Image Analysis and Compression: Renewed Focus on Texture

Thrasos Pappas

EECS, Northwestern University

ICIP, September 28, 2010

People

- **Jana Zujovic, NU**
- Pubudu Madhawa Silva, NU
- Lulu He, NU
- Guoxin Jin, NU
- Alan Brooks, Varian, Inc.
- Xiaonan Zhao, Google
- Junqing Chen, Aptina Imaging
- Dejan Depalov, HP
- Drew Seward, NU

- Huib de Ridder, TU Delft
- Rene van Egmond, TU Delft

- **David Neuhoff, Univ. of Michigan**
- Matt Reyes, Univ. of Michigan

- Fred Juang, Georgia Tech
- Soo Hyun Bae, SONY

- Jim West, Johns Hopkins Univ.
- Ilona Kretzschmar, CCNY
- Vicky Tartter, CCNY
- Karen Gourgey, Baruch College

  – SONY Labs, NSF, NNSA

Outline

- Image and Video Compression
  - “Structurally Lossless” Compression
- Content-Based Retrieval
- Visual to Tactile Conversion
- Multimodal Interfaces

- Structural Texture Similarity Metrics

Image and Video Compression

- Lossless
- Near-Threshold: Perceptually Lossless
  - Digital cameras, SDTV, HDTV, DVD, digital cinema
- Supra-Threshold: Visible Coding Artifacts
  - Video conferencing, video over Internet, wireless

Perceptually Lossless

Perceptually Lossless

PIC: 0.914 bits/pixel

[Safranek-Johnston'89]

Original

Perceptually Lossy (0.2 bits/pixel)



Perceptually Lossy (0.2 bits/pixel)

JPEG

PIC 8x8

SPIHT

Perceptually Lossy
Wireless Video with Error Concealment

Image and Video Compression

- Lossless

- Near-Threshold: Perceptually Lossless
  - Digital cameras, SDTV, HDTV, DVD

- ?

- Supra-Threshold: Visible Coding Artifacts
  - Video conferencing, video over Internet, wireless

“Structurally Lossless” Compression

“Structurally Lossless” Compression

“Structurally Lossless” Compression

“Structurally Lossless” Compression

More than 20% pixels are different

MSE = 392; PSNR = 22.2

“Structurally Lossless” Compression

Image and Video Compression

- Lossless
- Near-Threshold: Perceptually Lossless
  - Digital cameras, SDTV, HDTV, DVD, digital cinema
- "Structurally Lossless"
  - Images look different side-by-side, but …
  - Have similar quality
  - Cannot tell which is the original
- Supra-Threshold: Visible Coding Artifacts
  - Video conferencing, video over Internet, wireless

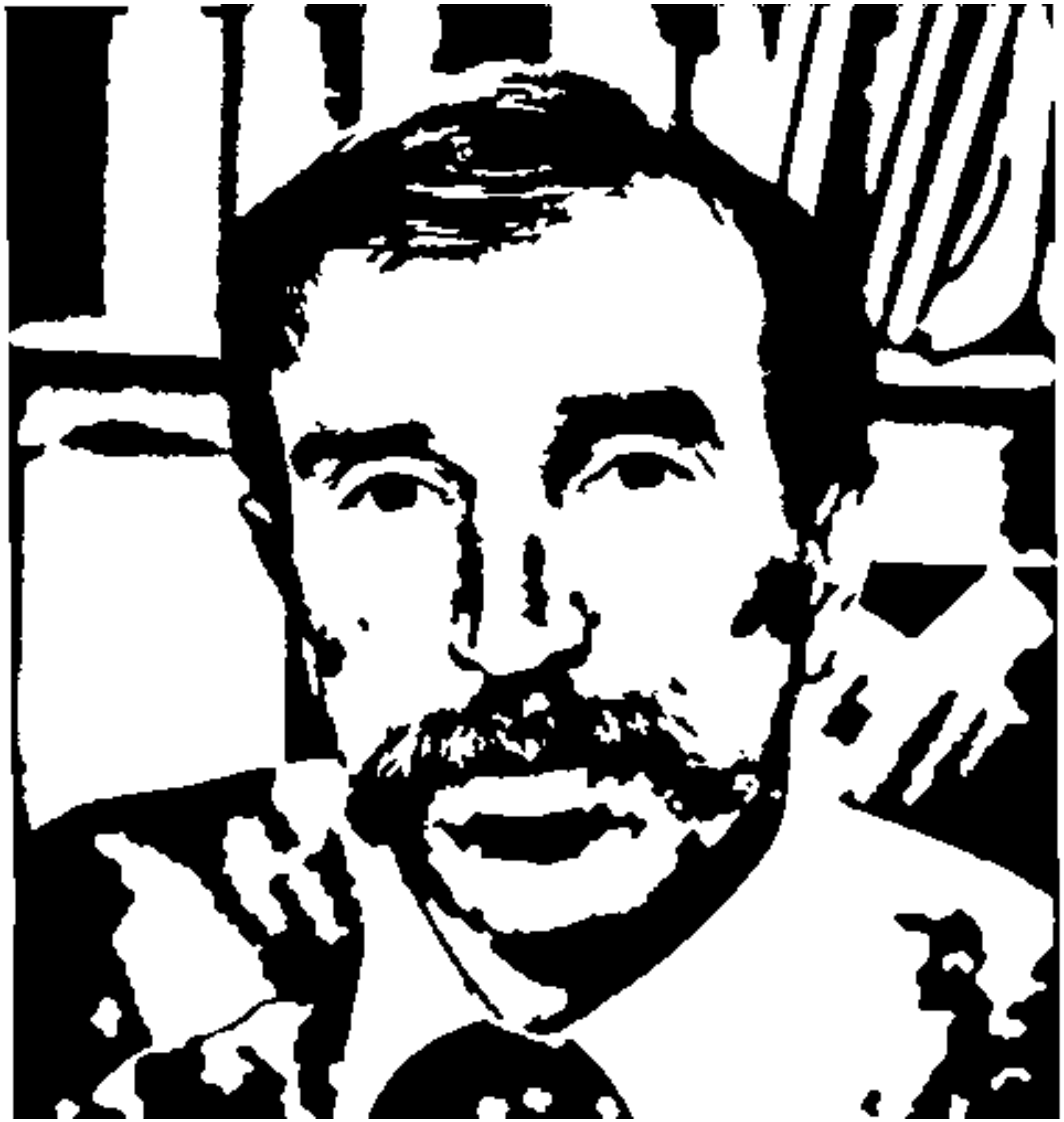
Pappas, Zujovic, Neuhoff, VICP'10

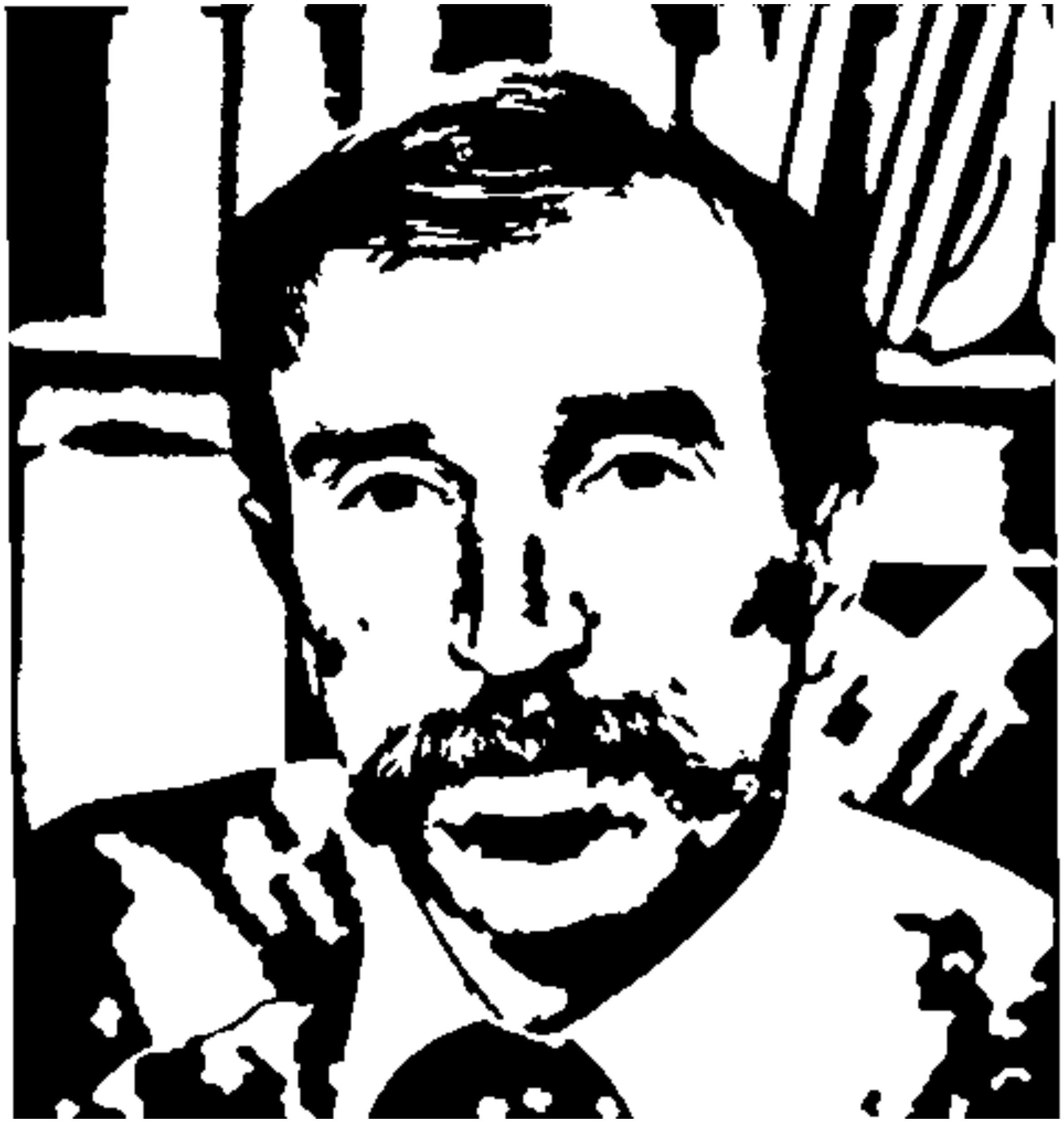
"Structurally Lossless" Compression

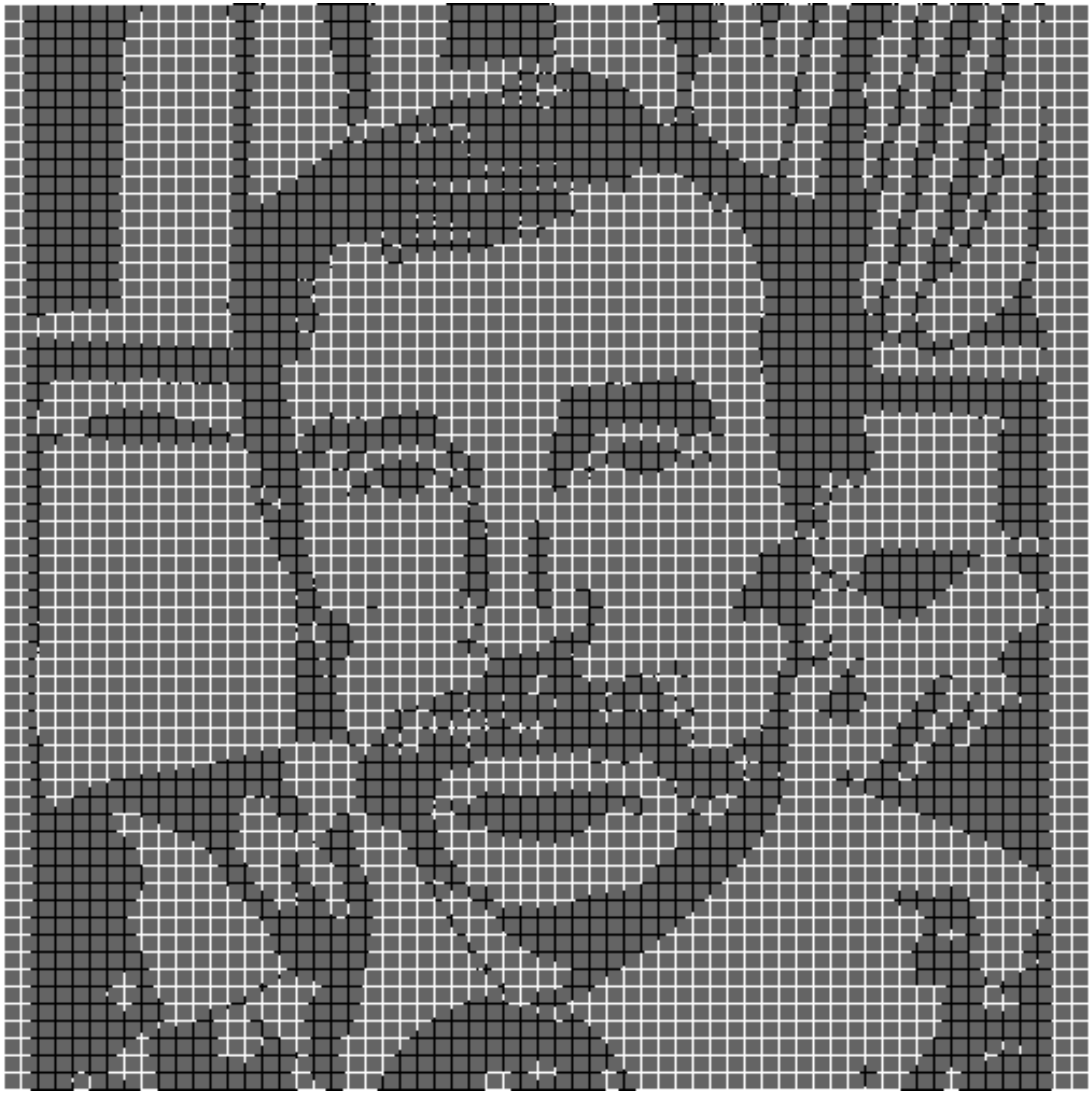
- Key Idea: Use spatial and temporal prediction to encode textured areas
- Develop new metrics that allow substantial point-by-point variations, which do not affect perceived texture quality (SSIM – Bovik *et al.*)
- Use texture blending to avoid blocking artifacts (image quilting – Efros&Freeman'01)
- Use an encoding scheme that is amenable to better texture prediction (Quadtree predictive coding – Teng & Neuhoff'95, Holt & Neuhoff'02, H-264)

Discussion

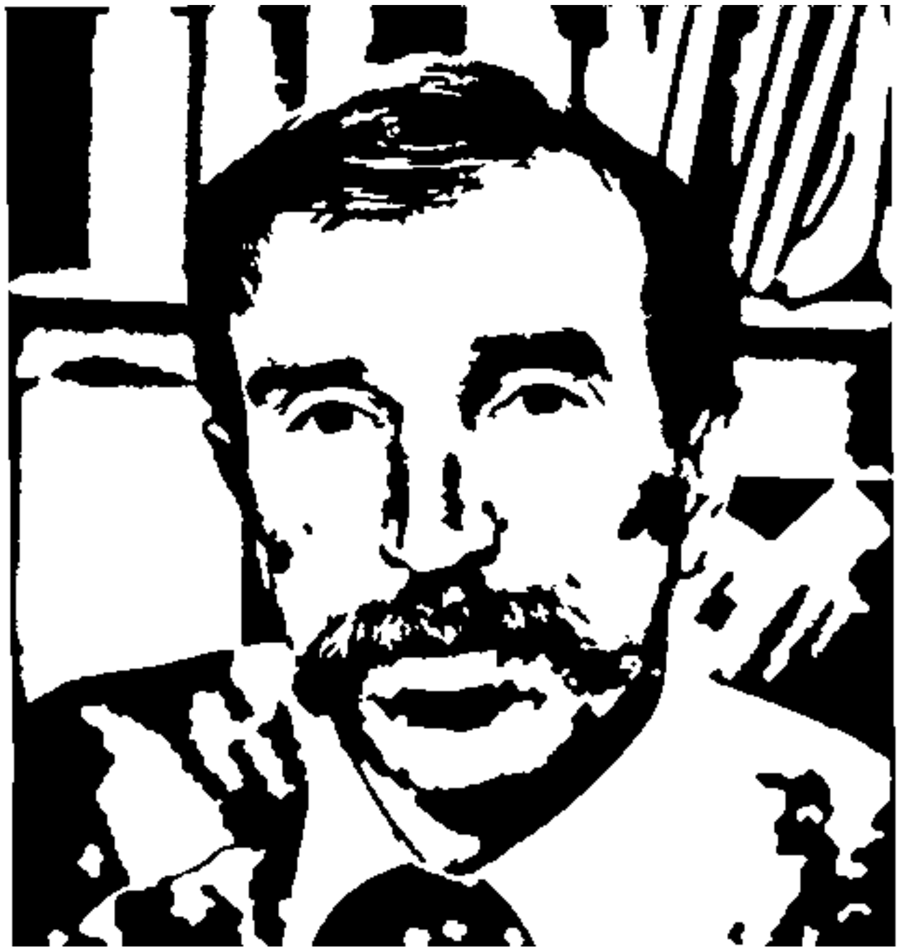
- Spatial and temporal (motion comp.) prediction is the key to success of video compression algorithms (e.g., H-264).
- Current metrics do not allow good prediction of textured areas.
- Transform/subband/wavelet coding techniques cannot efficiently encode texture regions.



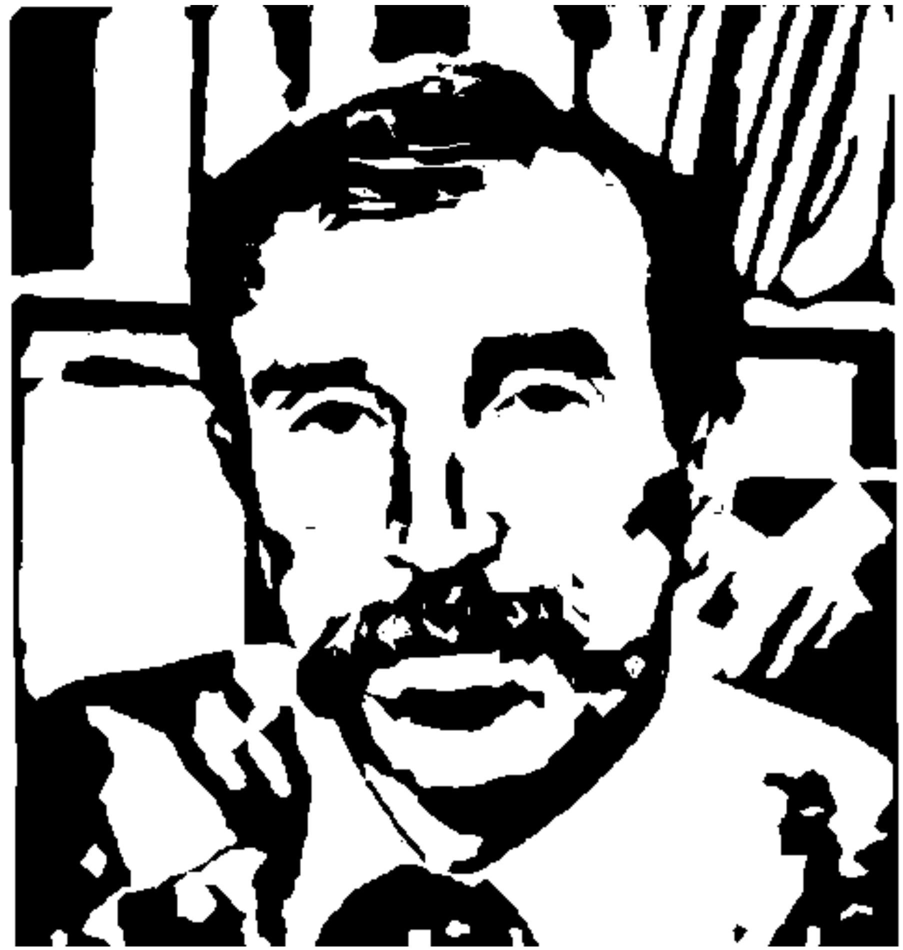




Structurally Lossless Bilevel Compression

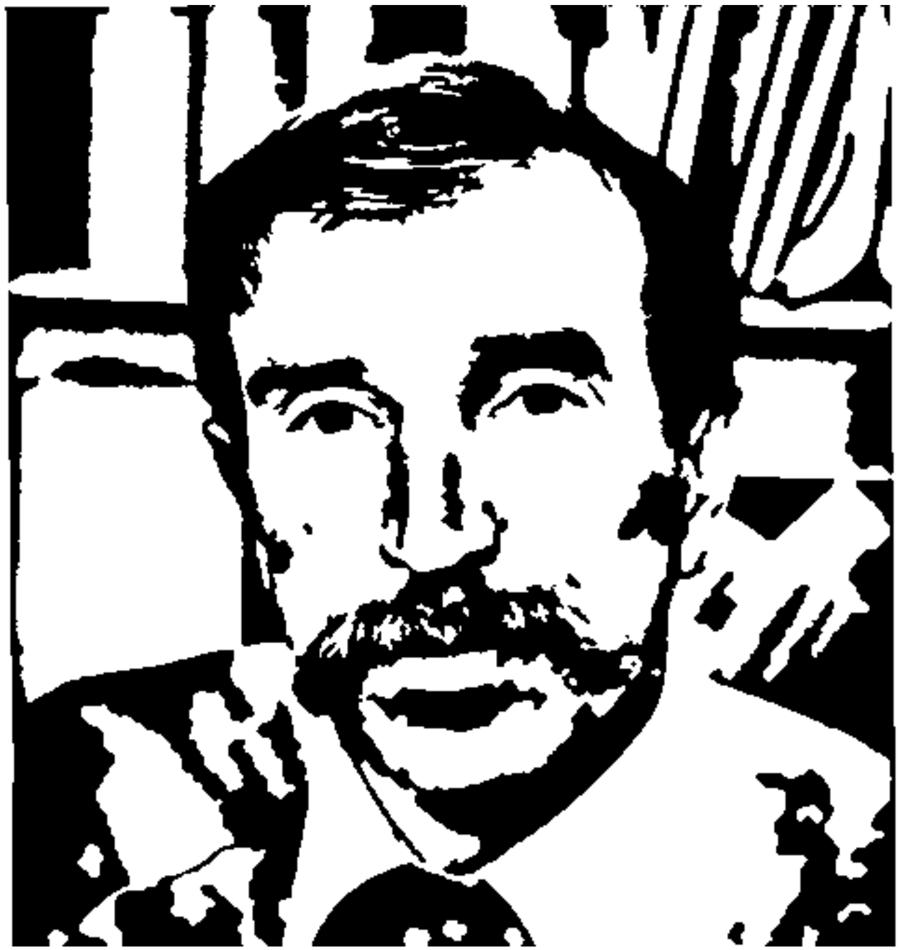


original



8x8 specification

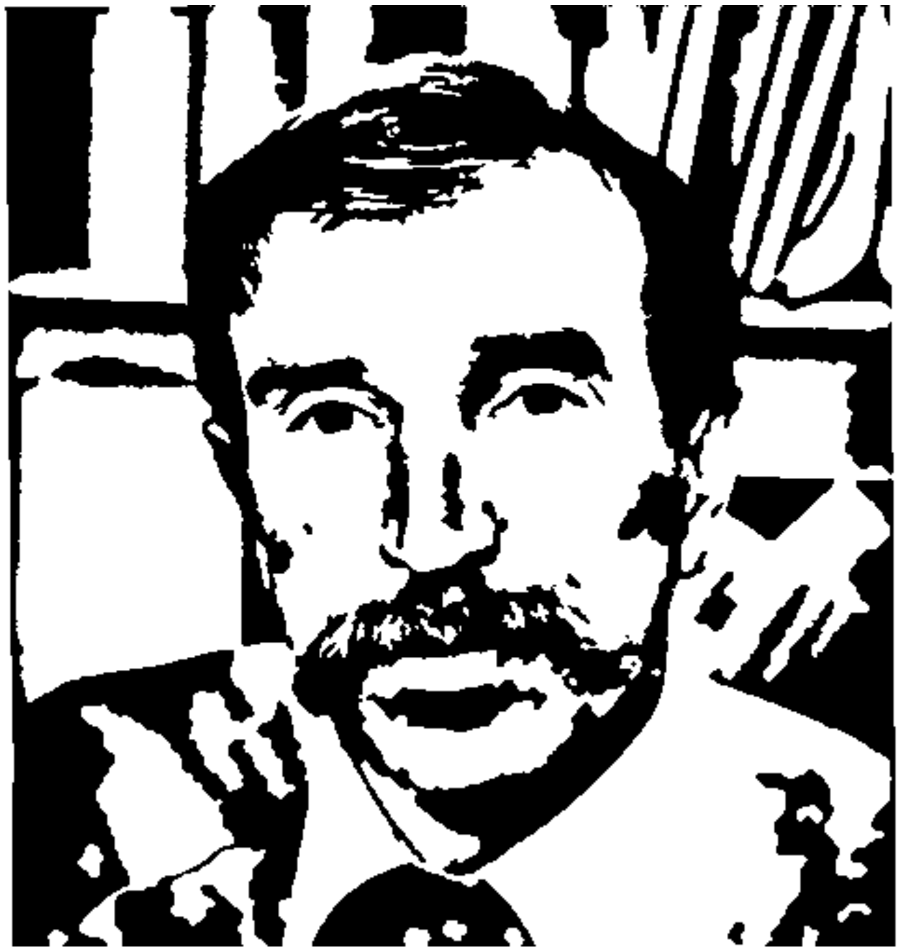
Structurally Lossless Bilevel Compression



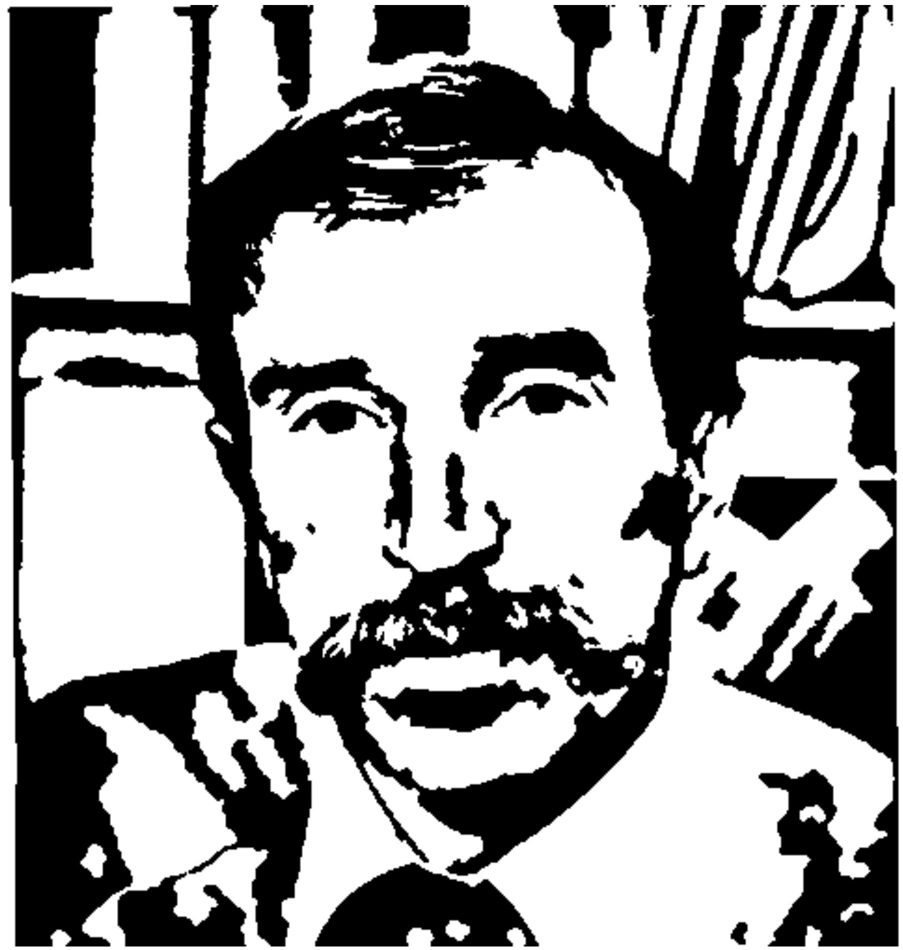
original

4x4 specification

Structurally Lossless Bilevel Compression

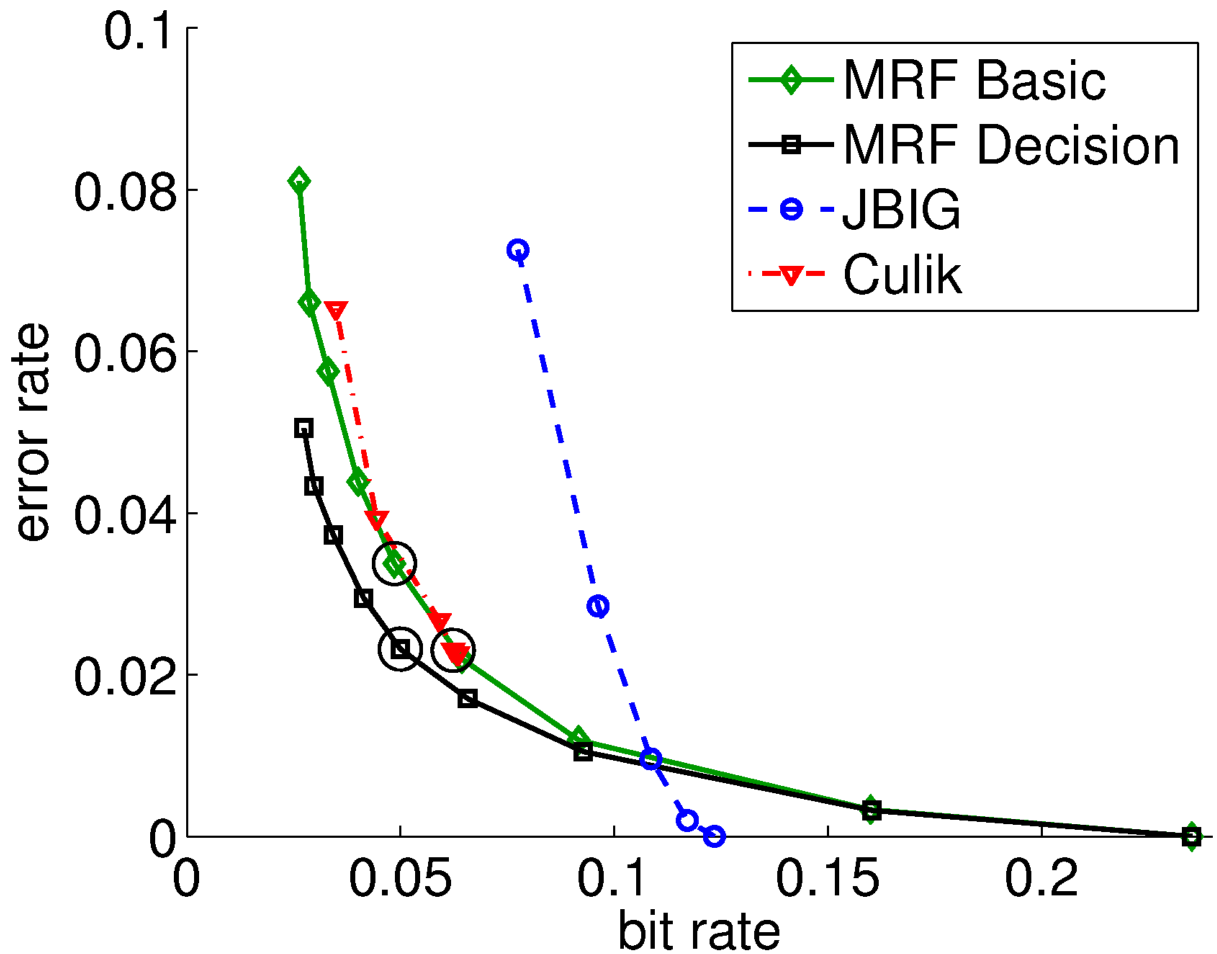


original



2x2 specification

Structurally Lossless Bilevel Compression



Reyes, Zhao, Neuhoff, Pappas, ICIP’07, HVEI’08

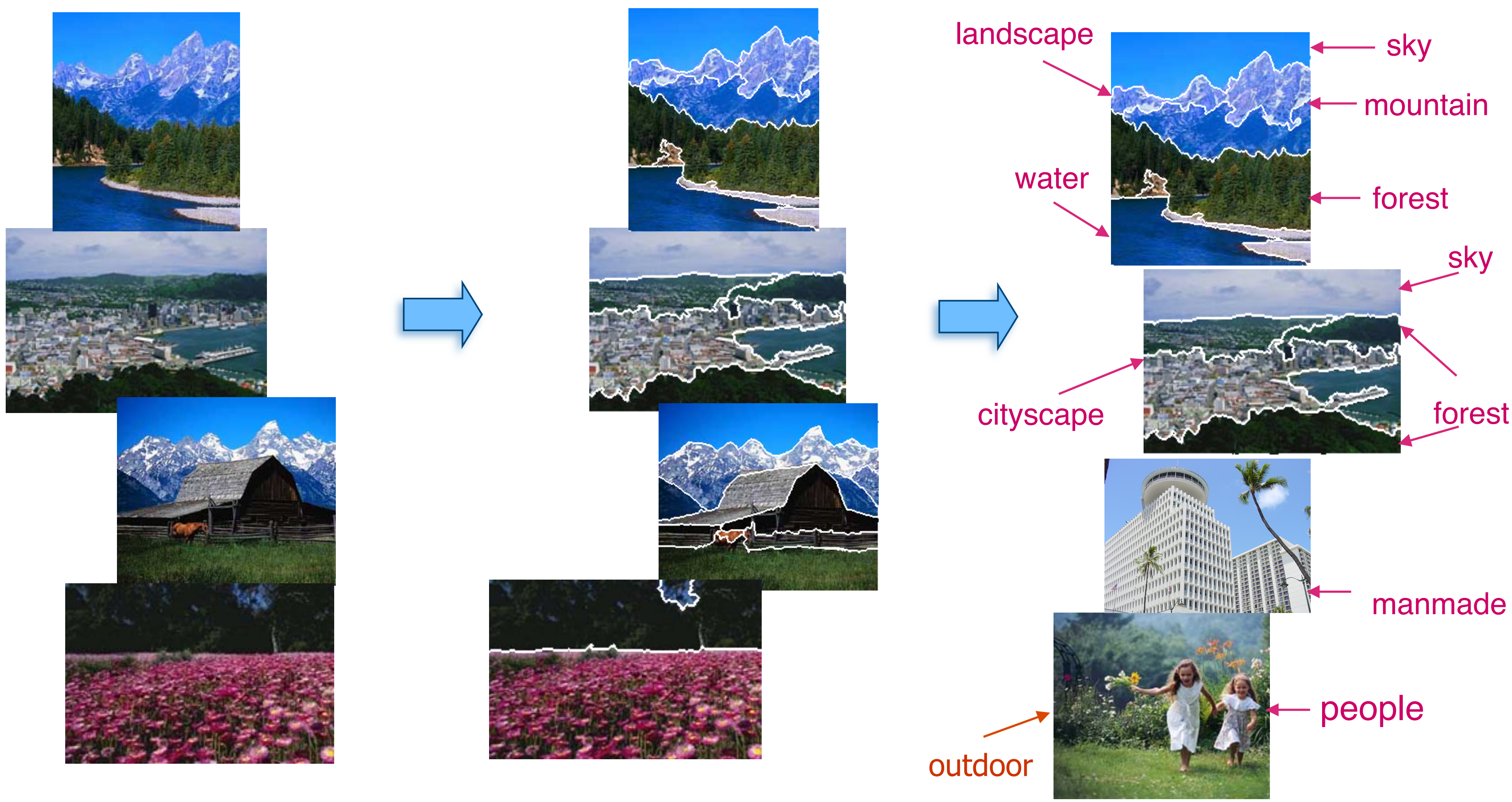
Content-based Retrieval

- Computer vision: focus on objects rather than material perception and texture [Adelson, HVEI'01]

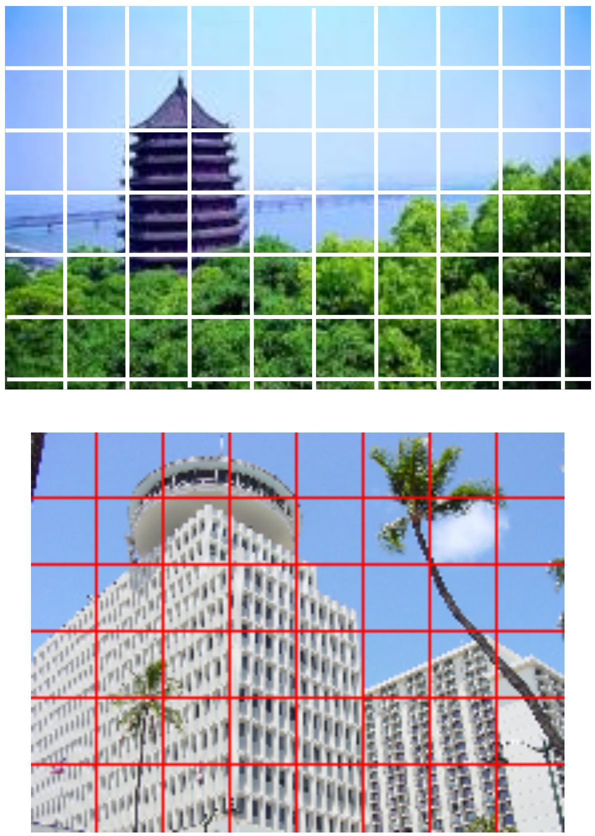
- Alternative: rely on perceptually uniform regions as medium level descriptors for extraction of semantic information [Pappas *et al.*, Comm. Mag.'07]

Segments as Medium Level Descriptors

Images | “Ideal” Segmentations | Semantic Categories



Block-partition vs. Segmentation



Rectangular Partition

Adaptive Perceptual
Color-Texture Segmentation

Segmentation Results





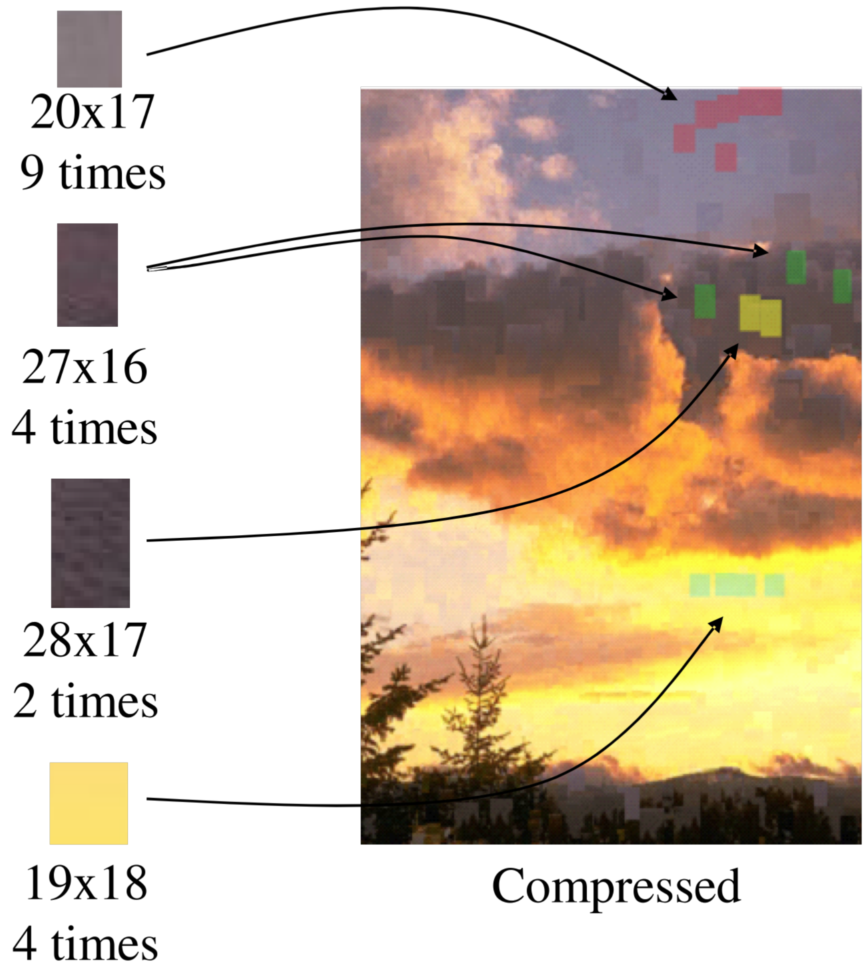






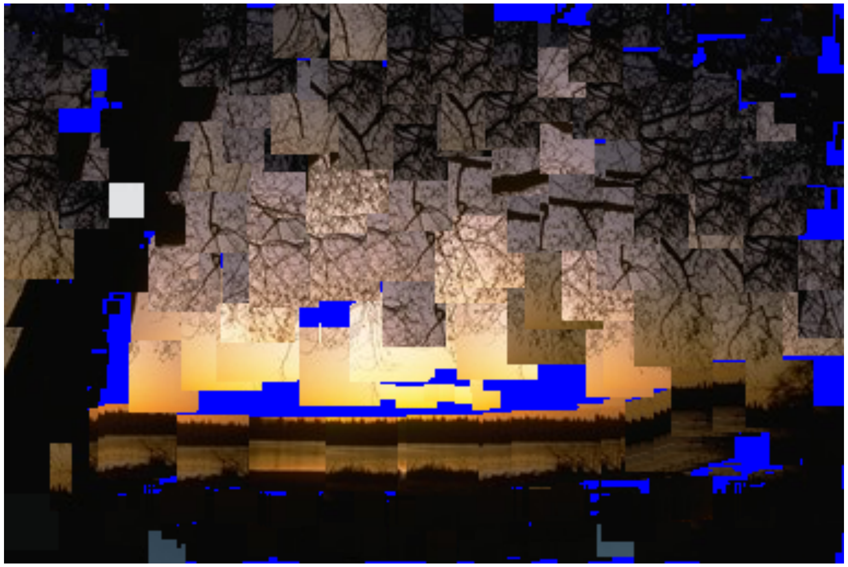
Incremental Parsing [Bae & Juang]

- Extension of Lempel-Ziv [LZ'78] to 2-D, lossy case
- Inherently universal
- Asymptotically optimal
- Compression [IP trans.'08]
- CBR [IP trans.'10]
  - Incr. parsing + LSA
- Need texture similarity metric



Compressed

Incremental Parsing with STSIM2



Semantic Information Extraction

(at Segment level)

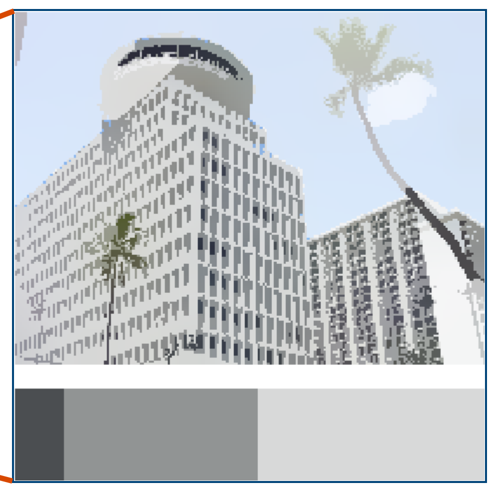
original

Dominant Colors (ACA)

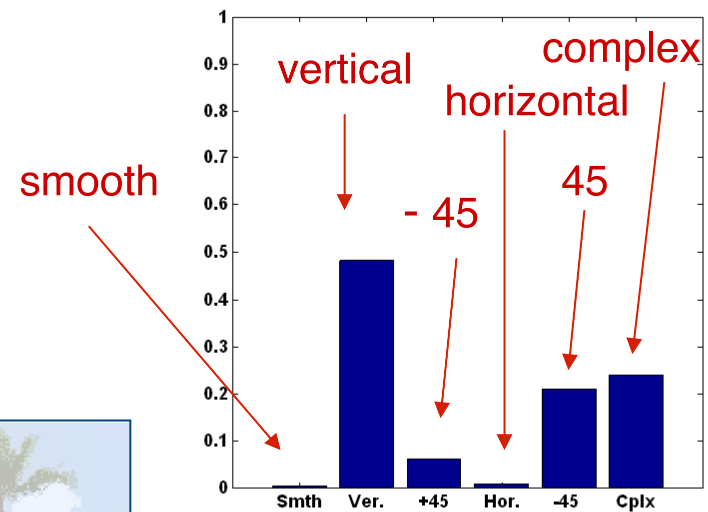
segment 1

segment 2

segment 3



Dominant Colors & Percentages



Spatial Texture

Pappas, Chen, Depalov, CommMag'07

Semantic Information Extraction

(at Segment level)

original

Dominant Colors (ACA)

segment 1

segment 2

segment 3

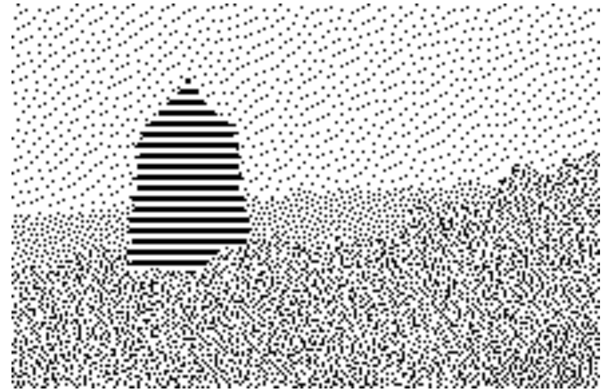
Or direct comparison via texture similarity metric

Pappas, Chen, Depalov, CommMag'07

Visual to Tactile Mapping

Original Color Image

Segmentation



Tactile Display

- Present picture as a collection of segments with different tactile textures
- Perceptually distinct tactile textures
- Tactile device models (paper, plastic, etc.)
- Mapping of visual to tactile textures
- Haptic space perception and scene perception
  - Advantages over line drawings

Need texture similarity metrics (visual and tactile)

Pappas *et al.*, HVEI'09

Braille

Letters:

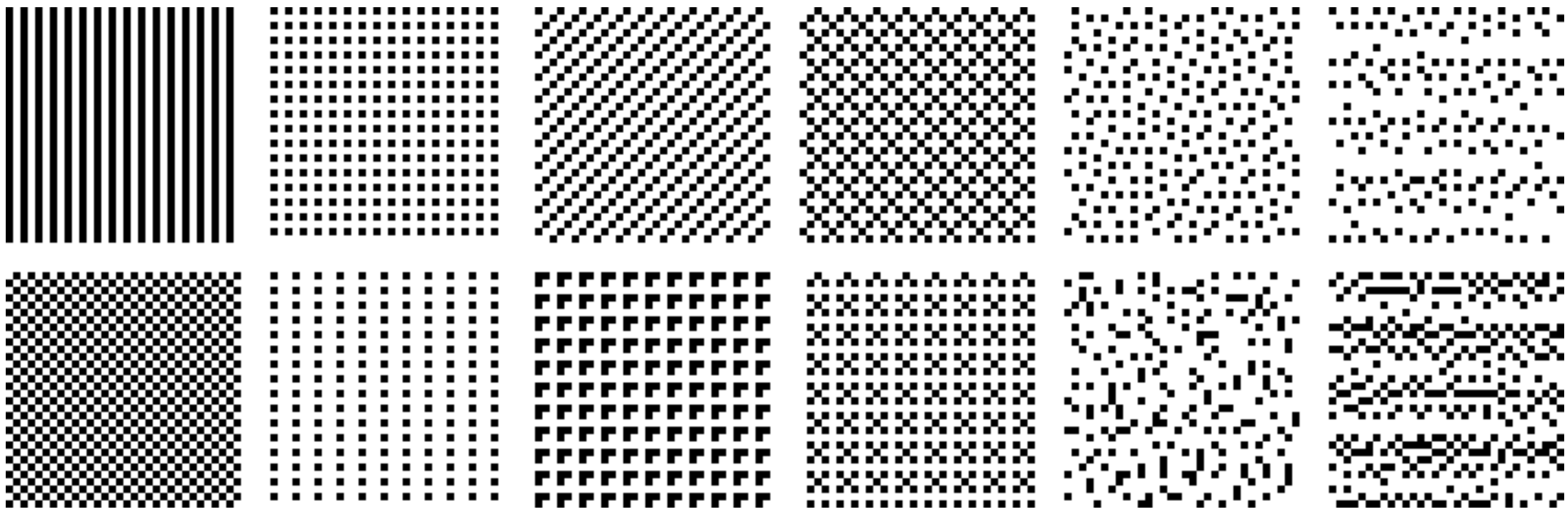
a ⠁	b ⠃	c ⠉	d ⠙	e ⠑	f ⠋	g ⠛	h ⠓	i ⠊	j ⠚
k ⠅	l ⠇	m ⠍	n ⠝	o ⠕	p ⠏	q ⠟	r ⠗	s ⠎	t ⠞
u ⠥	v ⠧	w ⠺	x ⠭	y ⠽	z ⠵				

Numbers:

1 ⠼⠁	2 ⠼⠃	3 ⠼⠉	4 ⠼⠙	5 ⠼⠑	6 ⠼⠋	7 ⠼⠛	8 ⠼⠓	9 ⠼⠊	0 ⠼⠚

Dot diameter: 1.3 mm
Dot spacing: 2.5 mm
Dot Height: 0.5 mm

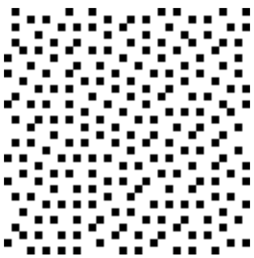
Tactile Pattern Generation



- Generate perceptually distinct tactile textures
- Leverage existing techniques: digital halftoning
- Visual patterns: Minimize visibility of halftone-induced textures
- Tactile patterns: Accentuate texture characteristics

Dot diameter: 0.8 mm   Spacing: 0.8 mm   Height: 0.4 mm

Visual Versus Tactile Pattern Perception



- Visually pleasing blue noise pattern
  - Floyd-Steinberg error diffusion
  - High frequency noise, less visible
- Tactile impression: smooth, boring

- Visually less pleasing
  - Error diffusion with weight perturbations
  - Contains more low frequencies
- Tactile impression: interesting, exciting

- Visually impaired and blind subjects
- Visually blocked subjects

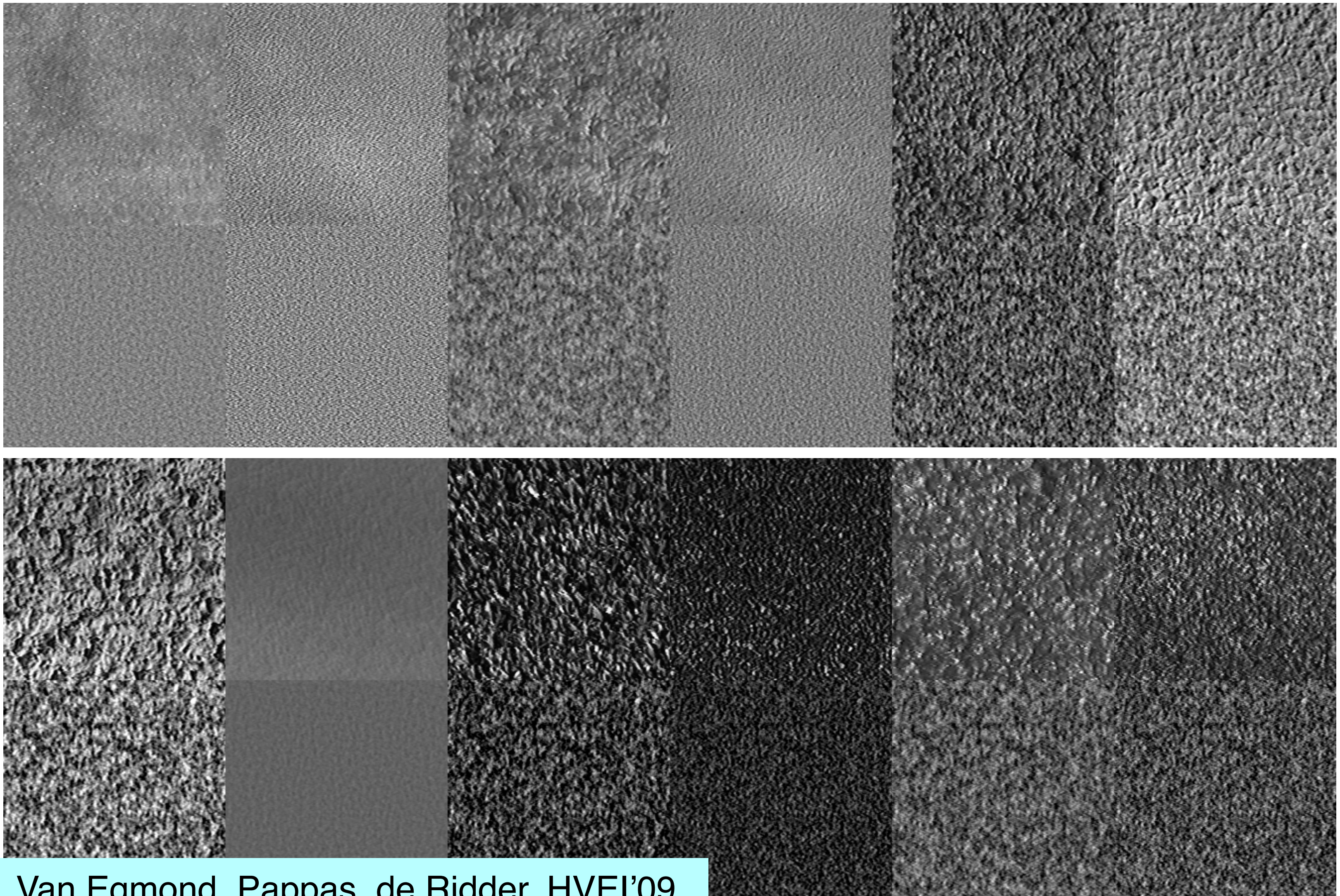
Multimodal Interactions

- Virtual reality
- Immersive environments

Need texture similarity metrics (visual, acoustic, tactile)

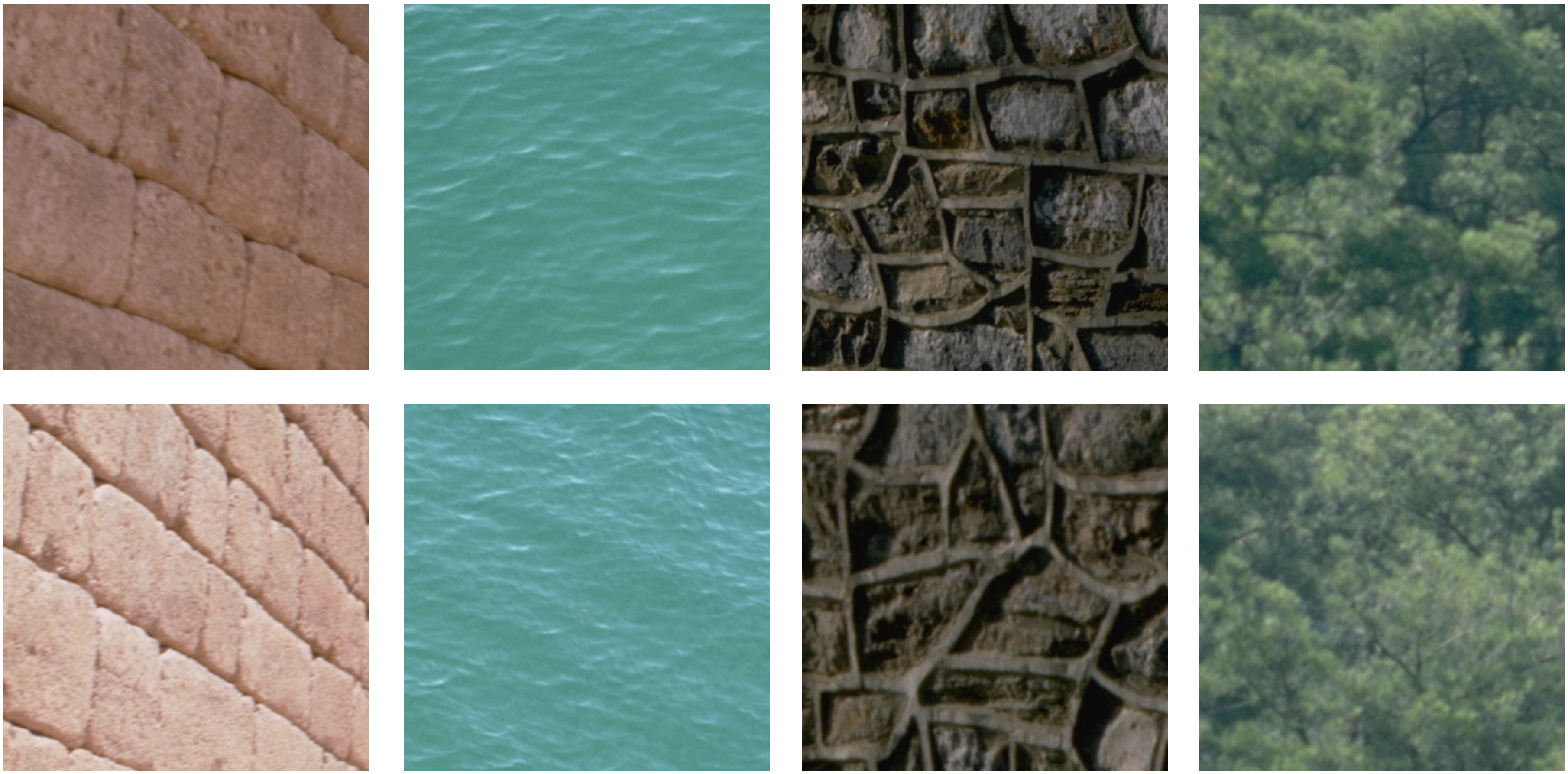
- Tactile-acoustic pattern perception
- Visual-acoustic pattern perception
  - Van Egmond, *et al.*, HVEI'09
- Visual, tactile, acoustic perceptual dimensions: roughness, regularity, directionality
- …

Visual Roughness Perception



Van Egmond, Pappas, de Ridder, HVEI'09

Texture Similarity

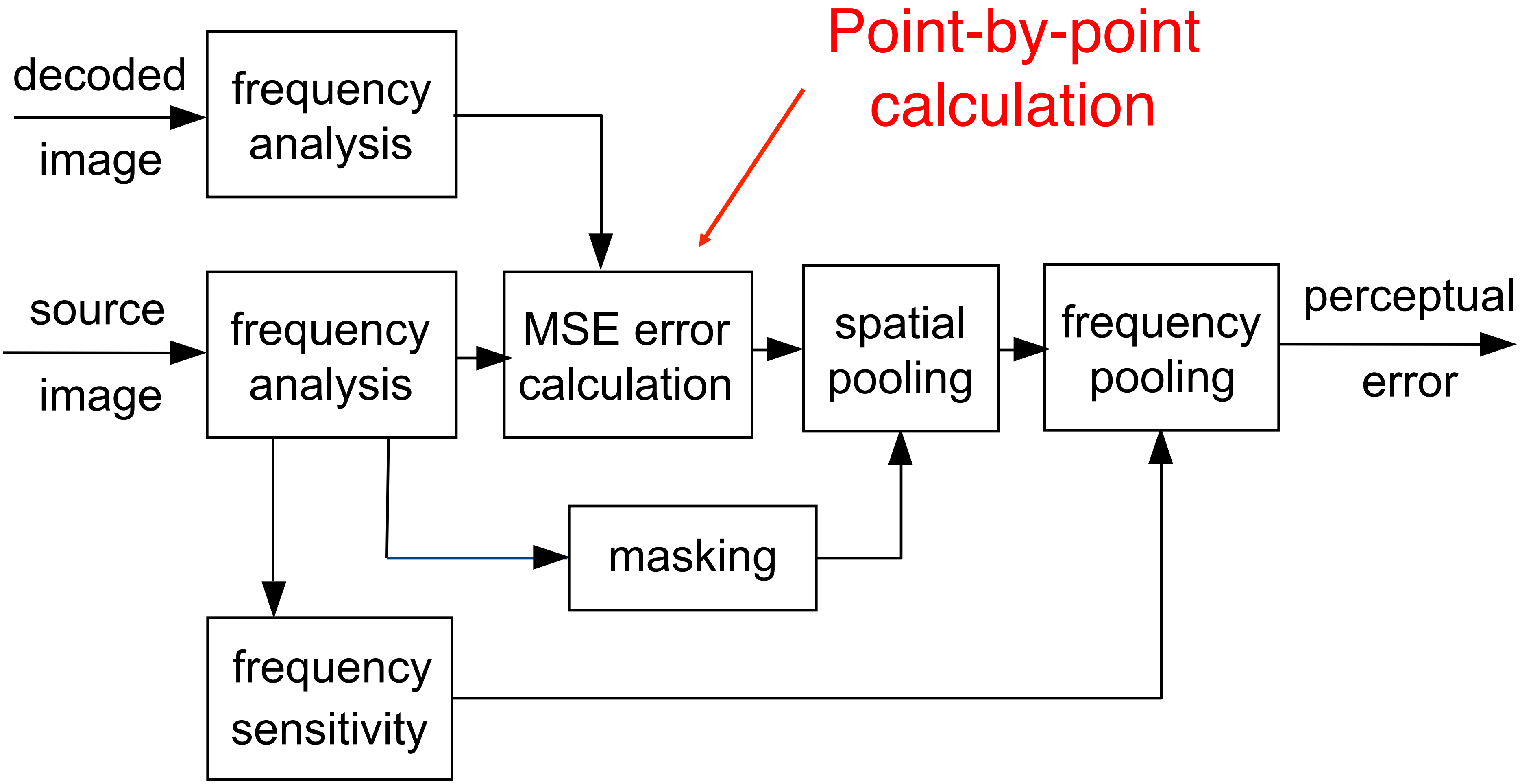


Quality Metrics

- MSE / PSNR
- Perceptually Weighted Metrics
  - Explicit models of HVS
    - Subband sensitivities, contrast/luminance masking
  - Mostly near-threshold
- Point-by-point calculations

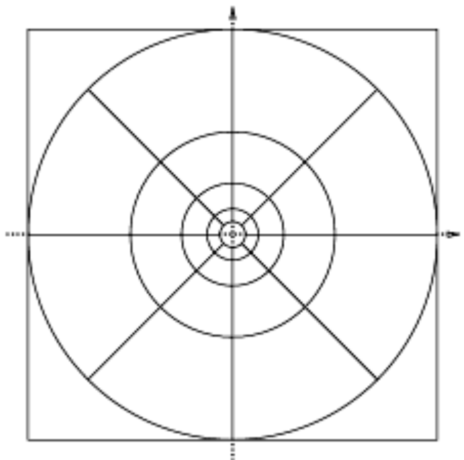
- Structural Similarity Metrics
  - Implicit models of high level HVS properties (extract "structure" information)
  - Insensitive to lighting changes
  - Allow significant pixel-by-pixel distortion
  - Effective in supra-threshold applications
- Region Statistics

Perceptual Quality Metrics

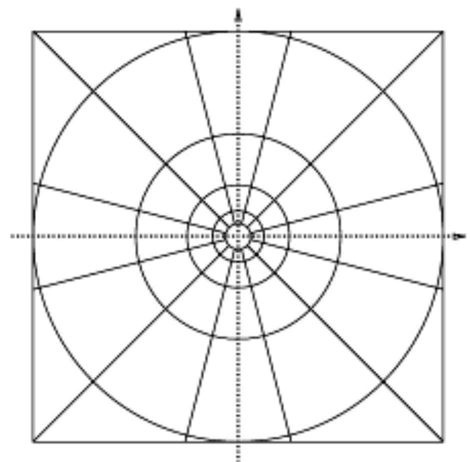


Frequency Analysis

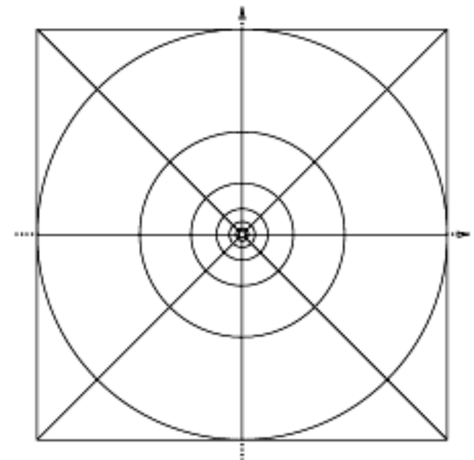
Cortex (Watson'87)



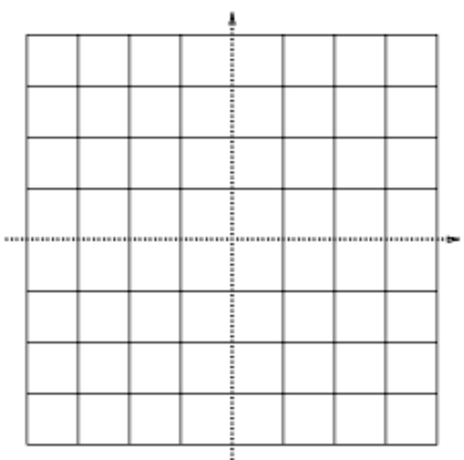
Cortex (Daly'92)



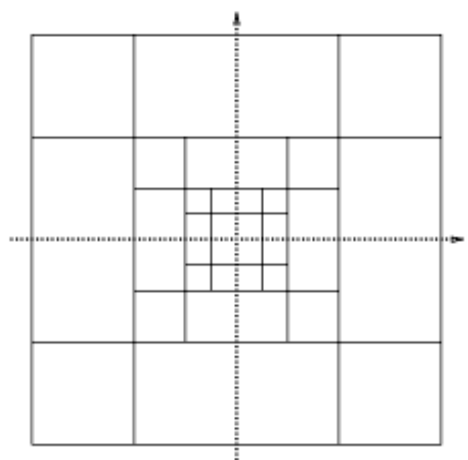
Lubin'91



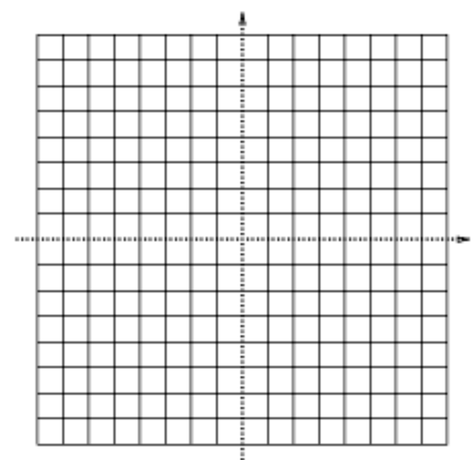
4x4 Subband



Wavelet



8x8 DCT



SSIM Quality Metrics

luminance measurement is a function of the *mean* intensity, related to luminance masking (a.k.a. Weber's law)

contrast measurement is a function of *standard deviation* and *variance*, related to contrast masking (the mean is removed)

structure measurement is a function of *covariance* (the variance is normalized)

image patches being compared

$$l(\mathbf{x}, \mathbf{y}) = \frac{2\,\mu_x\,\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}$$

$$c(\mathbf{x}, \mathbf{y}) = \frac{2\,\sigma_x\,\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}$$

$$s(\mathbf{x}, \mathbf{y}) = \frac{\sigma_{xy} + C_3}{\sigma_x\,\sigma_y + C_3},$$

$$\mathrm{SSIM}(\mathbf{x}, \mathbf{y}) = [l(\mathbf{x}, \mathbf{y})]^{\alpha} \cdot [c(\mathbf{x}, \mathbf{y})]^{\beta} \cdot [s(\mathbf{x}, \mathbf{y})]^{\gamma}$$

Based on papers by Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli

Spatial SSIM Metric

image patches

mean

covariance

$$S(\mathbf{x}, \mathbf{y}) = \frac{(2\,\mu_x\,\mu_y + C_1)\,(2\,\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)\,(\sigma_x^2 + \sigma_y^2 + C_2)}$$

small constants

variance

CW-SSIM Metric

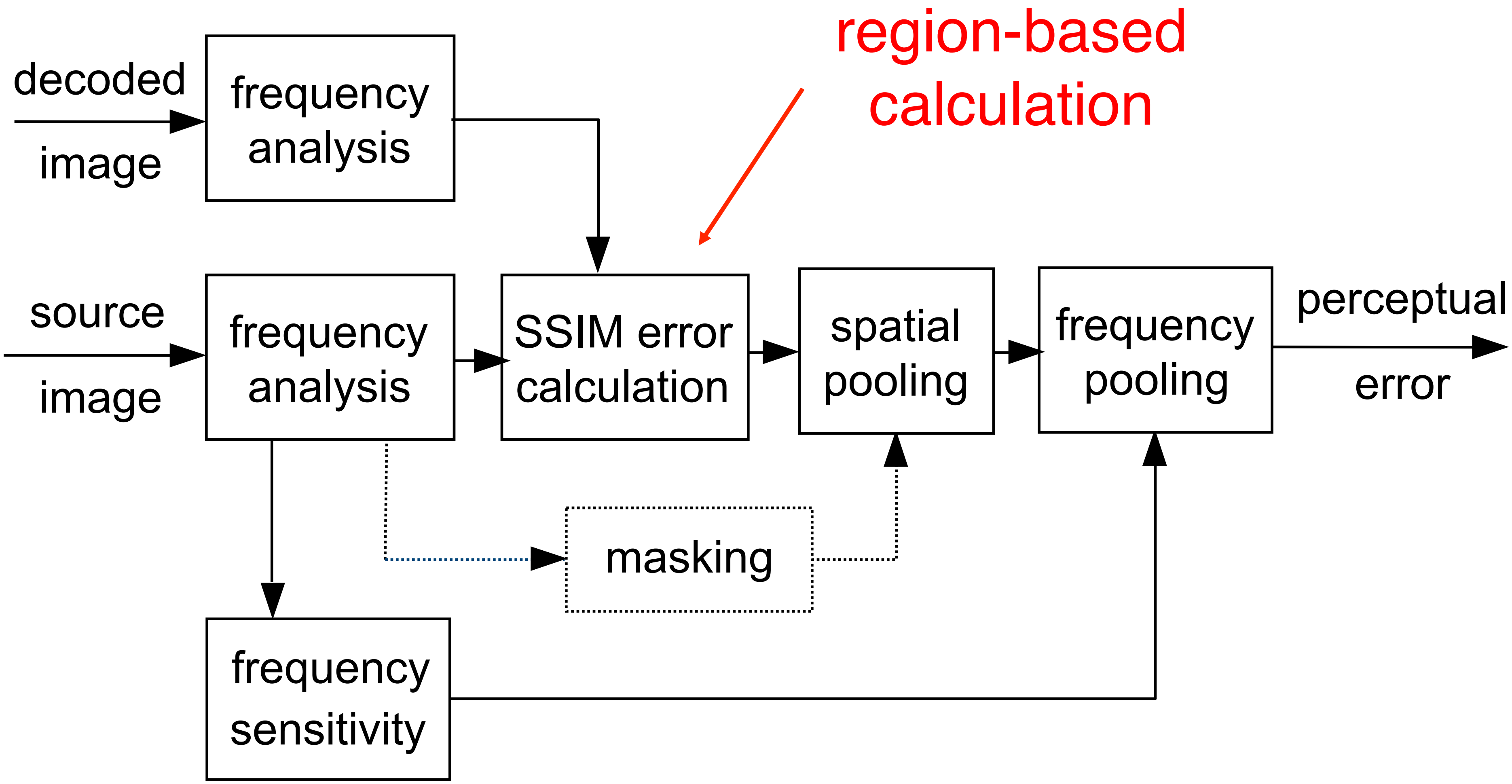
note: wavelet coefficients are zero mean because the filters are bandpass

small constant

$$\tilde{S}(\mathbf{c}_x, \mathbf{c}_y) = \frac{2\,|\sum_{i=1}^{N} c_{x,i}\,c_{y,i}^*| + K}{\sum_{i=1}^{N} |c_{x,i}|^2 + \sum_{i=1}^{N} |c_{y,i}|^2 + K}$$

complex wavelet coefficients

WCWSSIM (Perceptually-Weighted)



New SSIMs for Texture

$$l(\mathbf{x},\mathbf{y}) = \frac{2\,\mu_x\,\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}$$

$$c(\mathbf{x},\mathbf{y}) = \frac{2\,\sigma_x\,\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}$$

- Compare local image statistics
- Add spatial correlations
- Add correlations across subbands
- Add color composition

$$s(\mathbf{x},\mathbf{y}) = \frac{\sigma_{xy} + C_3}{\sigma_x\,\sigma_y + C_3},$$

- Point-by-point
- Modify to allow relative shifts in the two image patches

$$\mathrm{SSIM}(\mathbf{x},\mathbf{y}) = [l(\mathbf{x},\mathbf{y})]^{\alpha} \cdot [c(\mathbf{x},\mathbf{y})]^{\beta} \cdot [s(\mathbf{x},\mathbf{y})]^{\gamma}$$

Based on papers by Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli

Texture Similarity Metrics

- No point-by-point comparisons
  - Drop structure term
- Local image statistics
  - Include more parameters as texture descriptors
  - First order correlation coefficients
- Texture synthesis [Portilla&Simoncelli'00]
- Image statistics and the perception of surface qualities [Motoyoshi,Nishida,Sharan,Adelson'07]

Texture Synthesis [Portilla-Simoncelli'00]

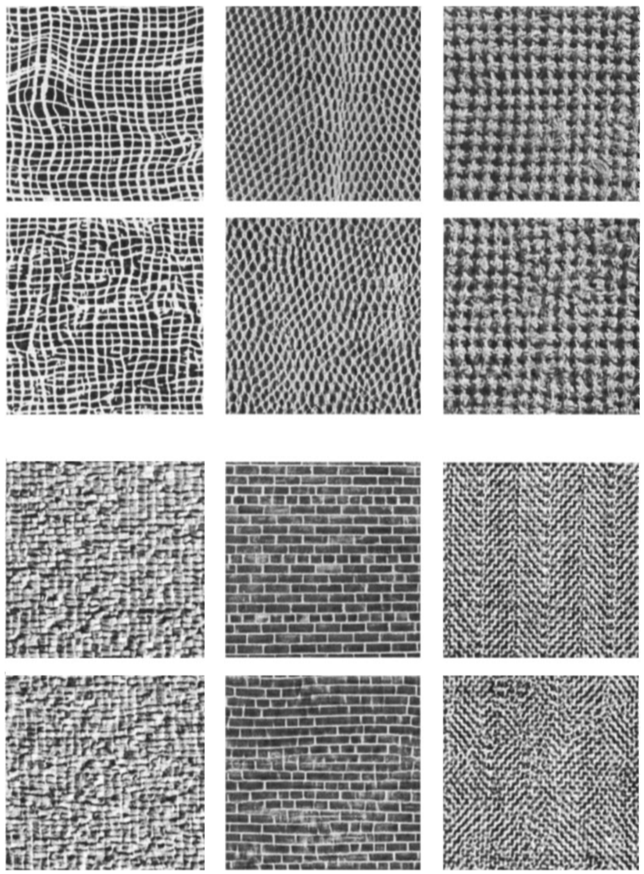


Figure 14. Synthesis results on photographic pseudo-periodic textures. See caption of Fig. 12.

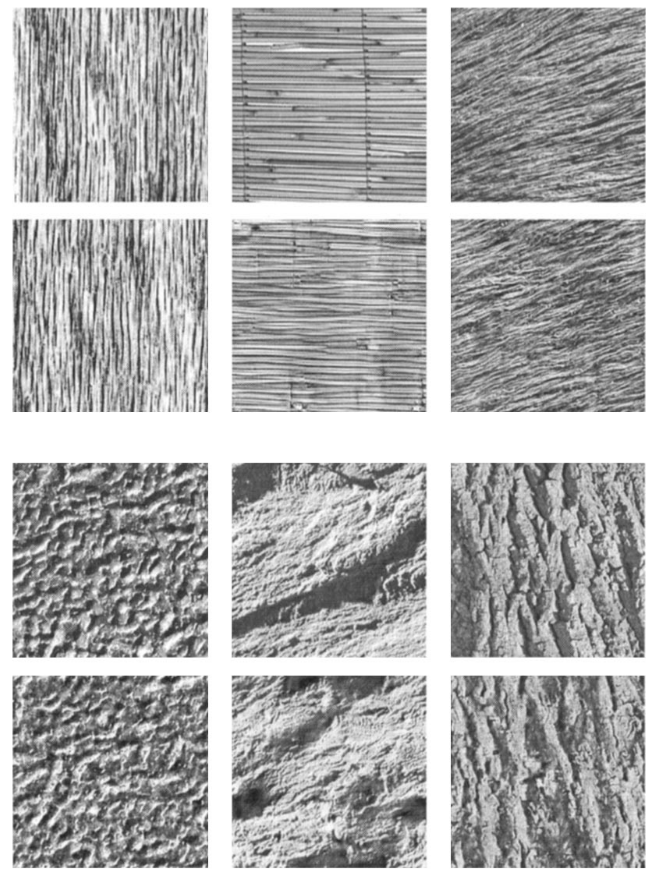


Figure 15. Synthesis results on photographic aperiodic textures. See caption of Fig. 12.

First-order Correlation Coefficients

Horizontal $$\rho_x(0,1) = \frac{E[(x_{ij} - \mu_x)(x_{i(j+1)} - \mu_x)]}{\sigma_x^2}$$

Vertical $$\rho_x(1,0) = \frac{E[(x_{ij} - \mu_x)(x_{(i+1)j} - \mu_x)]}{\sigma_x^2}$$

Diagonal $$\rho_x(1,1) = \frac{E[(x_{ij} - \mu_x)(x_{(i+1)(j+1)} - \mu_x)]}{\sigma_x^2}$$

Anti-diagonal $$\rho_x(1,-1) = \frac{E[(x_{ij} - \mu_x)(x_{(i+1)(j-1)} - \mu_x)]}{\sigma_x^2}.$$

New SSIM Terms

- Original SSIM terms:

$$l(\mathbf{x},\mathbf{y}) = \frac{2\,\mu_x\,\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}$$

$$c(\mathbf{x},\mathbf{y}) = \frac{2\,\sigma_x\,\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}$$

$$s(\mathbf{x},\mathbf{y}) = \frac{\sigma_{xy} + C_3}{\sigma_x\,\sigma_y + C_3},$$

- New SSIM terms need new comparison formula because $\rho_x(0,1)$ takes values in interval [-1,1].

New SSIM Terms

Horizontal: $C_{0,1}(x,y) = 1 - 0.5\,|\rho_x(0,1) - \rho_y(0,1)|^p$

Vertical: $C_{1,0}(x,y) = 1 - 0.5\,|\rho_x(1,0) - \rho_y(1,0)|^p$

Diagonal: $C_{1,1}(x,y) = 1 - 0.5\,|\rho_x(1,1) - \rho_y(1,1)|^p$

Anti-diagonal: $C_{1,-1}(x,y) = 1 - 0.5\,|\rho_x(1,-1) - \rho_y(1,-1)|^p$

- Symmetric with respect to x and y
- Values in [0,1]
- Unique maximum when x=y

Higher-order Statistics

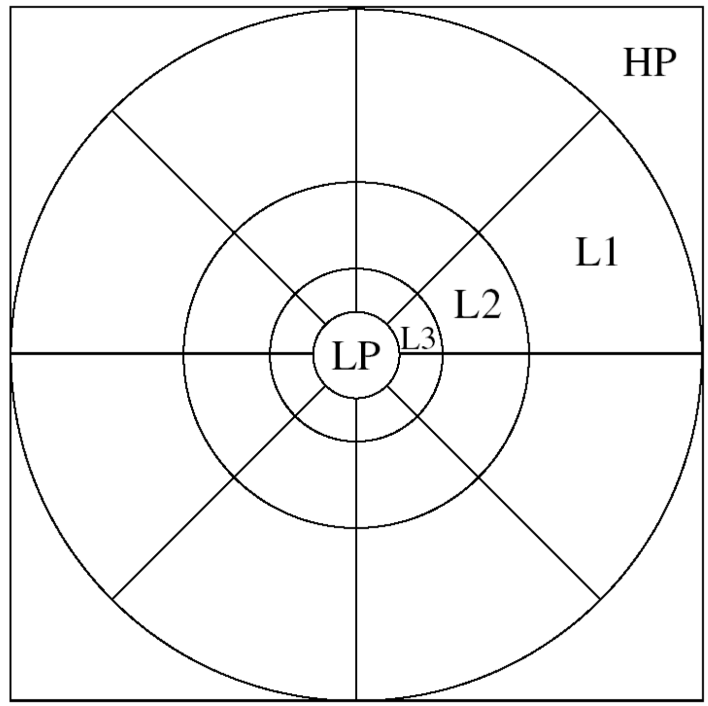
- Higher-order autocovariances
  - Equivalent to first-order autocovariances of decimated images (aliasing)

$$\Updownarrow$$

- First-order autocovariances of filtered and decimated images (no aliasing)

Multi-Resolution Analysis

- Steerable filters



- Compute the metric value at each location and resolution

STSIM [Zhao et al., ICIP'08]

- Keep L, C terms (complex domain)
- Add spatial correlation terms $c_{0,1}$ and $c_{1,0}$

$$\rho_x^k(0,1) = \frac{E\{(x_{k,i,j} - \mu_x)(x_{k,i,j+1} - \mu_x)\}}{\sigma_{x_k}^2}$$

$$c_{0,1}^k(\mathbf{x},\mathbf{y}) = 1 - 0.5\left(|\rho_x^k(0,1) - \rho_y^k(0,1)|\right)^p$$

$$\boxed{Q_{\mathrm{stsim}}^k(\mathbf{x},\mathbf{y}) = l^k(\mathbf{x},\mathbf{y})^{\frac{1}{4}} c^k(\mathbf{x},\mathbf{y})^{\frac{1}{4}} c_{0,1}^k(\mathbf{x},\mathbf{y})^{\frac{1}{4}} c_{1,0}^k(\mathbf{x},\mathbf{y})^{\frac{1}{4}}}$$

k - k^{th} subband, $p = 1$

Error Pooling

- Additive error pooling
  - Over space and frequency
    + radial and angular orientation
  - Consistent with CWSSIM
  - Can incorporate perceptual weights (WCWSSIM)
- Multiplicative error pooling
  - Multiply across frequency
  - Normalize by taking m-th root (m subbands)
  - Spatial average
  - Weakest matches dominate

STSIM2 [Zujovic *et al.*, ICIP'09]

- Add inter-subband cross-correlations

$$\rho_x^{k,l}(0,0) = \frac{E\{(|x_{k,i,j}| - \mu_{x_k})(|x_{l,i,j}| - \mu_{x_l})\}}{\sigma_{x_k}\sigma_{x_l}}$$

$$c_{0,0}^{k,l}(\mathbf{x},\mathbf{y}) = 1 - 0.5\left(|\rho_x^{k,l}(0,0) - \rho_y^{k,l}(0,0)|\right)^p$$

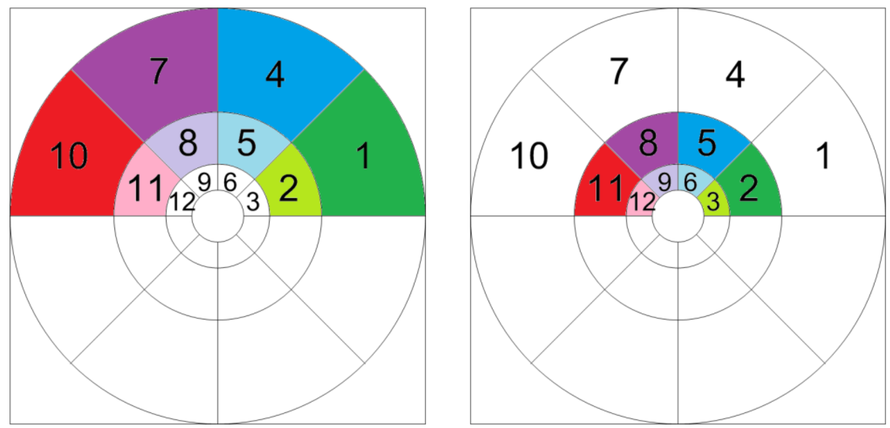
- Pooling

$$p = 1$$

$$\boxed{Q_t(\mathbf{x},\mathbf{y}) = \frac{1}{N_t}\left(\sum_k Q_{\text{stsim}}^k(\mathbf{x},\mathbf{y}) + \sum_{k,l} c_{0,0}^{k,l}(\mathbf{x},\mathbf{y})\right)}$$

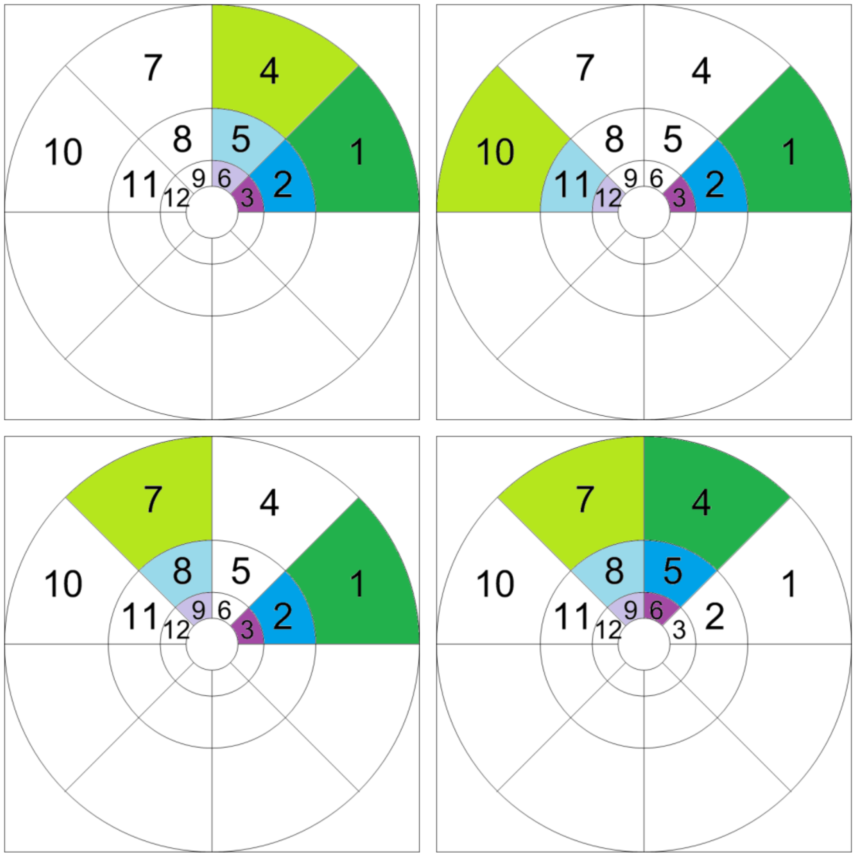
STSIM2

- 3-scale, 4-orientation decomposition
- Add inter-subband correlations
  - Adjacent scales, for same orientation (e.g., 2&1, 3&2, 5&4)
  - Across orientations, for same scale (e.g., 1&4, 1&7, 1&10)



STSIM2

- 3-scale, 4-orientation decomposition
- Add inter-subband correlations
  - Adjacent scales, for same orientation (e.g., 2&1, 3&2, 5&4)
  - Across orientations, for same scale (e.g., 1&4, 1&7, 1&10)



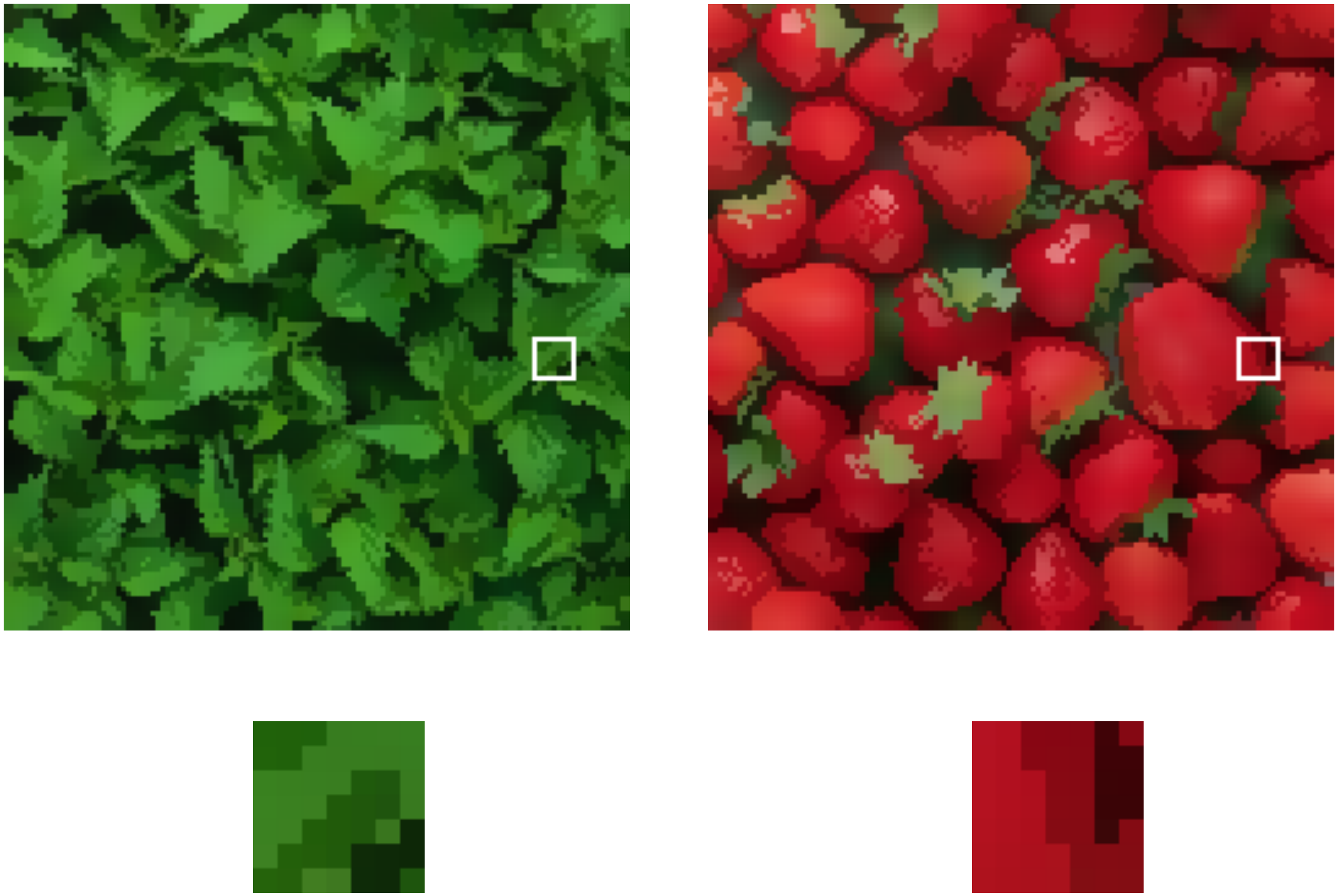
Color Composition Similarity

[Zujovic *et al.*, ICIP'09]

- Traditional methods
  - Raw color histogram comparisons
- Our approach
  - Remove unnecessary color detail
    - + Extract dominant colors
    - + Using adaptive clustering [Pappas'92] as a prefilter
  - Use more sophisticated distance metric
    - + EMD [Rubner'00], OCCD [Mojsilovic'02]
  - Use "perceptually uniform" color space (L*a*b*)

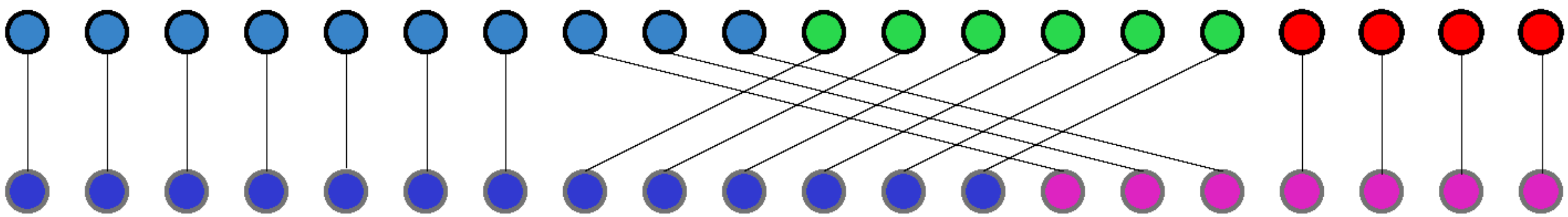
Color Composition Similarity

Color Composition Similarity



Optimal Color Composition Distance

- Minimum cost graph matching problem
- Quantize percentages of colors into “units”
- Example: 5% units = 20 units total



Mojsilovic, *et al.*, TIP’02

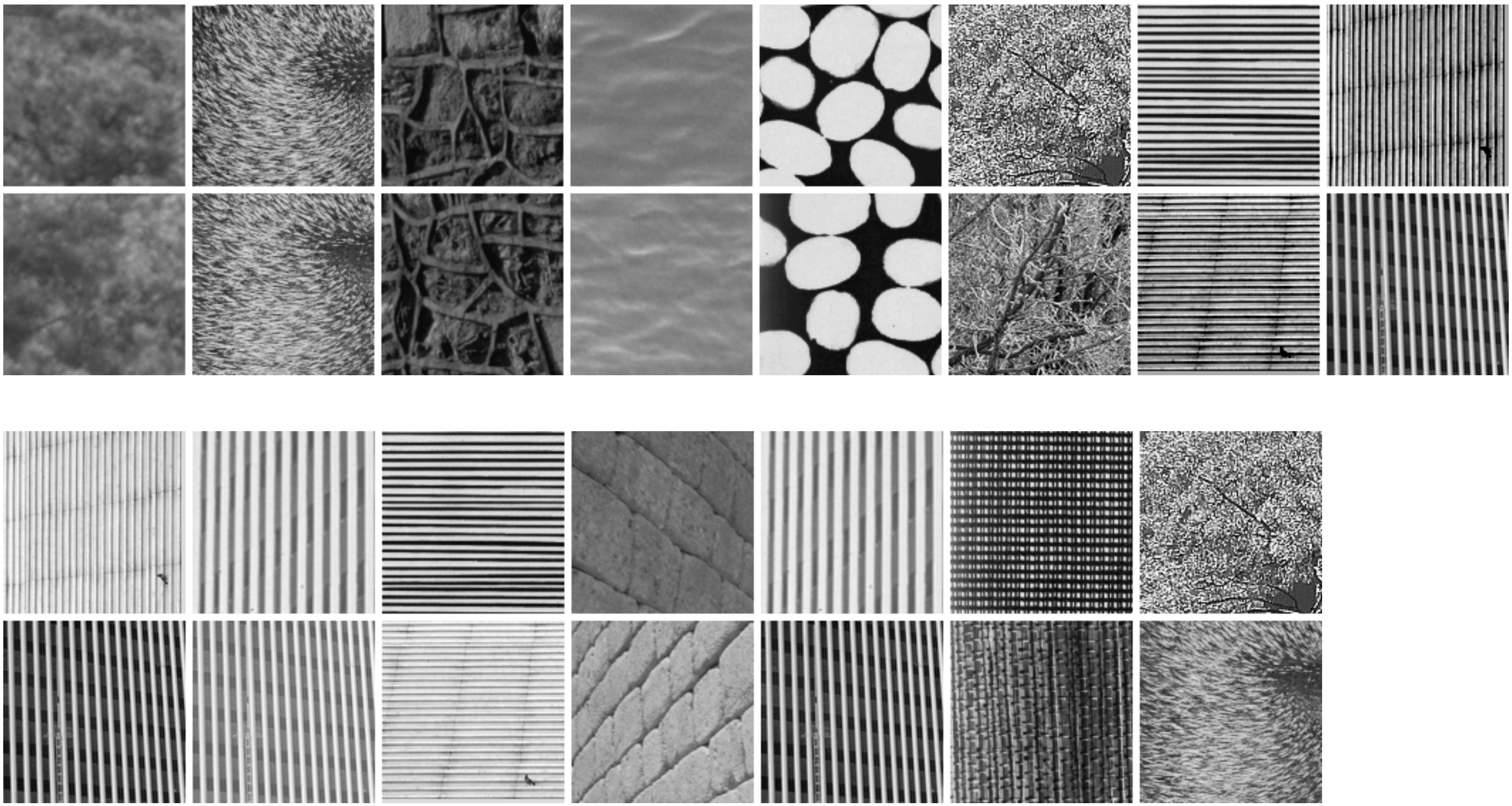
Color Composition Similarity

- Compute OCCD distance over small windows
- This gives us a map of distances OCCD
- Similarity map is OCCDs = 1 – OCCD

Texture Similarity Metric Evaluation

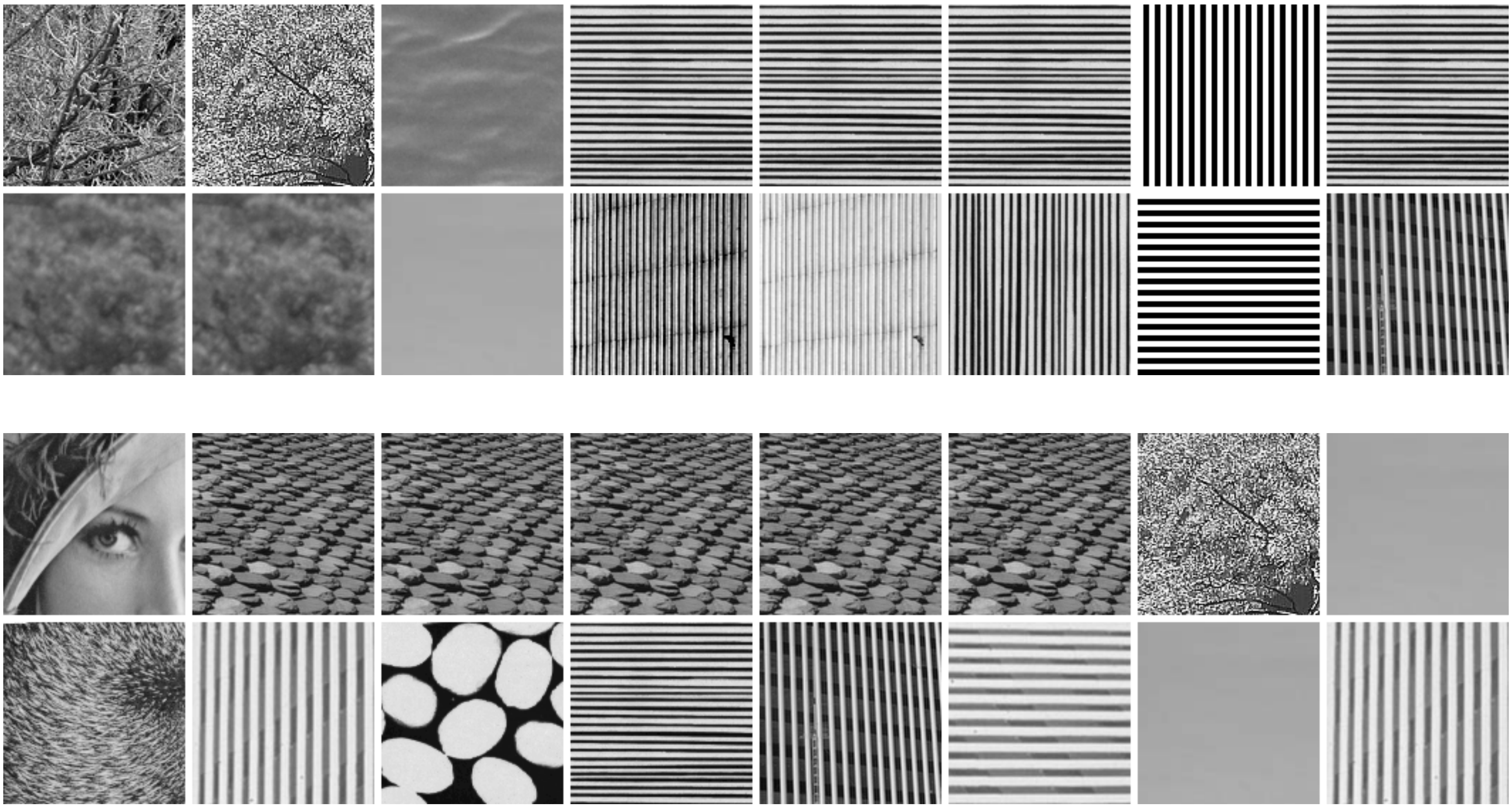
- Texture similarity metrics vs. image quality metrics
- Compression: Quantify distortion
- Retrieval: Distinguish similar-dissimilar
  - Needs extensive subjective tests
- Retrieval: Known-Item Search
  - Easy to obtain a lot of data (assuming uniform textures)
  - Goal: Find metric that accurately retrieves textures similar to query
  - Similarity defined as coming from same bigger texture image
  - Evaluation criteria: Precision@1, MRR, MAP, Precision-Recall plot

Texture Database: Similar Pairs



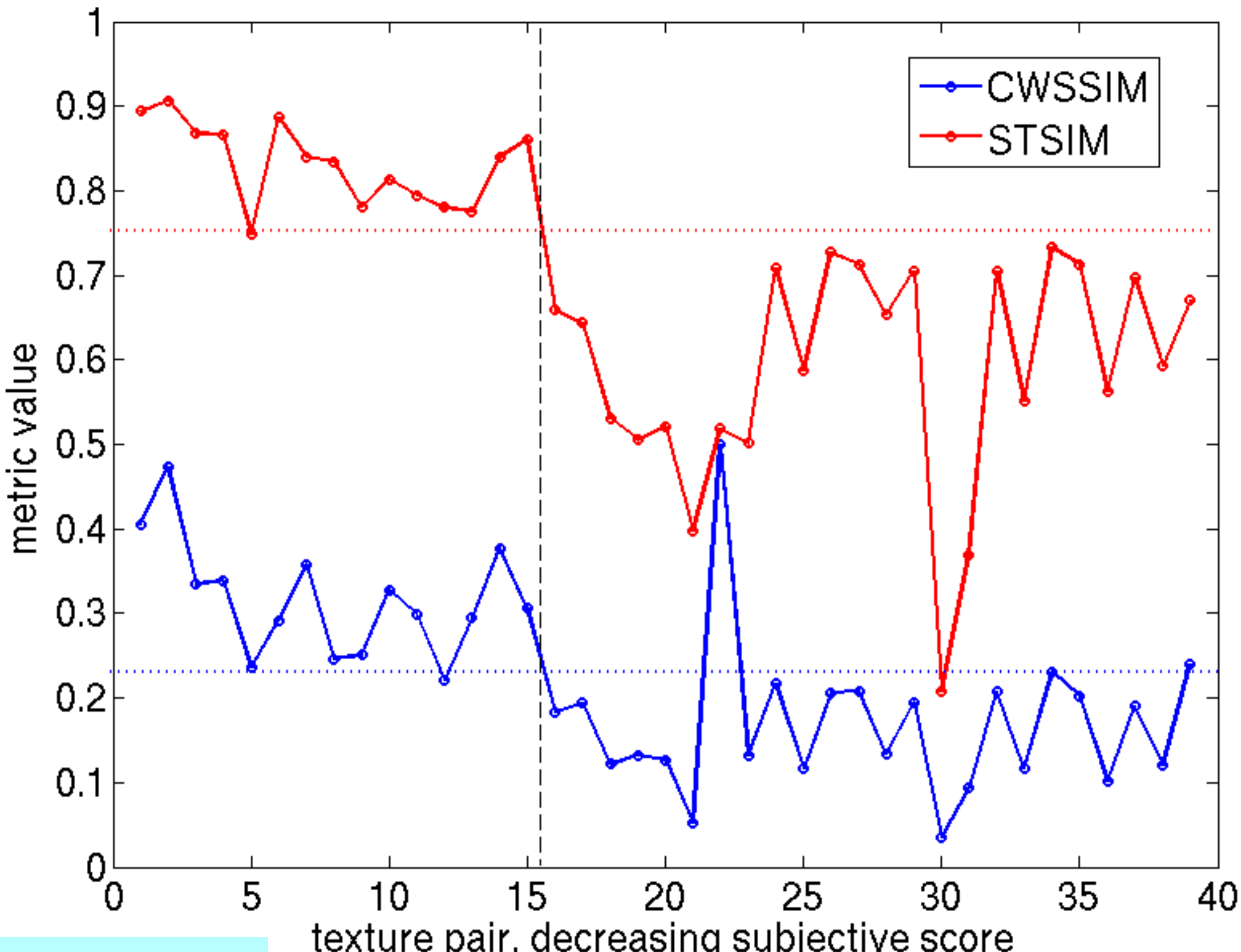
Zhao, *et al.*, ICIP'08

Texture Database: Dissimilar Pairs



Zhao, *et al.*, ICIP'08

STSIM - Additive Error Pooling



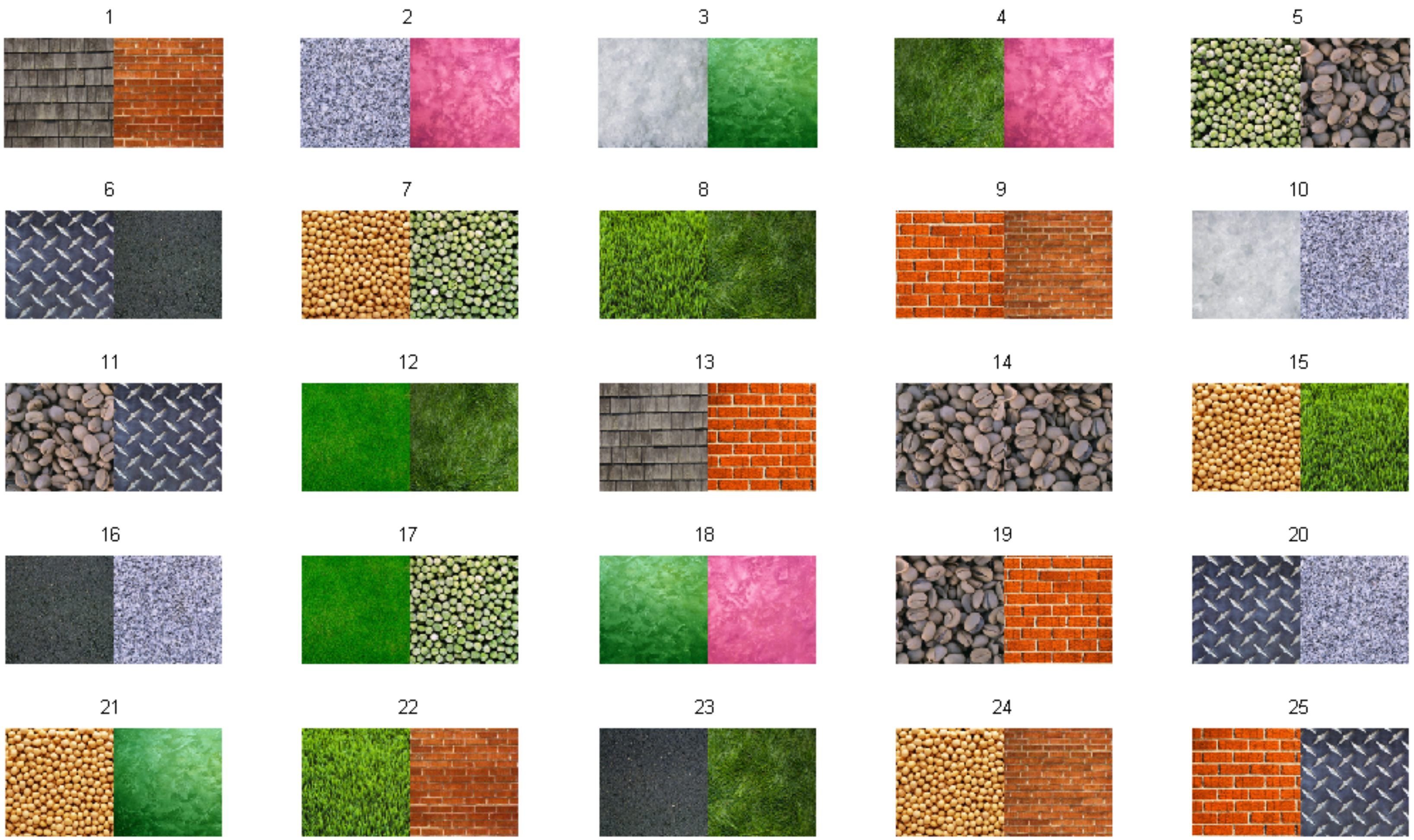
Zhao, *et al.*, ICIP'08

STSIM2 - Statistical Ranking Comparisons

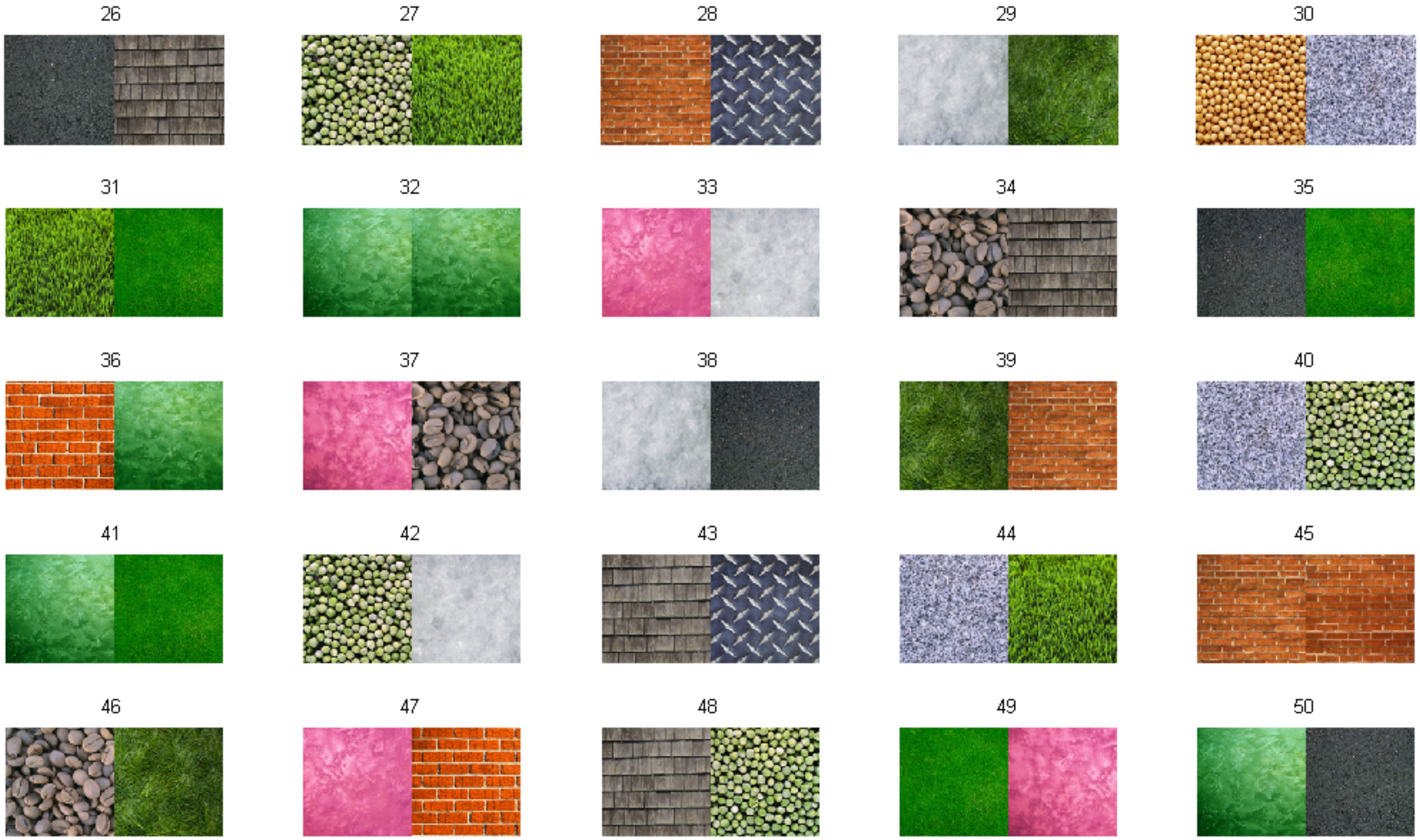
- Informal subjective test, 50 pairs of textures from pool of 30 different images
- Performance evaluation: Kendall’s and Spearman’s statistical tests (ranking comparisons)
- Variables: subsets of cross-correlation terms, weights for texture similarity and color similarity

Zujovic, *et al.*, ICIP’09

Experiment and Results



Experiment and Results



Experiment and Results

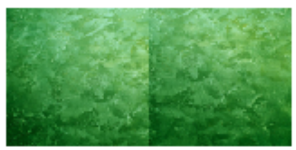
- Best combinations:
  - use all of the new texture similarity maps
  - use 0.6 for texture, 0.4 for color, additive app.
- STSIM2 coefficient of correlation: 0.66
- PSNR: 0.28
- SSIM: 0.51
- STSIM: 0.60
- Human Performance: When correlating one human's grades to other humans' averages, the average correlation is 0.79

Experiment and Results

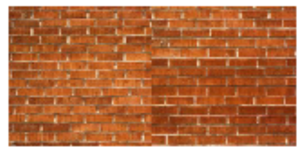
- 5 most similar pairs, by humans:

14

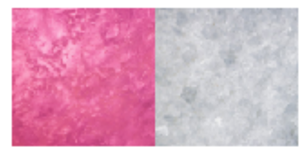
32



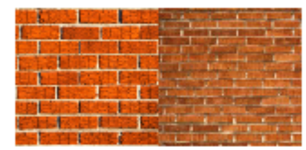
45



33

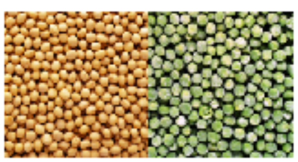


9



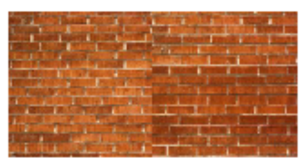
- 5 most similar pairs, according to metric:

7



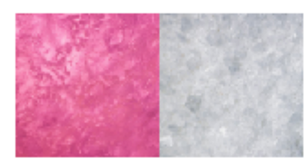
32

45

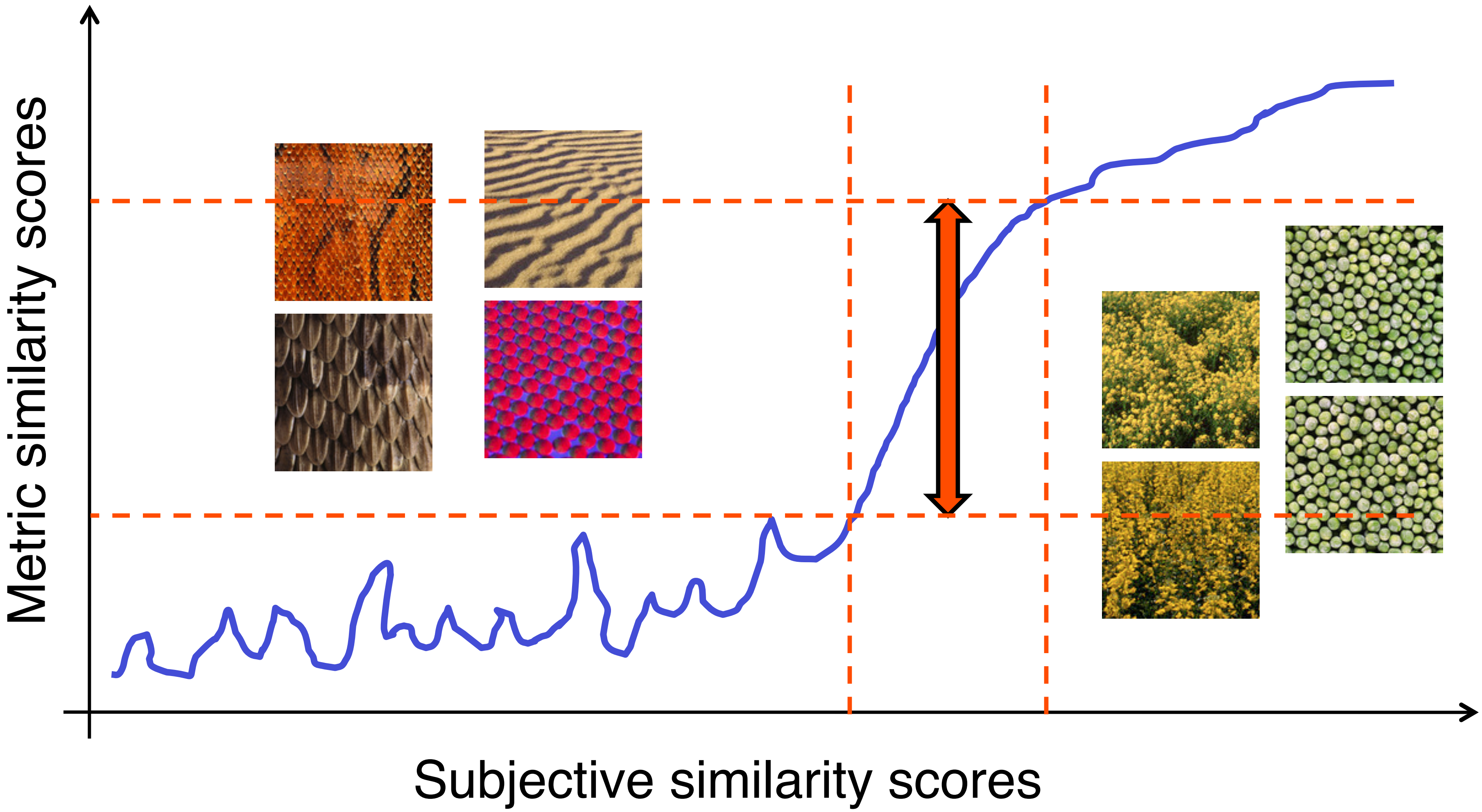


23

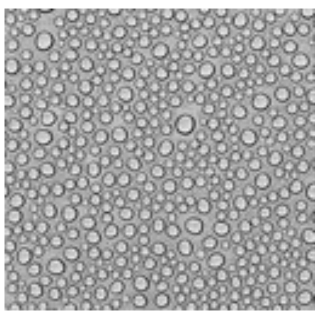
33



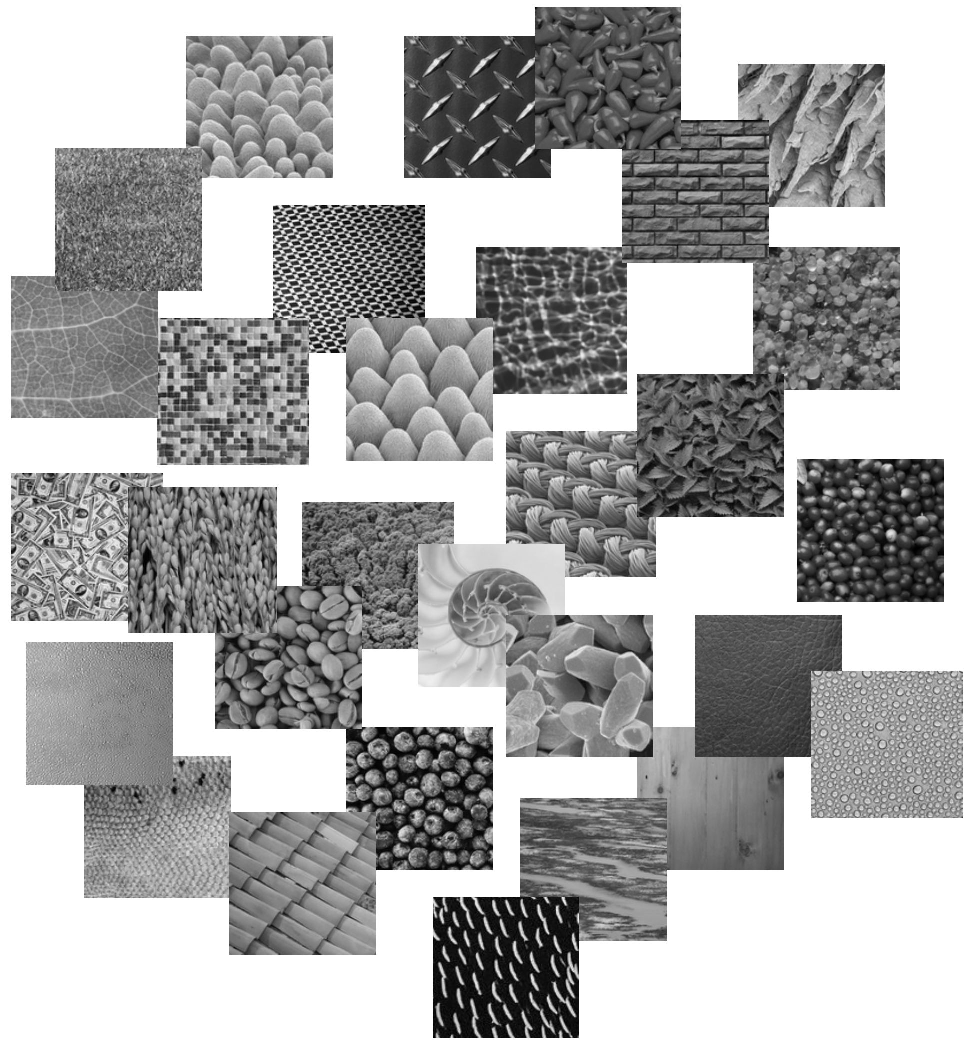
Ideal Metric Performance



Known-Item Search
Problem: Find Query in Database

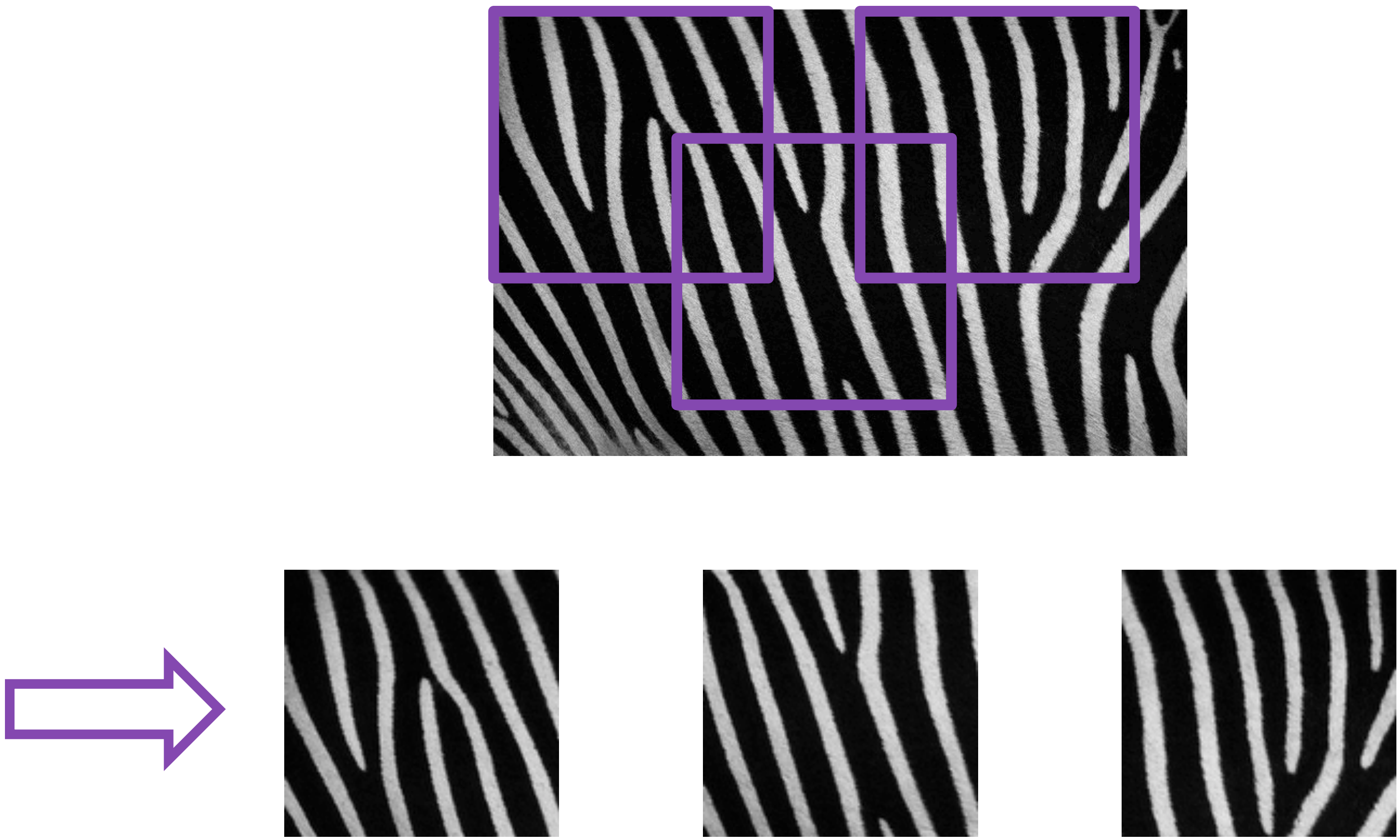


?

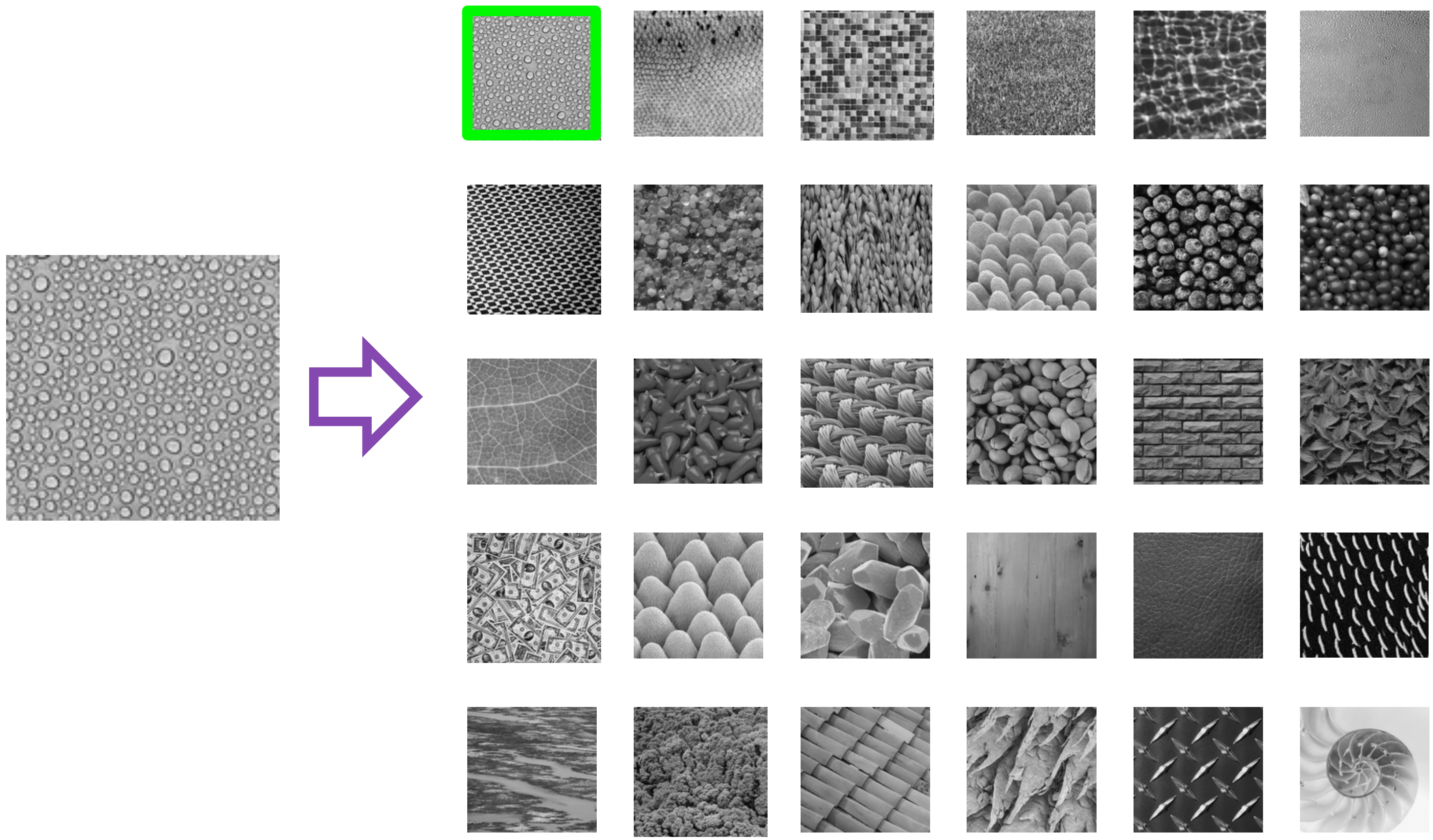


Zujovic, *et al.*, MMSP'09

Building the Database

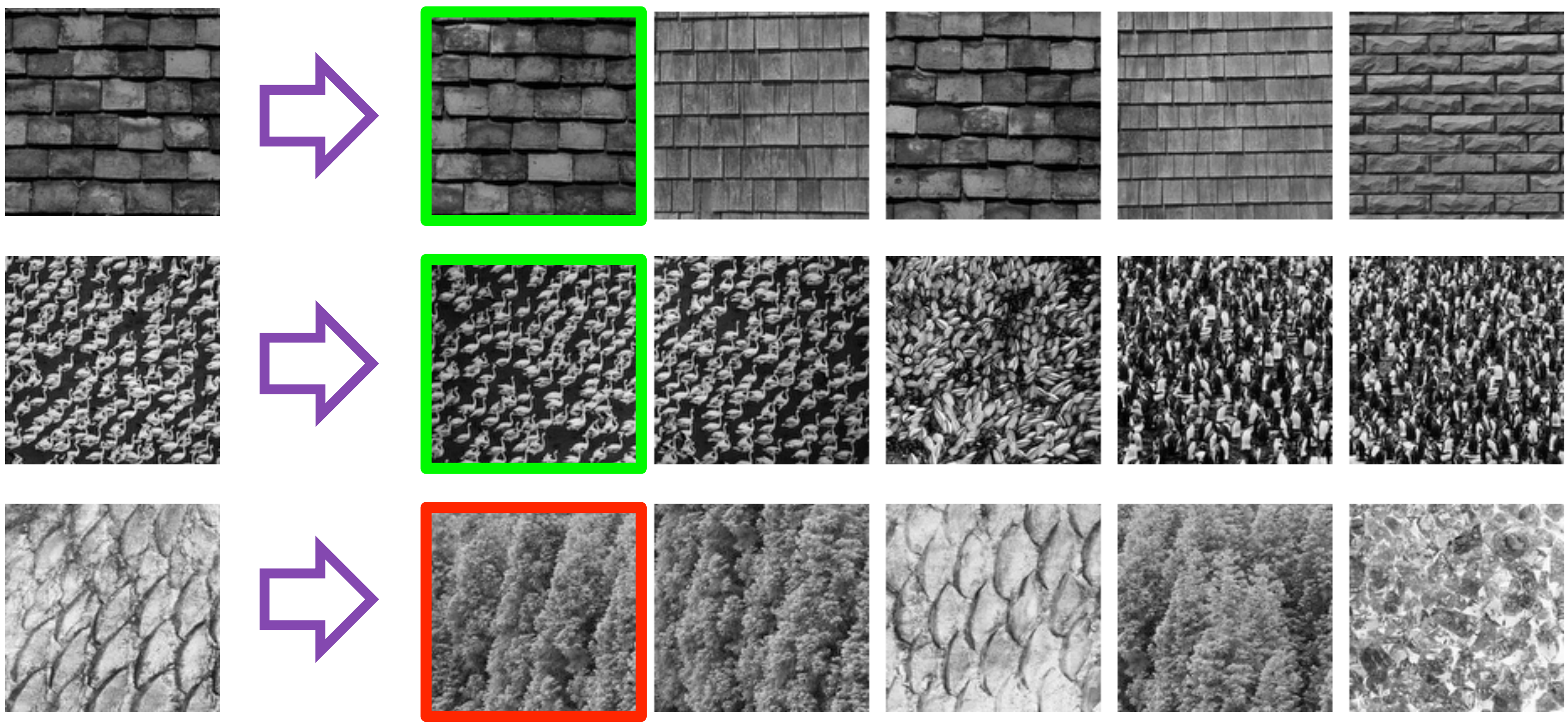


Order Database by similarity



Precision at One

- Measures how many times the first retrieved texture was the correct one

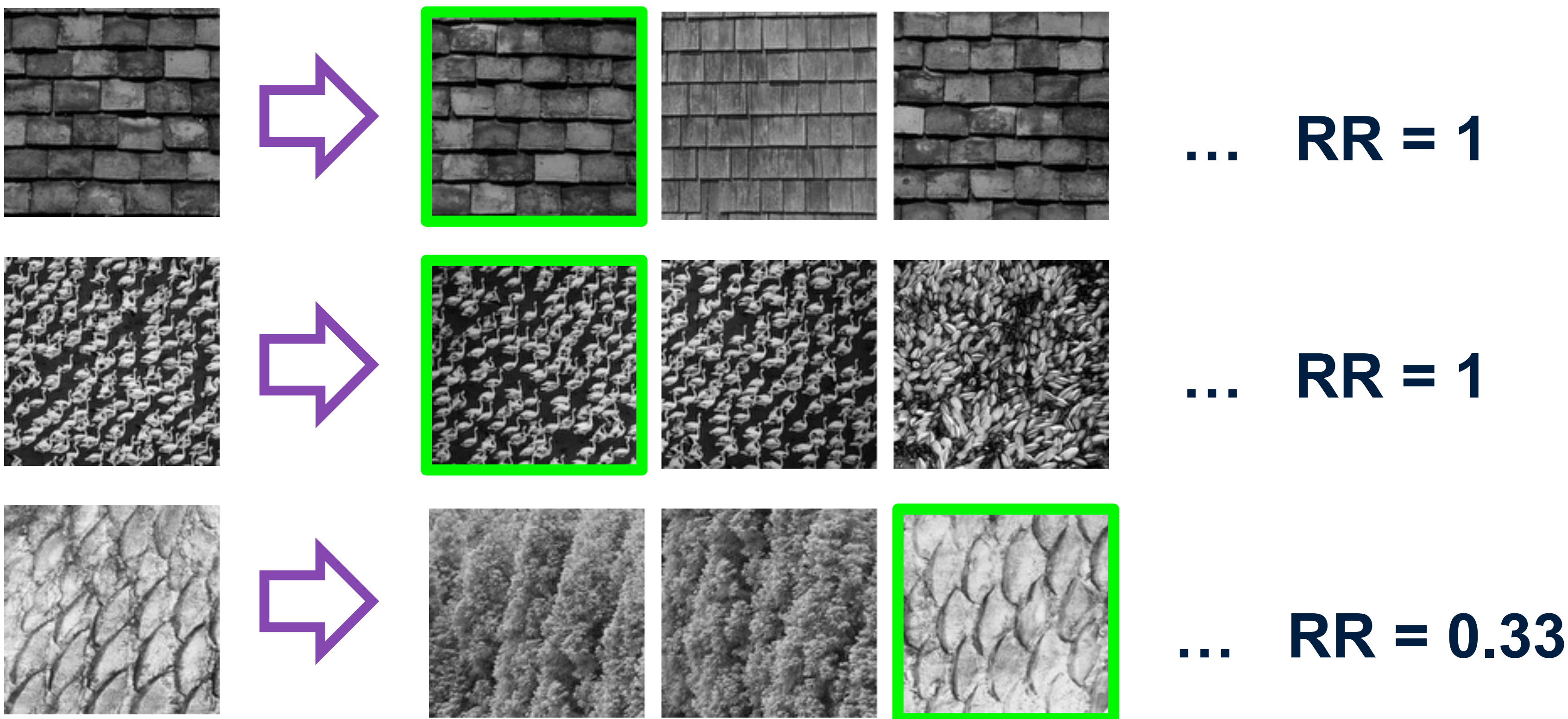


- **Precision at one is computed as percentage**

	PSNR	SSIM	CWSSIM	STSIM2
Precision @1 (%)	6.02	8	63.64	**77.67**

Mean Reciprocal Rank (MRR)

- Measures the average inverse rank of the first correct retrieved image

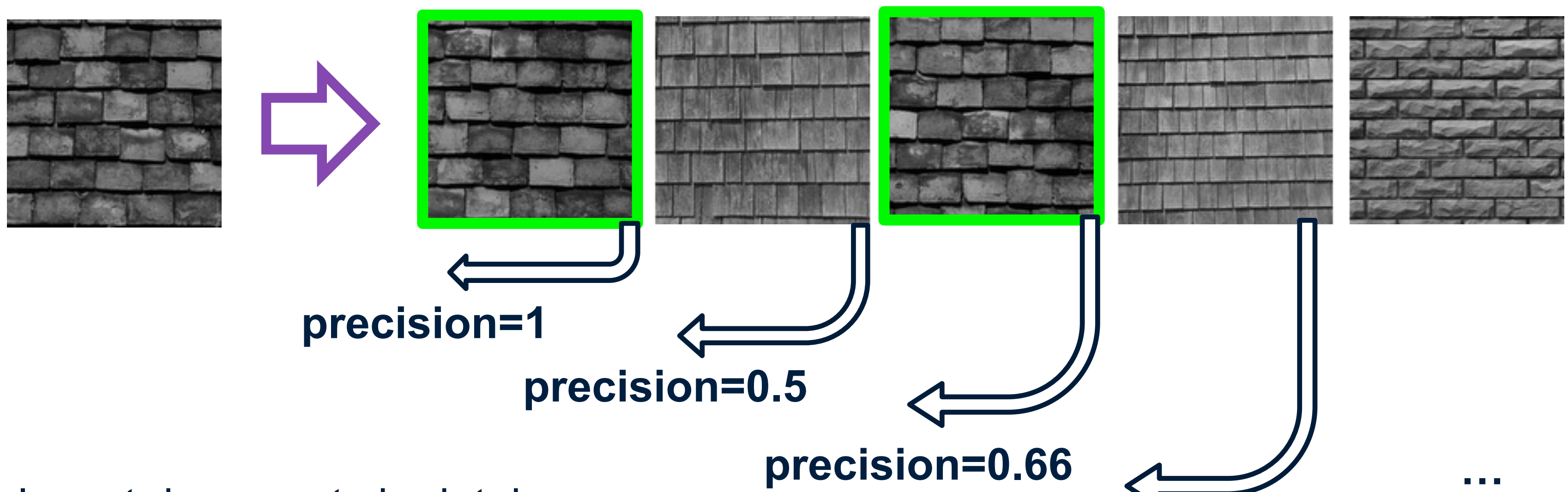


- **MRR is the mean of all reciprocal ranks RR**

	PSNR	SSIM	CWSSIM	STSIM2
MRR	0.10	0.11	0.71	**0.83**

Mean Average Precision (MAP)

- Measures average precision when cutoff is made at 1^{st}, 2^{nd}, … N^{th} retrieved image



two relevant documents in database

indicator that document was relevant

$$AP = 0.5*(1*1 + 0.5*0 + 0.66*1 + 0.5*0 + \dots) = 0.83$$

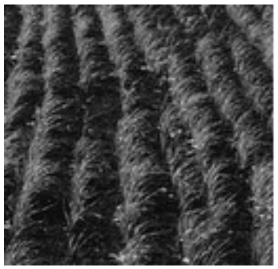
precision after first retrieved document

- **MAP is the mean of all average precisions AP**

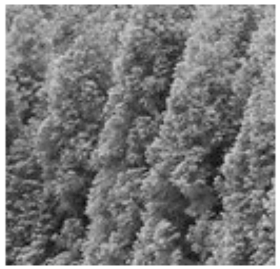
	PSNR	SSIM	CWSSIM	STSIM2
MAP	0.095	0.06	0.62	**0.75**

Failure Examples

query



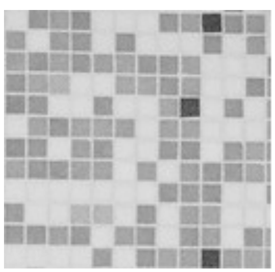
1st match



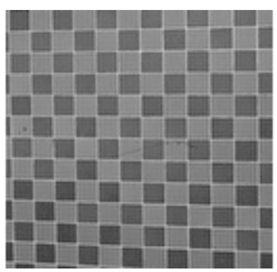
2nd match

Failure due to gray-level difference

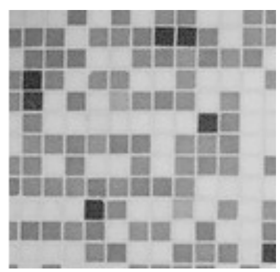
query



1st match



4th match



Failure due to large variation in color

Failure Examples

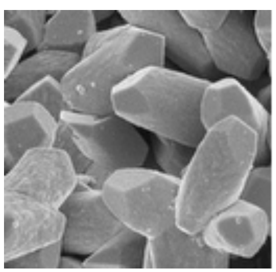
query

1st match

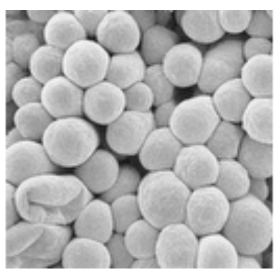
2nd match

Failure due to strong texture elements and scale issues

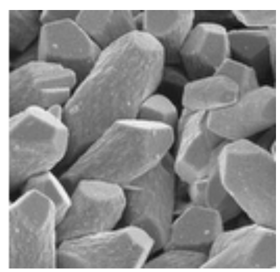
query



1st match



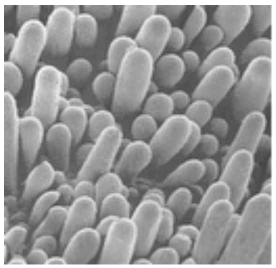
7th match



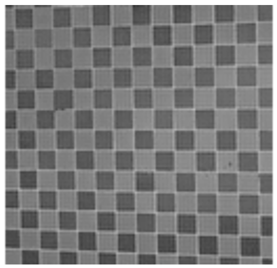
Failure due to scale issues

Failure Examples

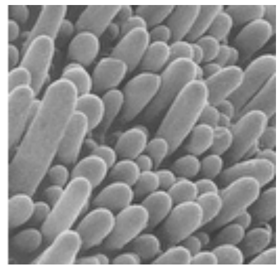
query



1st match

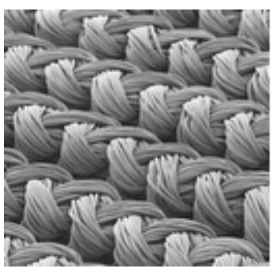


2nd match



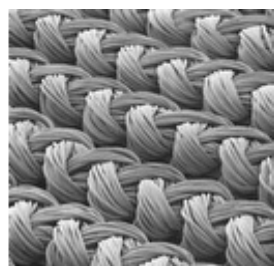
Failure due to inability to capture difference in texture orientation

query



1st match

14th match



Failure due to inability to capture periodicity in texture


Thank you
Questions?