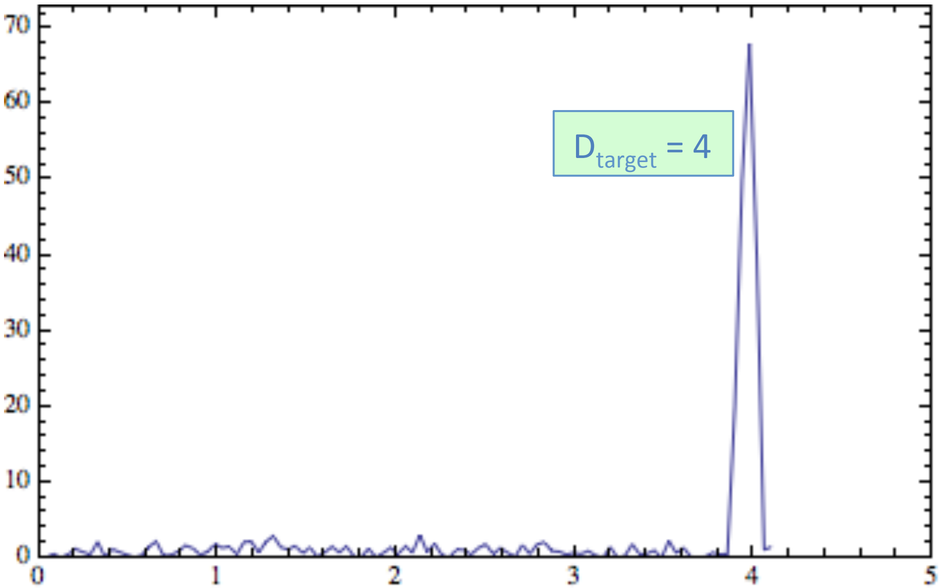
slacks (in block-wise DCT)

$D_{wat} = 0.24$

Since DCT is a linear transform, Watson's perceptual difference is linear wrt scaling of the embedded pattern => may easily select α providing arbitrary perceptual difference:

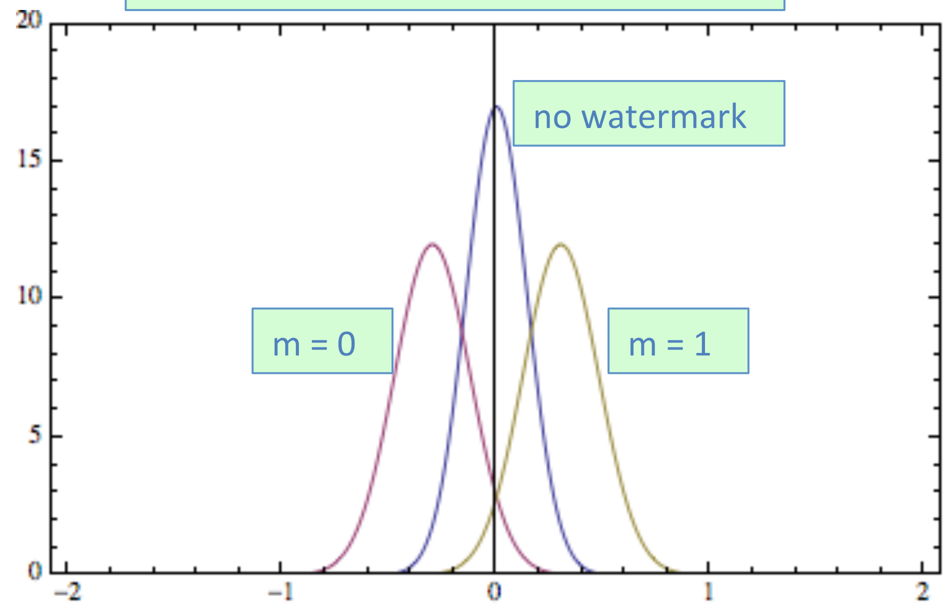
$$D_{wat}\left(\vec{c}_w = \vec{c}_0 + \alpha\vec{c}_m, \vec{c}_0\right) = \left[\sum_i \left|\frac{C_w[i] - C_o[i]}{s[i]}\right|^p\right]^{1/p} = \left[\sum_i \left|\frac{\alpha C_m[i]}{s[i]}\right|^p\right]^{1/p} = \alpha D_{wat}\left(\vec{c}_m, \vec{c}_0\right)$$

Simplest adaptation: choose embedding strength α to reach some pre-defined value D_{target}:



Distribution over perceptual distance

Typical histograms over detection value



[Cox et al '08]

Perceptual shaping

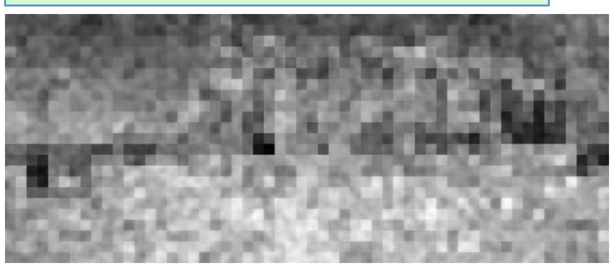
Idea: amplify mark where it is well hidden, attenuate where it is noticeable

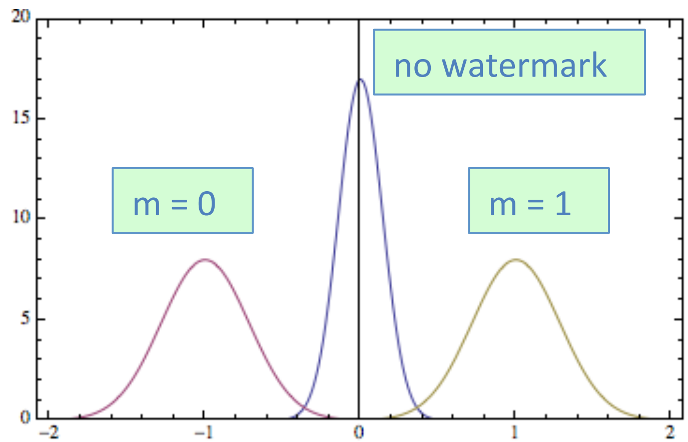
- Need to determine perceptual slack per each term of work in some marking space
- Watson model: slack s[i, j, k] per each DCT coefficient. Other models: pixel values, frequencies in Fourier space, wavelet transform coefficients etc.
- While embedding, scale added pattern according to slacks, for detection, invert scaling

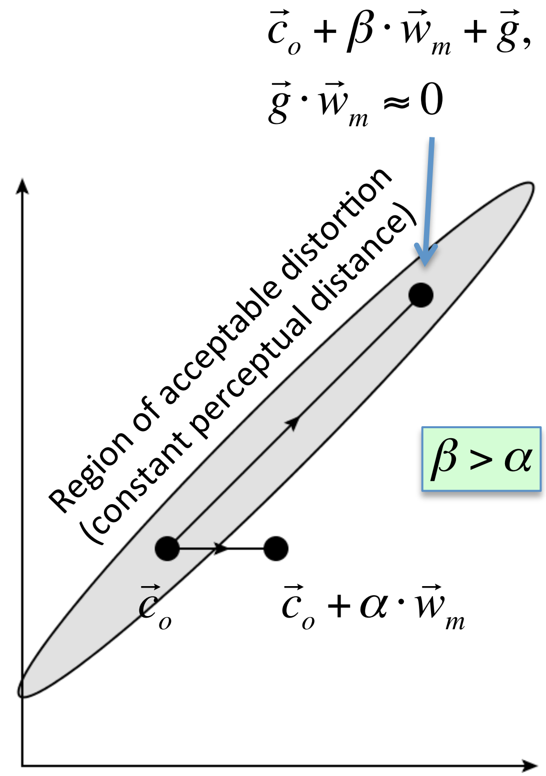
Problem: detector does not have original work to compute slacks!

- One solution: detector computes slacks for the received work c_{wn}, which can be approximately the same as the original values
- Another solution: consider shaping as another distortion, do not perform any inversion at detector
- Compensate higher "noise" by adjusting detection value to a higher level than without perceptual shaping
- Typical result [Cox et al '08]:

Perceptually shaped watermark







Optimal perceptual shaping

Assume watermarking system: detection value - linear correlation,
perceptive distance: L_p-norm, re-scaled energy-conserving linear transform:

$$z_{lc}(\vec{c},\vec{w}_m) = \sum_i c[i]\cdot w_m[i], \qquad D\left(\vec{c}_w,\vec{c}_0\right) = \left[\sum_i \left|\frac{C_w[i]-C_o[i]}{s[i]}\right|^p\right]^{1/p}$$

Task #1: maximize robustness, keeping constant perceptive distortion

$$\arg\max_{\vec{w}_a} z_{lc}(\vec{c}_w,\vec{w}_m) = \arg\max_{\vec{w}_a} z_{lc}(\vec{c}_o+\vec{w}_a,\vec{w}_m) = \arg\max_{\vec{w}_a} z_{lc}(\vec{w}_a,\vec{w}_m) \Leftrightarrow \arg\max_{\vec{W}_a} z_{lc}(\vec{W}_a,\vec{W}_m)$$

$$D(\vec{c}_o,\vec{c}_o+\vec{w}_a) = D_t = const \Leftrightarrow \sum_i \left(\frac{W_a[i]}{s[i]}\right)^p = D_t^p = const$$

Solution: need to find $W_a[i]$, => common optimization, constraint via Lagrange multiplier:

$$\frac{\partial z_{lc}(\vec{W}_a,\vec{W}_m)}{\partial W_a[i]} - \lambda\cdot\frac{D(\vec{c}_o,\vec{c}_o+\vec{w}_a)^p}{\partial W_a[i]} = 0, \qquad W_m[i] - \lambda\cdot p\cdot\frac{W_a[i]^{p-1}}{s[i]^p} = 0,$$

and finally

$$\vec{W}_a = \alpha\cdot\vec{W}_s, \quad \vec{W}_s[i] = \left(W_m[i]\cdot s[i]^p\right)^{1/(p-1)}, \quad \alpha = \frac{D_t}{D\left(\vec{c}_o,\vec{c}_o+\vec{w}_s\right)}.$$

Optimal perceptual shaping

Task #2: minimize perceptual distance, keeping detection value constant

$$z_{lc}(\vec{c}_w, \vec{w}_m) = z_t = z_{lc}(\vec{c}_o + \vec{w}_a, \vec{w}_m) = z_{lc}(\vec{c}_o, \vec{w}_m) + z_{lc}(\vec{W}_a, \vec{W}_m),$$

or, equivalently, $z_{lc}(\vec{W}_a, \vec{W}_m) = const = z_t' = z_t - z_{lc}(\vec{c}_o, \vec{w}_m).$

Solution: almost identical, result:

$$\vec{W}_a = \alpha \cdot \vec{W}_s, \quad \vec{W}_s[i] = \left(W_m[i] \cdot s[i]^p\right)^{1/(p-1)}, \quad \alpha = \frac{z_t - z_{lc}(\vec{c}_o, \vec{w}_m)}{z_{lc}\left(\vec{w}_m, \vec{w}_s\right)}.$$

Effect on real images:
improvement from 89% embedding efficiency (naïve shaping) to 95% [Cox et al, '08].

Non-linear detection values, more complex robustness definition, perceptual distance:

- generally, no closed-form solution, need numerical optimization in each case.
- However, optimal shaping as described above helps also with e.g. normalized correlation!

There are lots of further topics in watermarking – left for your own discovery ☺