Presentation:
Reviews and talks on ICME 2016

Cho-Ying Wu

Disp Lab

Graduate Institute of
Communication Engineering
National Taiwan University

December 07, 2016

Introduction

ICME is a high-class conference on Multimedia.

Multimedia ?

Graphics

Speech

Vision

Display

Data Mining

Video

Introduction

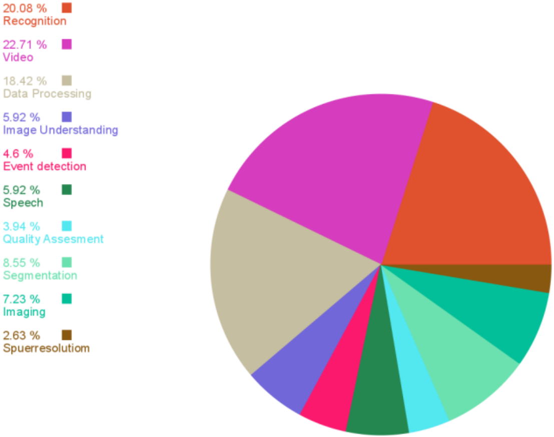
Top conference on multimedia and related

CVPR / ICCV, ECCV, ACM MM

Generally, ICME is suitable for MS(30%), PhD(65%) students to challenge. Recruitments from industrial are also a lot.

Introduction

Topics this year : (Video and Vision are the most popular in the trend)



Roadmap

1 Impressive Talks

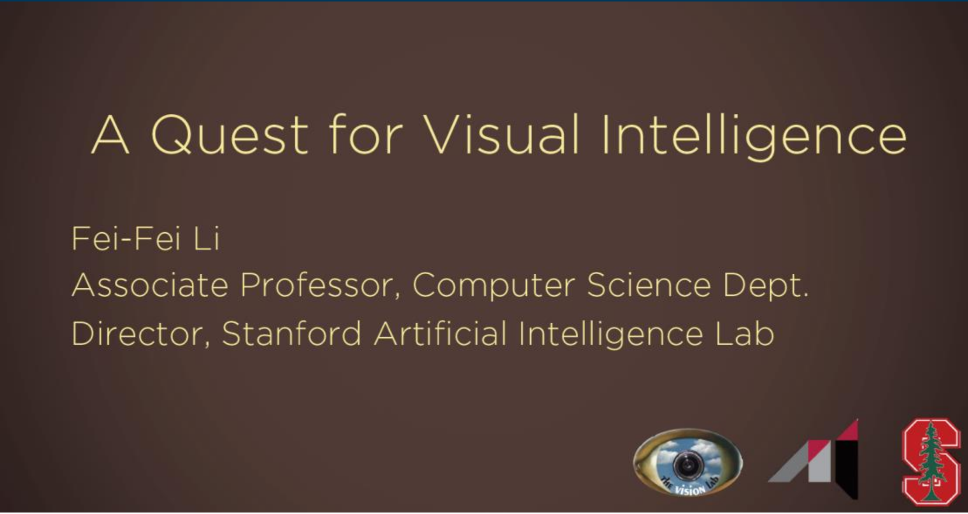
(1) Fei-Fei Li, Associate Professor, "A Quest for Visual Intelligence."Computer Science Dept. Director, Stanford Artificial Intelligence Lab.

(2) Mathias Wien, "High Efficiency Video Coding – Coding Tools and Specification: HEVC V3 and Coming Developments"

2 Interesting Works

(1) Gene Cheung, Xian Ming Liu "Graph Signal Processing for Image Compression and Restoration," Nii.
(2) Xian Ming Liu, "Random Walk Graph Laplacian based Smoothness Prior for Soft Decoding of JPEG Images," HIT.

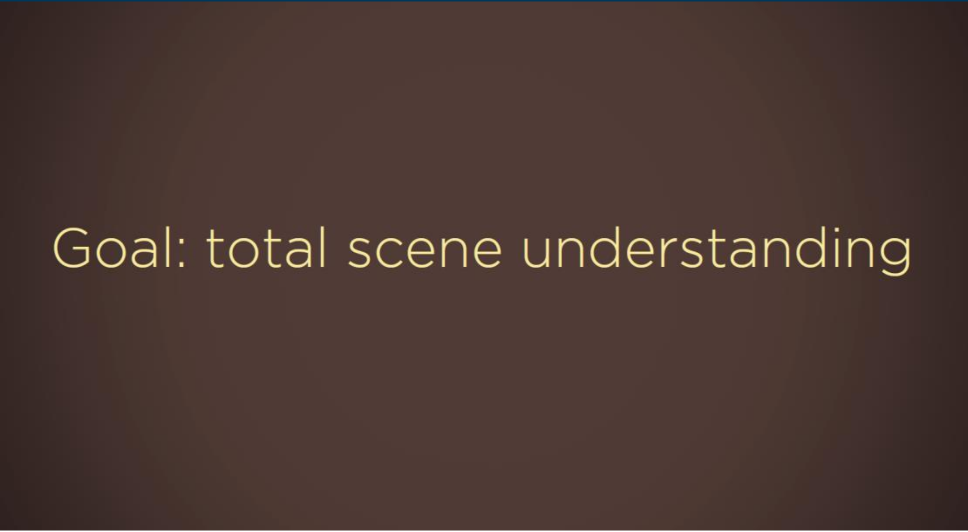
Fei-Fei Li



A Quest for Visual Intelligence

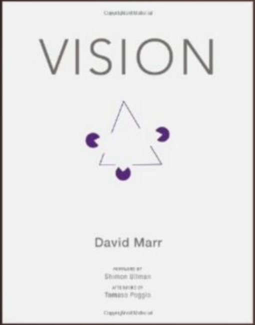
Fei-Fei Li
Associate Professor, Computer Science Dept.
Director, Stanford Artificial Intelligence Lab

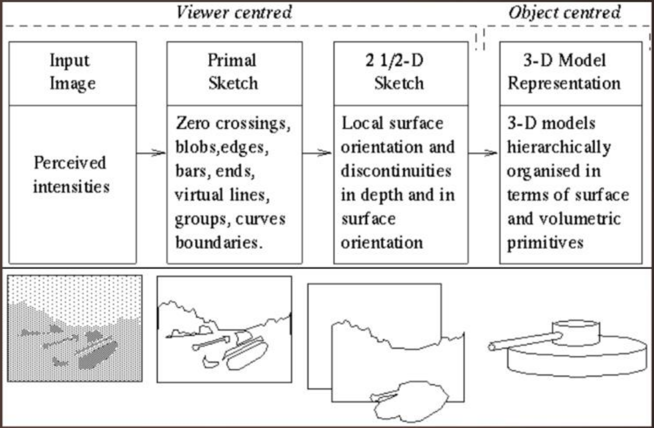
Fei-Fei Li



Goal: total scene understanding

Fei-Fei Li





Viewer centred

Object centred

Input Image

Perceived intensities

Primal Sketch

Zero crossings, blobs,edges, bars, ends, virtual lines, groups, curves boundaries.

2 1/2-D Sketch

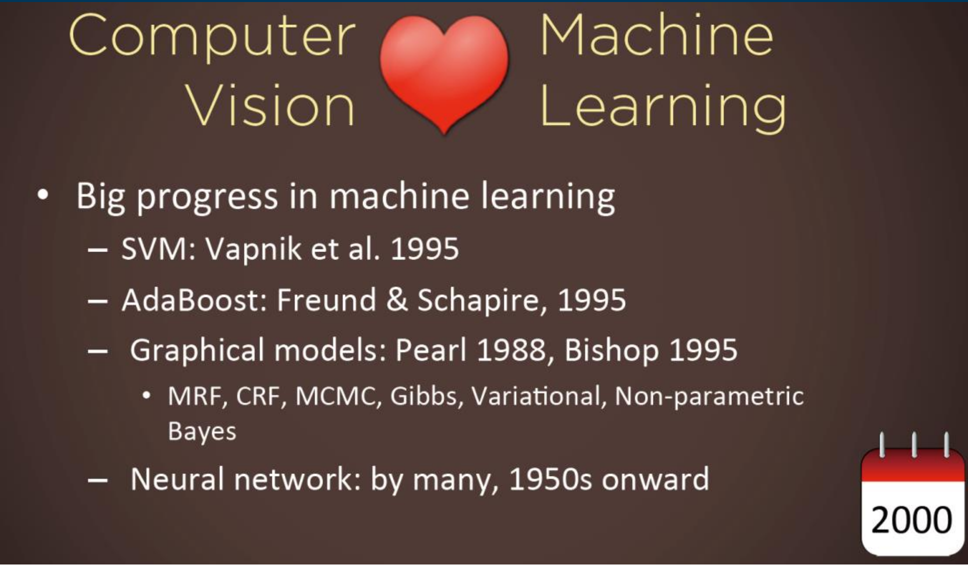
Local surface orientation and discontinuities in depth and in surface orientation

3-D Model Representation

3-D models hierarchically organised in terms of surface and volumetric primitives

D. Marr, 1979

Fei-Fei Li

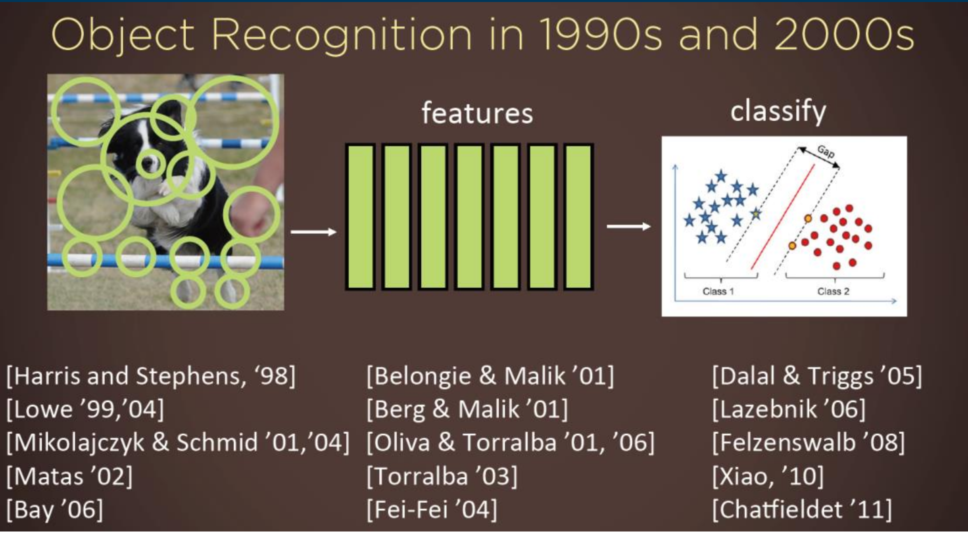


Computer Vision ❤ Machine Learning

- Big progress in machine learning
  - SVM: Vapnik et al. 1995
  - AdaBoost: Freund & Schapire, 1995
  - Graphical models: Pearl 1988, Bishop 1995
    - MRF, CRF, MCMC, Gibbs, Variational, Non-parametric Bayes
  - Neural network: by many, 1950s onward

2000

Fei-Fei Li



Object Recognition in 1990s and 2000s

features

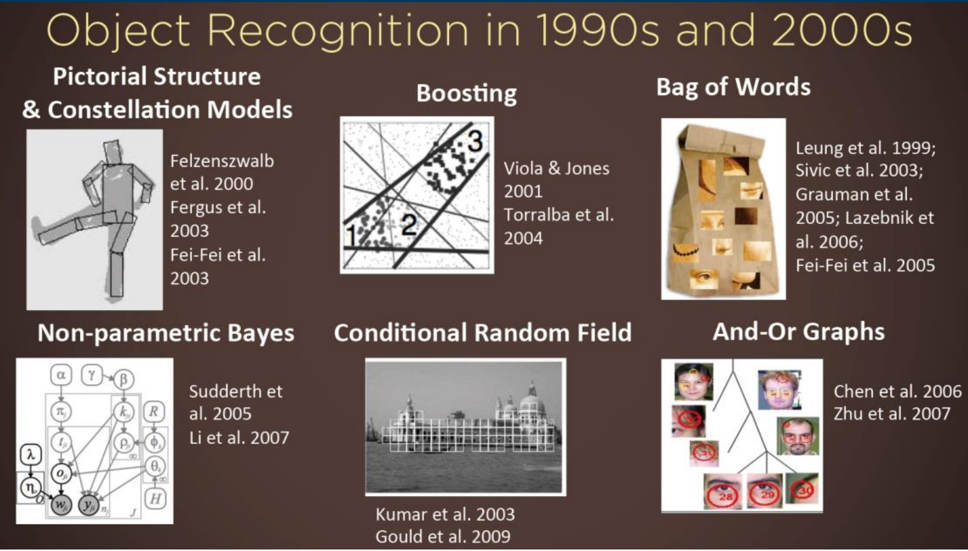
classify

[Harris and Stephens, ‘98]
[Lowe ’99,’04]
[Mikolajczyk & Schmid ’01,’04]
[Matas ’02]
[Bay ’06]

[Belongie & Malik ’01]
[Berg & Malik ’01]
[Oliva & Torralba ’01, ’06]
[Torralba ’03]
[Fei-Fei ’04]

[Dalal & Triggs ’05]
[Lazebnik ’06]
[Felzenswalb ’08]
[Xiao, ’10]
[Chatfieldet ’11]

Fei-Fei Li



Object Recognition in 1990s and 2000s

Pictorial Structure & Constellation Models

Felzenszwalb et al. 2000
Fergus et al. 2003
Fei-Fei et al. 2003

Boosting

Viola & Jones 2001
Torralba et al. 2004

Bag of Words

Leung et al. 1999; Sivic et al. 2003; Grauman et al. 2005; Lazebnik et al. 2006; Fei-Fei et al. 2005

Non-parametric Bayes

Sudderth et al. 2005
Li et al. 2007

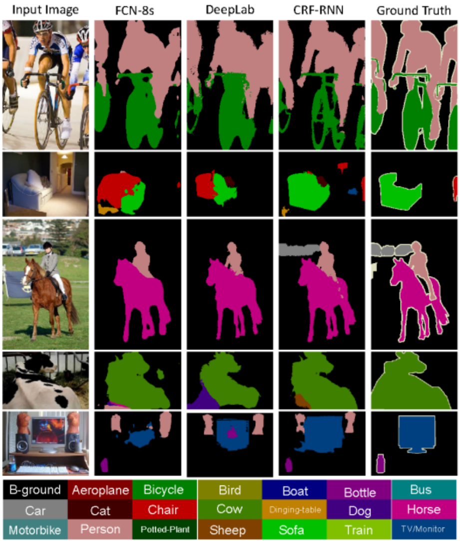
Conditional Random Field

Kumar et al. 2003
Gould et al. 2009

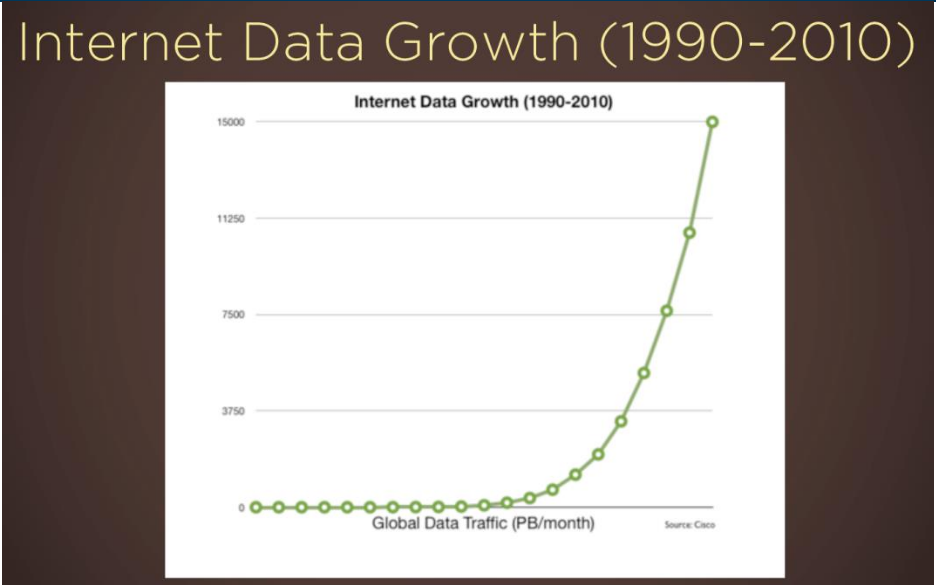
And-Or Graphs

Chen et al. 2006
Zhu et al. 2007

Semantic Segmentation

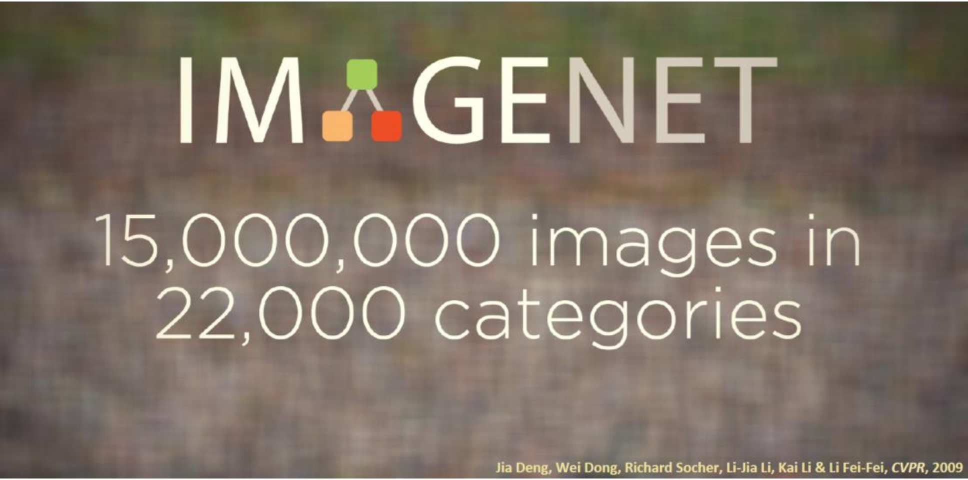


Fei-Fei Li



Internet Data Growth (1990-2010)

Fei-Fei Li

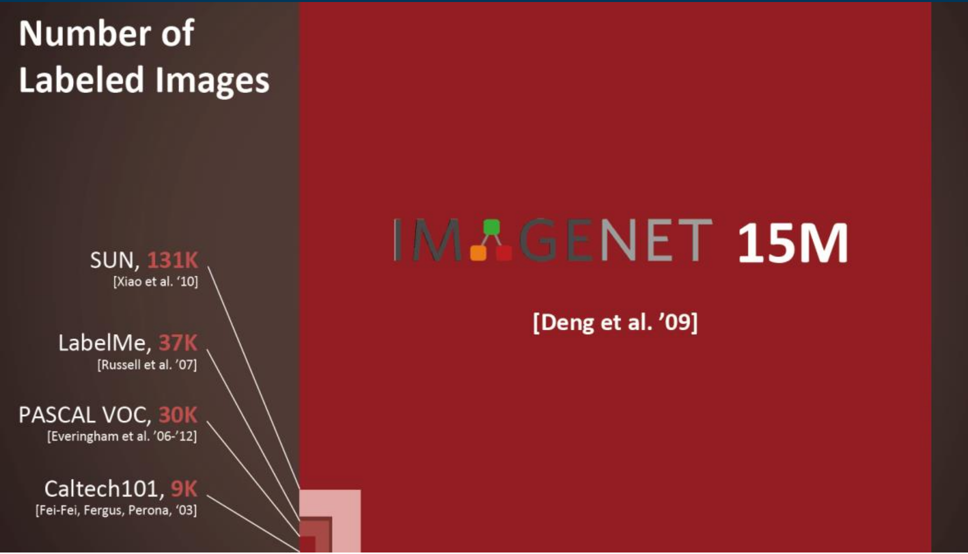


IMAGENET

15,000,000 images in
22,000 categories

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li & Li Fei-Fei, *CVPR*, 2009

Fei-Fei Li



Number of Labeled Images

SUN, 131K
[Xiao et al. '10]

LabelMe, 37K
[Russell et al. '07]

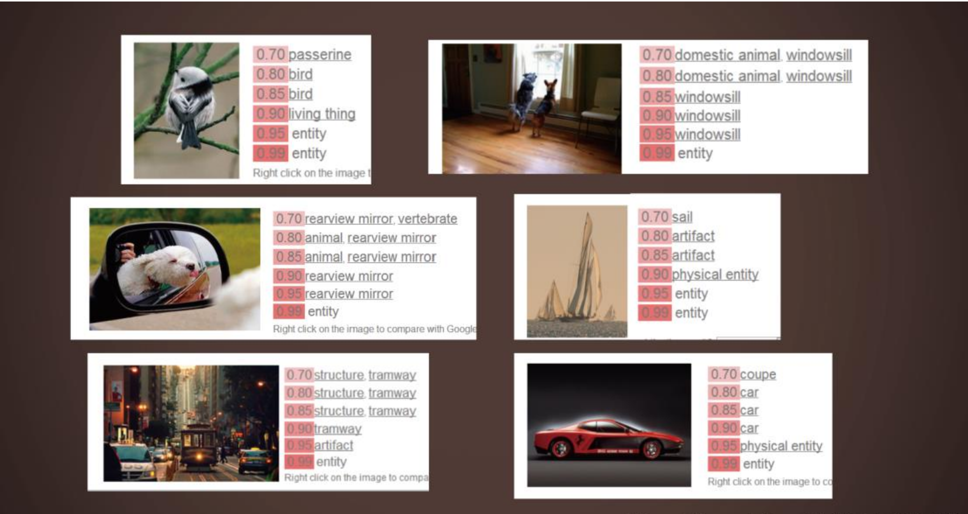
PASCAL VOC, 30K
[Everingham et al. '06-'12]

Caltech101, 9K
[Fei-Fei, Fergus, Perona, '03]

IMAGENET **15M**

[Deng et al. '09]

Fei-Fei Li



0.70 passerine
0.80 bird
0.85 bird
0.90 living thing
0.95 entity
0.99 entity
Right click on the image t

0.70 domestic animal, windowsill
0.80 domestic animal, windowsill
0.85 windowsill
0.90 windowsill
0.95 windowsill
0.99 entity

0.70 rearview mirror, vertebrate
0.80 animal, rearview mirror
0.85 animal, rearview mirror
0.90 rearview mirror
0.95 rearview mirror
0.99 entity
Right click on the image to compare with Google

0.70 sail
0.80 artifact
0.85 artifact
0.90 physical entity
0.95 entity
0.99 entity

0.70 structure, tramway
0.80 structure, tramway
0.85 structure, tramway
0.90 tramway
0.95 artifact
0.99 entity
Right click on the image to compa

0.70 coupe
0.80 car
0.85 car
0.90 car
0.95 physical entity
0.99 entity
Right click on the image to co

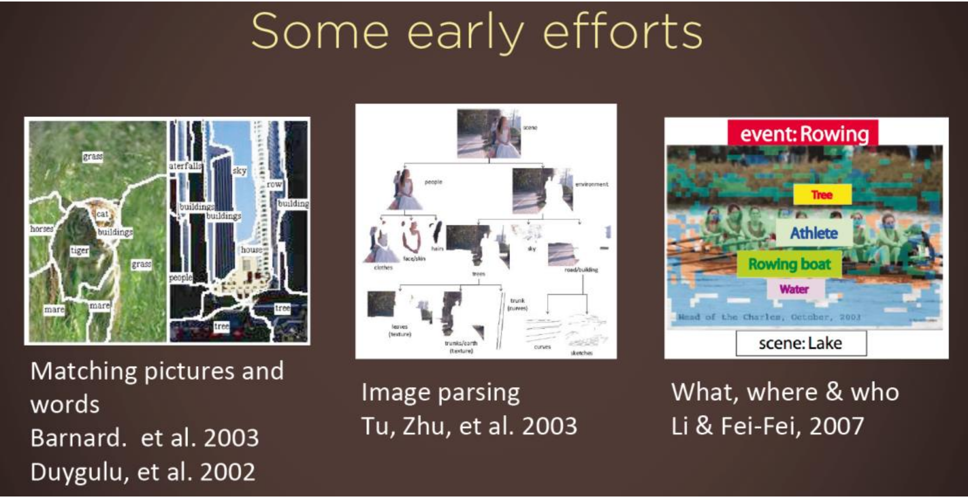
J. Deng, J. Krause, A. Berg, & L. Fei-Fei, *CVPR*, 2012

Fei-Fei Li

Goal: total scene understanding

Sub-goal 2: Gist of a scene
(aka Image Captioning)

Fei-Fei Li



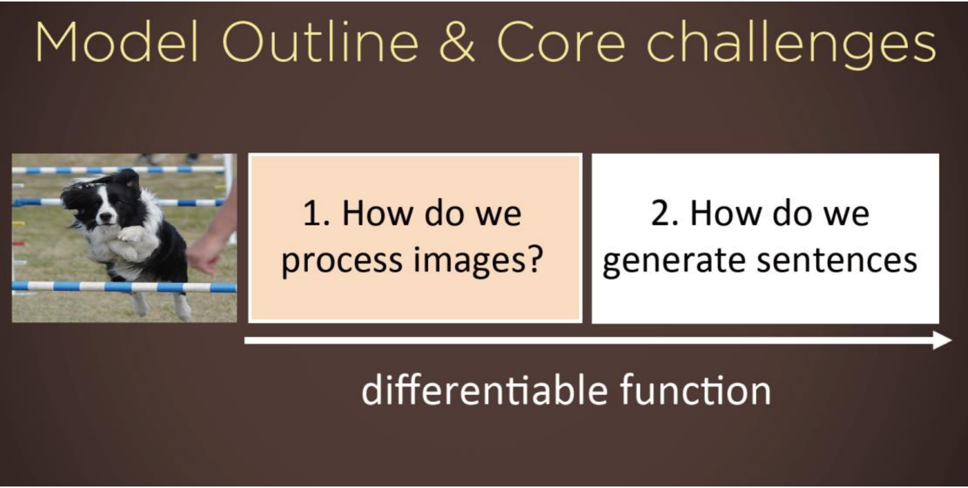
Some early efforts

Matching pictures and words
Barnard. et al. 2003
Duygulu, et al. 2002

Image parsing
Tu, Zhu, et al. 2003

What, where & who
Li & Fei-Fei, 2007

Fei-Fei Li



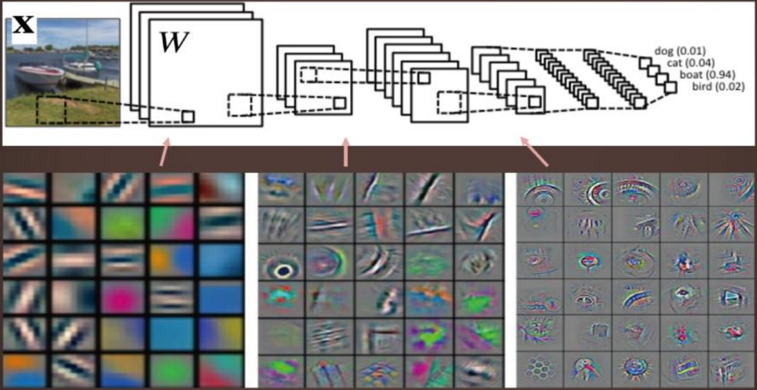
Model Outline & Core challenges

1. How do we process images?

2. How do we generate sentences

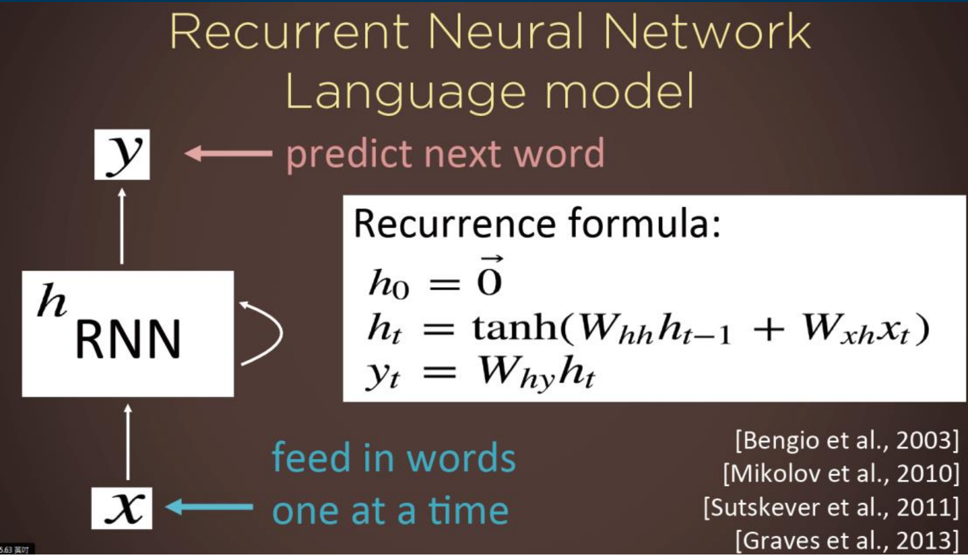
differentiable function

Convolutional Neural Network



[Zeiler & Fergus '13]

Fei-Fei Li



Recurrent Neural Network Language model

y ← predict next word

h RNN

x ← feed in words one at a time

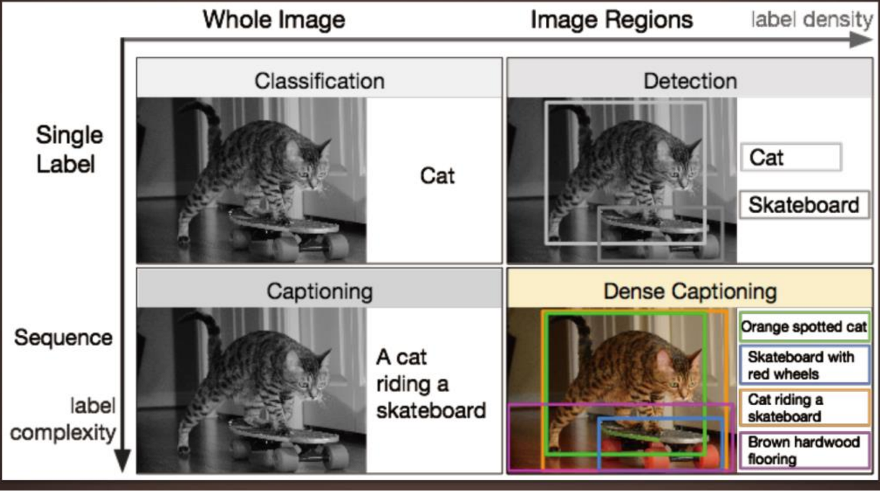
Recurrence formula:

$$h_0 = \vec{0}$$
$$h_t = \tanh(W_{hh}h_{t-1} + W_{xh}x_t)$$
$$y_t = W_{hy}h_t$$

[Bengio et al., 2003]
[Mikolov et al., 2010]
[Sutskever et al., 2011]
[Graves et al., 2013]

Fei-Fei Li

Dense Captioning



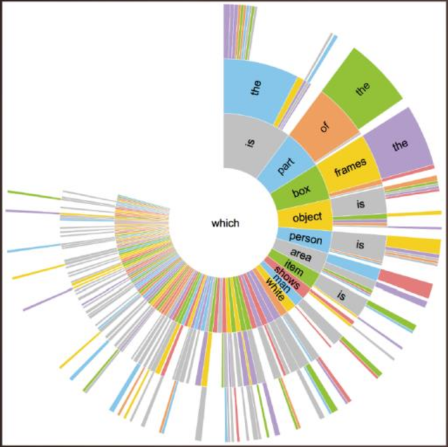
Fei-Fei Li

Finding regions given descriptions

"head of a giraffe"

Justin Johnson, Andrej Karpathy & Li Fei-Fei, *CVPR*, 2016

Fei-Fei Li



Q: Which item is used to cut items?

Q: Which doughnut has multicolored sprinkles?

Q: Which man is wearing the red tie?

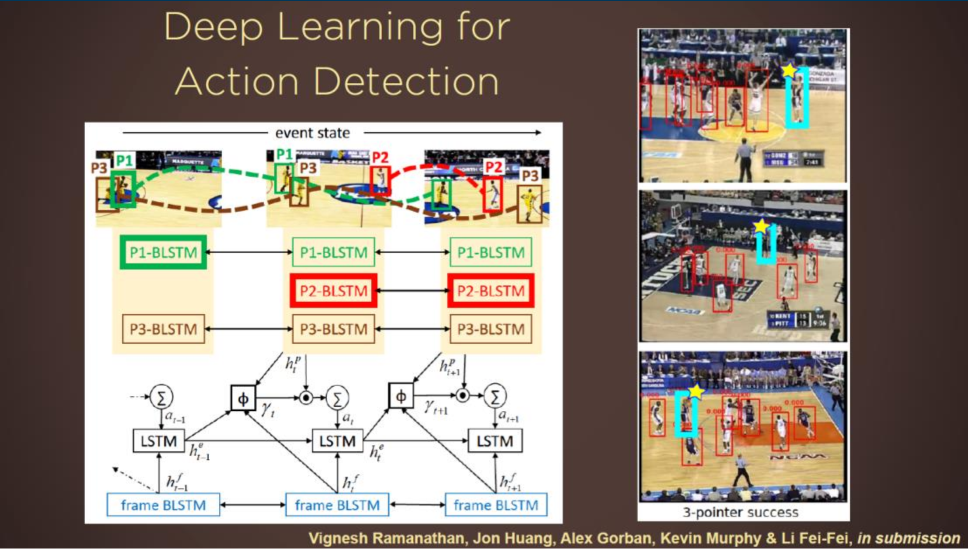
Q: Which pillow is farther from the window?

Q: Which step leads to the tub?

Q: Which is the small computer in the corner?

Yuke Zhu, Oliver Groth & Li Fei-Fei, *CVPR 2016*

Fei-Fei Li

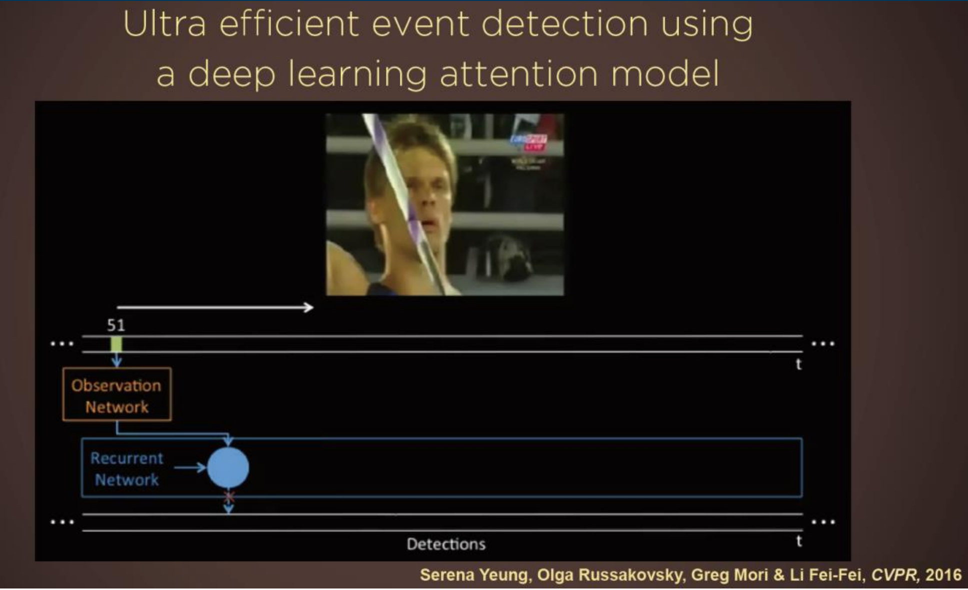


Deep Learning for Action Detection

3-pointer success

Vignesh Ramanathan, Jon Huang, Alex Gorban, Kevin Murphy & Li Fei-Fei, ***in submission***

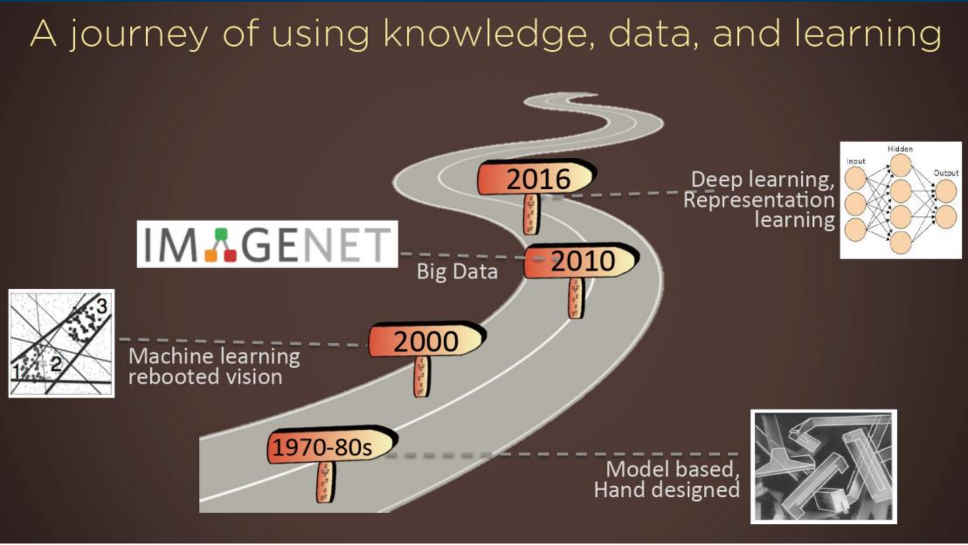
Fei-Fei Li



Ultra efficient event detection using a deep learning attention model

Serena Yeung, Olga Russakovsky, Greg Mori & Li Fei-Fei, *CVPR*, 2016

Fei-Fei Li



A journey of using knowledge, data, and learning

2016

Deep learning, Representation learning

2010

Big Data

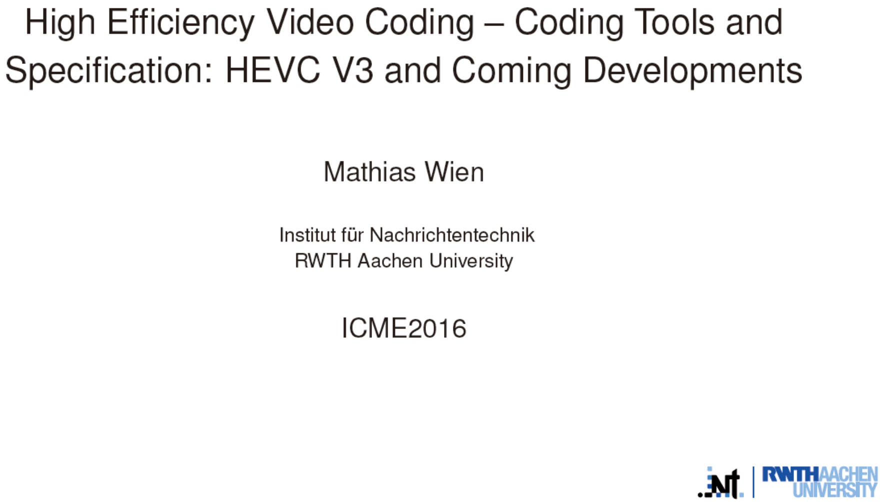
2000

Machine learning rebooted vision

1970-80s

Model based, Hand designed

Mathias Wien



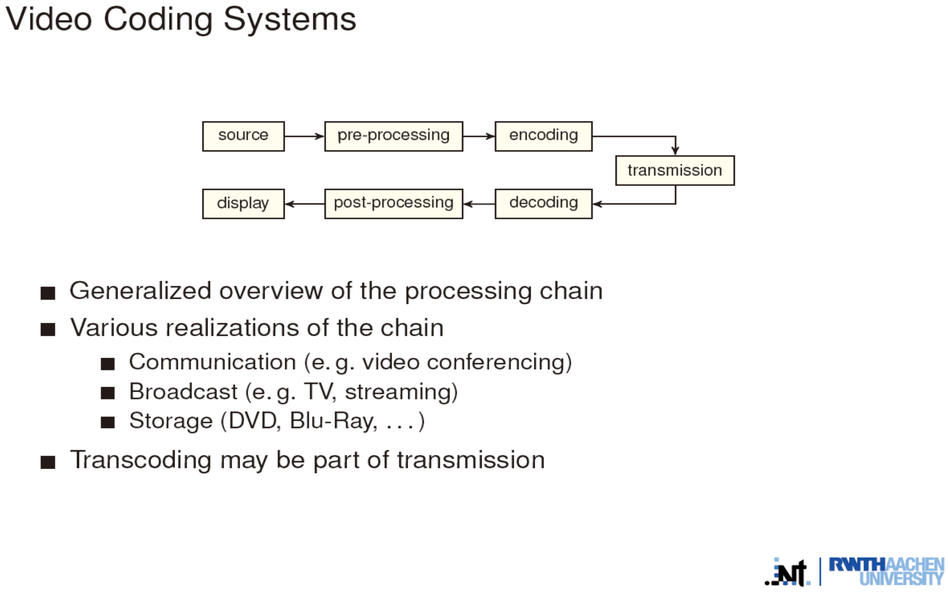
High Efficiency Video Coding – Coding Tools and Specification: HEVC V3 and Coming Developments

Mathias Wien

Institut für Nachrichtentechnik
RWTH Aachen University

ICME2016

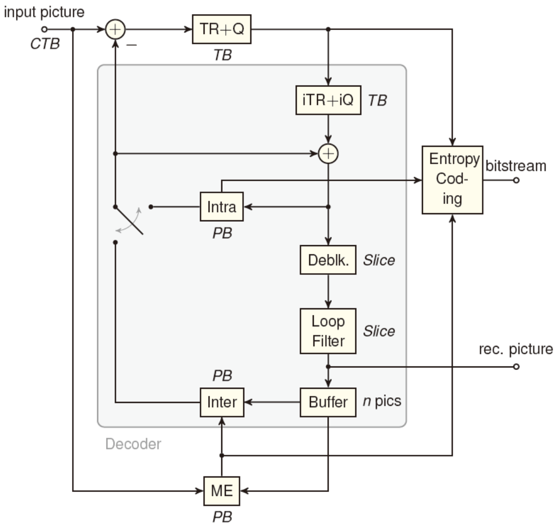
Mathias Wien



Video Coding Systems

- Generalized overview of the processing chain
- Various realizations of the chain
  - Communication (e. g. video conferencing)
  - Broadcast (e. g. TV, streaming)
  - Storage (DVD, Blu-Ray, . . .)
- Transcoding may be part of transmission

Hybrid Coding Scheme: Encoder

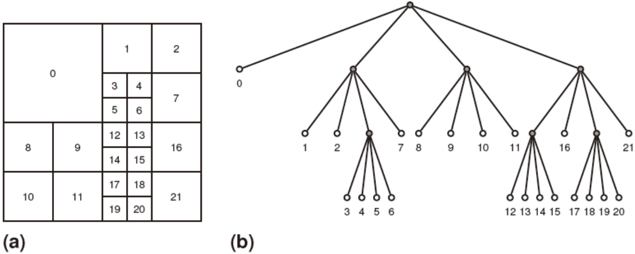


CTB	– Coding Tree Block
ME	– Motion Estimation
PB	– Prediction Block
Q	– Quantization
TB	– Transform Block
TR	– Transform

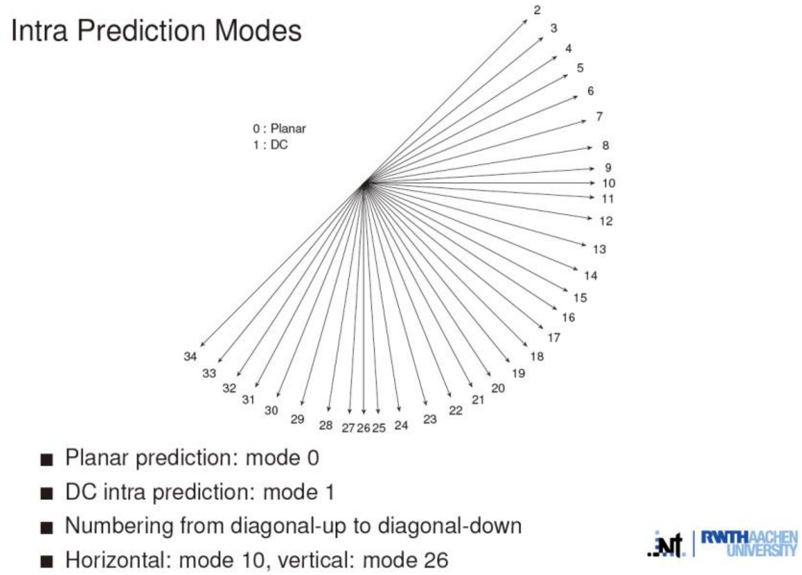
Mathias Wien

Coding Tree Blocks and Coding Blocks (CBs)

- Quadtree partitioning of CTB into CBs
- If picture size not integer multiple of CTB size:
  Implicit CTB partitioning to meet picture size (must be multiple of 8×8 pixels)



Mathias Wien

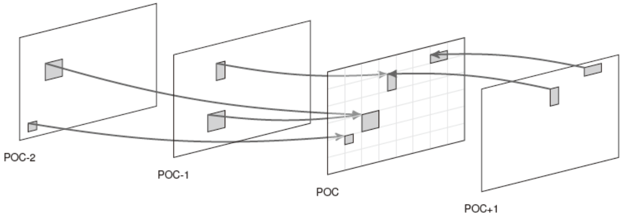


Intra Prediction Modes

- Planar prediction: mode 0
- DC intra prediction: mode 1
- Numbering from diagonal-up to diagonal-down
- Horizontal: mode 10, vertical: mode 26

Mathias Wien

Motion Compensated Prediction

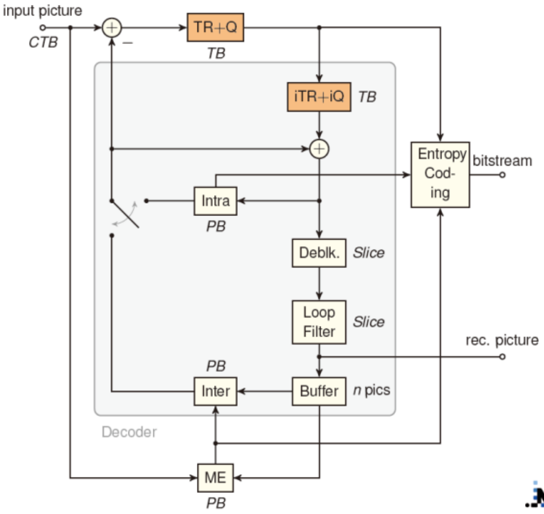


Prediction from reference picture lists

- Uni-prediction
  - P-slices only with List0, B-slices with List0 or List1
  - Minimum PB size 8 × 4 or 4 × 8
- Bi-prediction, only in B-slices
  - One predictor from List0, one predictor from List1
  - Minimum PB size 8 × 8

Mathias Wien

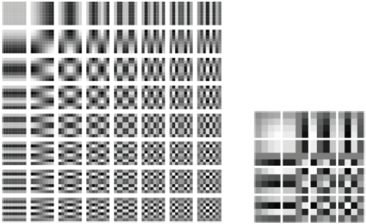
Residual Coding



Mathias Wien

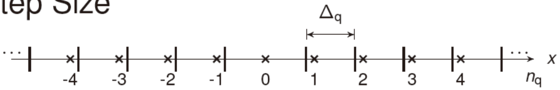
Core Transforms

- Transform block sizes 4×4, 8×8, 16×16, and 32×32
- Integer approximations of the DCT-II transform matrix
- Additionally, integer approximation of the DST-VI transform matrix
- 'Single-norm' design per transform block size → simple quantizer implementation
- Not all perfectly orthogonal, leakage below normalization threshold



Mathias Wien

Quantizer Step Size



- Quantizer step size Δ_q derived from quantization parameter QP
- Logarithmic relation of quantizer step sizes
- Double step size every 6 QP

$$\Delta_q(QP+1) = \sqrt[6]{2} \cdot \Delta_q(QP)$$

- Definition: $\Delta_q = 1$ for $QP = 4$, thereby

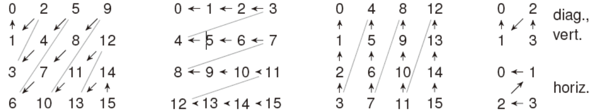
$$\Delta_{q,0} \in \left\{2^{-\frac{4}{6}}, 2^{-\frac{3}{6}}, 2^{-\frac{2}{6}}, 2^{-\frac{1}{6}}, 1, 2^{\frac{1}{6}}\right\}$$

- Quantizer step sizes for $QP > 5$

$$\Delta_q(QP) = \Delta_{q,0}(QP \bmod 6) \cdot 2^{\left\lfloor \frac{QP}{6} \right\rfloor}$$

Mathias Wien

Transform Sub-Block Scanning



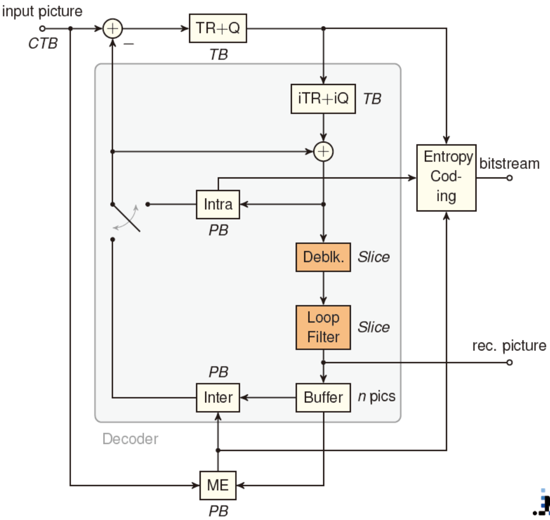
(a) 4×4 diagonal **(b)** 4×4 horizontal **(c)** 4×4 vertical **(d)** 8×8 TSB scans

- Partitioning of transform block into 4×4 transform sub-blocks (TSBs)
- Scan direction selectable for 4×4 and 8×8 blocks, diagonal otherwise
- Scan direction in TSB depending on (intra) prediction mode
- Level distribution: 'trailing ones' expected towards higher frequencies
  - Scan used in *inverse* direction
  - Start with expected '1' values

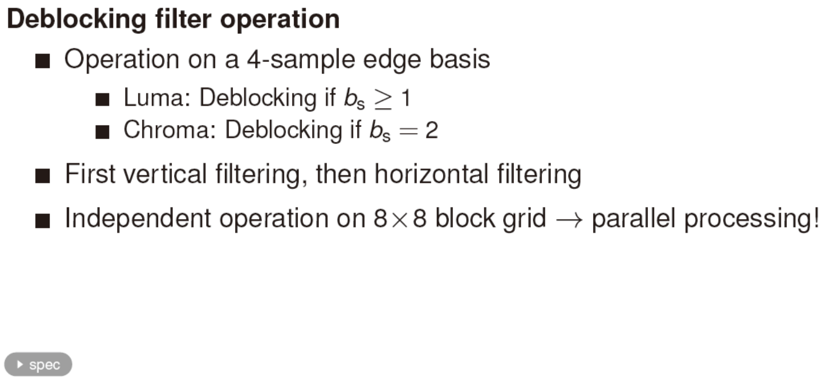
Level: quantized transform coefficient level n_q

Mathias Wien

Loop Filtering



Deblocking Filter Operation

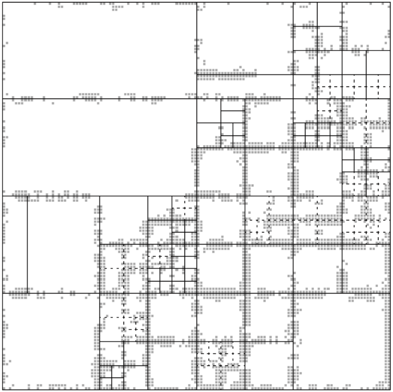


Deblocking filter operation

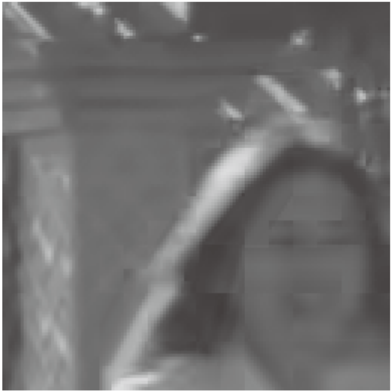
- Operation on a 4-sample edge basis
  - Luma: Deblocking if $b_s \geq 1$
  - Chroma: Deblocking if $b_s = 2$
- First vertical filtering, then horizontal filtering
- Independent operation on 8×8 block grid $\rightarrow$ parallel processing!

Mathias Wien

Deblocking Filter Example



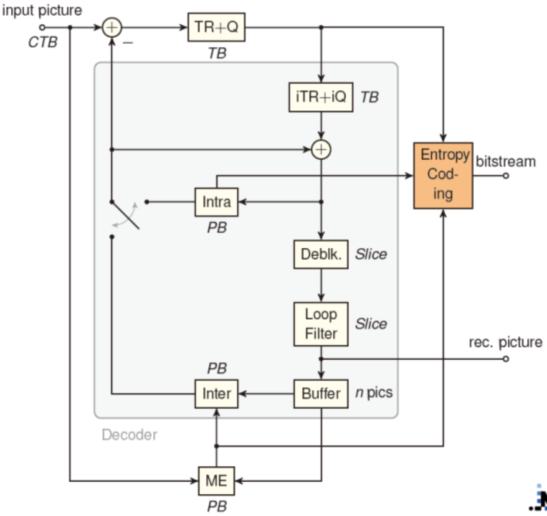
(c) Structure, deblocked samples



(d) Reconstruction without deblocking

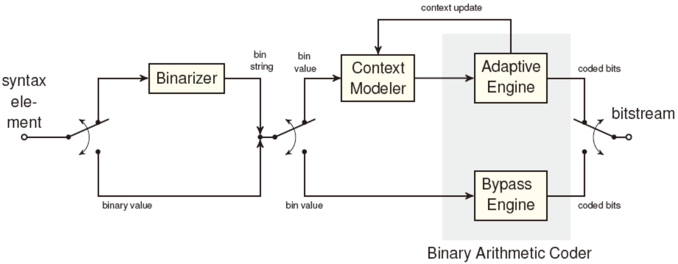
Mathias Wien

Entropy Coding



Mathias Wien

Context-Based Adaptive Binary Arithmetic Coding – CABAC



- Binarization
- Context model selection
- Binary arithmetic coding
- Optimized binarization design, reduced number of non-bypass bins compared to H.264 | AVC

Mathias Wien

Extensions of HEVC

Range extensions (HEVC V.2/Ed. 2, 10/2014)

V.x=ITU-Version, Ed.y=ISO/ICE Edition

- Extended color formats (4:2:2, 4:4:4)
- Extended bit depth

Scalable extensions(HEVC V.2/Ed. 2, 10/2014)

- Simple, multiloop approach, no modifications on tool-level
- Supports spatial, SNR scalability

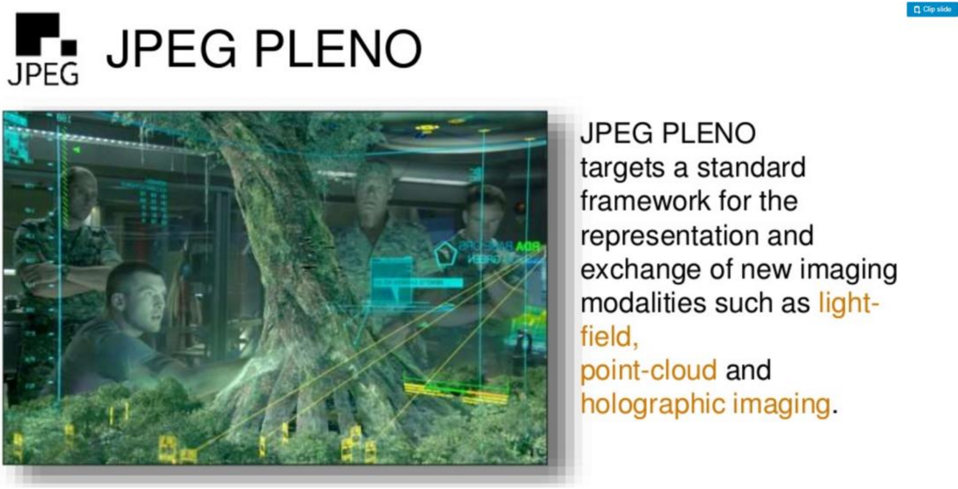
Multi-view (HEVC V.2/Ed. 2, 10/2014)
3D extensions (HEVC V.3/Ed. 3, 04/2015)

- Stereo / multi-view coding
- Multi-view with depth coding

Screen Content Coding (V.4/Ed. 3, 2016)

- Dedicated tools for this content type

Mathias Wien



JPEG PLENO

JPEG PLENO targets a standard framework for the representation and exchange of new imaging modalities such as light-field, point-cloud and holographic imaging.

Mathias Wien

23rd International Conference on Image Processing
September 25 – 28, 2016, Phoenix, Arizona

SUMMARY AND NEXT STEPS FOR THE IMAGE COMPRESSION CHALLENGE

Evangelos Alexiou , Irene Viola,
Lukas Krasula, Thomas Richter, Tim Bruylants, Antonio Pinheiro, Karel Fliegel,
Martin Rerabek, Athanassios Skodras, Peter Schelkens and Touradj Ebrahimi

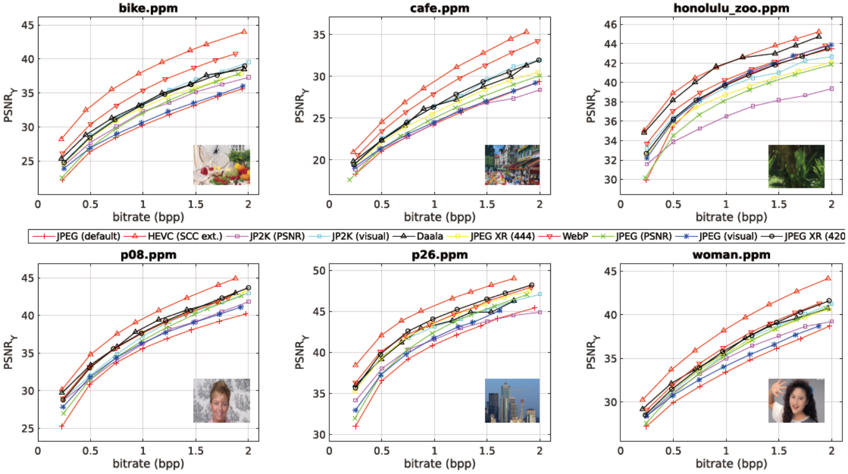
A contribution from Qualinet to JPEG call for information

23rd International Conference on Image Processing, September 25 – 28, 2016, Phoenix, Arizona

Mathias Wien

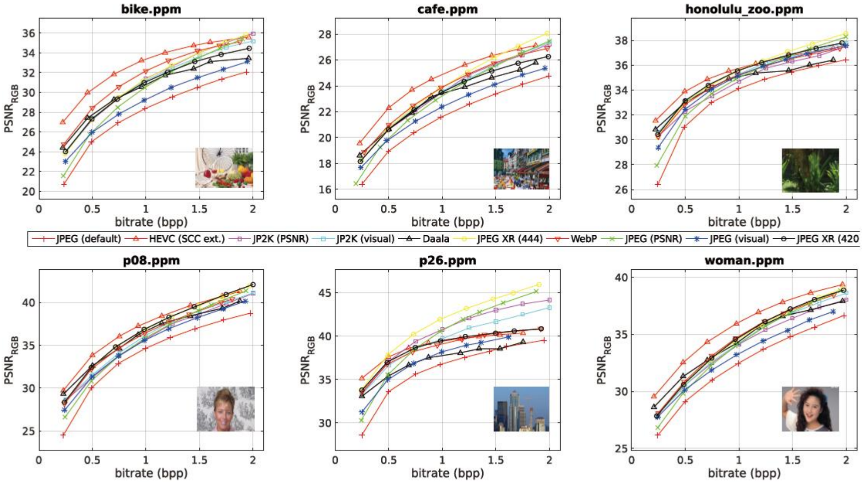
Objective evaluation: PSNR$_Y$ results

18

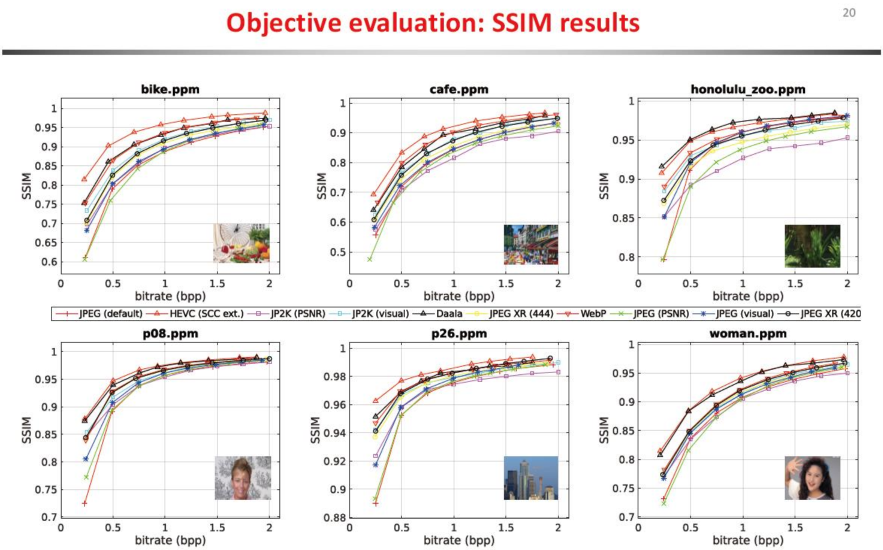


Mathias Wien

Objective evaluation: PSNR$_{RGB}$ results



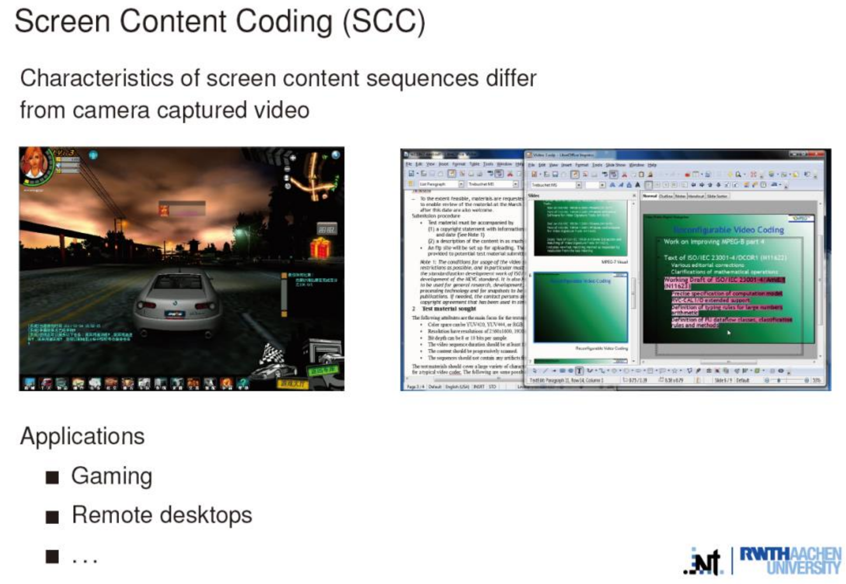
Mathias Wien



Objective evaluation: SSIM results

20

Mathias Wien



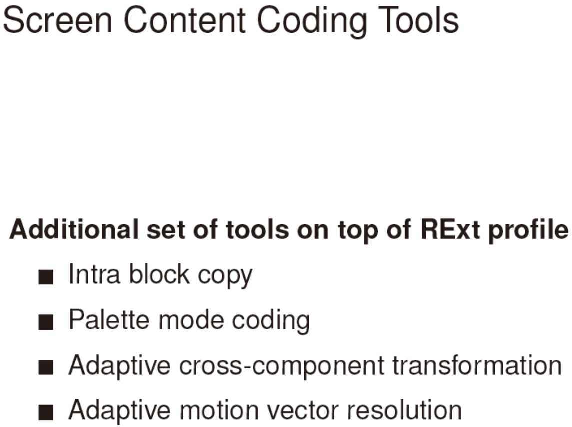
Screen Content Coding (SCC)

Characteristics of screen content sequences differ from camera captured video

Applications

- Gaming
- Remote desktops
- …

Mathias Wien



Screen Content Coding Tools

Additional set of tools on top of RExt profile

- Intra block copy
- Palette mode coding
- Adaptive cross-component transformation
- Adaptive motion vector resolution

Gene Cheung

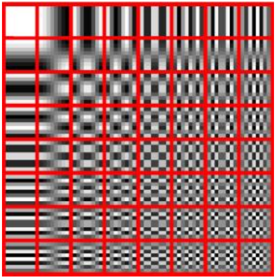
(1) Gene Cheung, Xian Ming Liu "Graph Signal Processing for Image Compression and Restoration," Nii.

Traditional signal property:

1. Discrete
2. Smooth
3. Band-limited frequency

Ex. DCT (Approx. of KLT)

$$X_k = \sum_{n=0}^{N-1} x_n \cos\left(\frac{\pi}{N}\left(n+\frac{1}{2}\right)k\right)$$



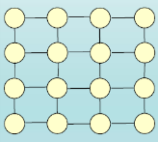
$$a = \Psi x$$

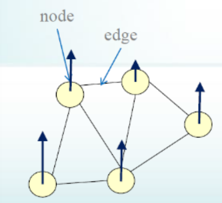
Dictionary representation inducing sparse coding

Gene Cheung

Graph Signal : Signal live on graph

Ex. Images: 2D – grid

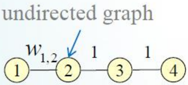




Research interest:

1. Sampling : how to efficiently acquire signal from graph
2. Representation : Given the graph signal, how to compactly represent it.
3. Signal Restoration : given the partial or noisy signal, how to recover the structure.

Graph Fourier Transform (GFT)



Graph Laplacian:

- *Adjacency Matrix A*: entry $A_{i,j}$ has *non-negative* edge weight $w_{i,j}$ connecting nodes *i* and *j*.

$$A = \begin{bmatrix} 0 & w_{1,2} & 0 & 0 \\ w_{1,2} & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

- *Degree Matrix D*: diagonal matrix w/ entry $D_{i,i}$ being sum of column entries in row *i* of A.

$$D_{i,i} = \sum_j A_{i,j}$$

$$D = \begin{bmatrix} w_{1,2} & 0 & 0 & 0 \\ 0 & w_{1,2}+1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

- *Combinatorial Graph Laplacian L*: $L = D\text{-}A$
  - *L* is *symmetric* (graph undirected).
  - *L* is a *high-pass* filter.
  - *L* is related to 2nd *derivative*.

$$L = \begin{bmatrix} w_{1,2} & -w_{1,2} & 0 & 0 \\ -w_{1,2} & w_{1,2}+1 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 1 \end{bmatrix}$$

$$L_{3,:}x = -x_2 + 2x_3 - x_4$$

$$f''(x) = \lim_{h \to 0} \frac{f(x+h) - 2f(x) + f(x-h)}{h^2}$$

Graph Fourier Transform (GFT)

- **Graph Fourier Transform** (GFT) is eigen-matrix of graph Laplacian *L*.

eigenvalue

$$L u_i = \lambda_i u_i$$

eigenvector

- Recall classical **Fourier Transform**: of function *f* is inner-product with *complex exponentials*:

inner-product of f and complex exp

$$\hat{f}(\xi) = \left\langle f, e^{2\pi i \xi t} \right\rangle = \int f(t) e^{-2\pi i \xi t} dt$$

  - Complex exponentials are ***eigen-functions*** of 1D Laplace operator:

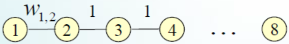
$$-\Delta\left(e^{2\pi i \xi t}\right) = \frac{\partial^2}{\partial t^2} e^{2\pi i \xi t} = (2\pi\xi)^2 e^{2\pi i \xi t}$$

- Analogously, GFT of graph-signal *f* is inner-product with ***eigenvectors*** of graph Laplacian L:

eigenvector of graph Laplacian

$$\hat{f}(\lambda_i) = \langle f, u_i \rangle = \sum_{n=1}^{N} f(n) u_i^*(n)$$

Graph Fourier Transform (GFT)

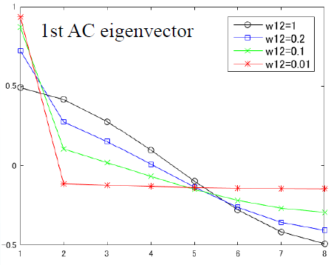


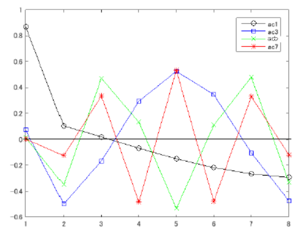
1. Sum of columns in L = **0** → constant eigenvector assoc. with $\lambda_0 = 0$
2. Edge weights affect shapes of eigenvectors.
3. Eigenvalues (≥ 0) as *graph frequencies*.
   - Constant eigenvector is DC component.
   - \# *zero-crossings* increases as λ increases.
4. GFT enables signal representation in graph frequency domain.

GFT

$$\alpha = \Psi x$$

5. "Smoothness", "band-limited" defined w.r.t. to graph frequencies.





Variants of Graph Laplacians

- **Graph Fourier Transform** (GFT) is eigen-matrix of graph Laplacian *L*.

eigenvalue

$$L u_i = \lambda_i u_i$$

eigenvector

- Other definitions of graph Laplacians:
  - **Normalized** graph Laplacian:

$$L_n = D^{-1/2} L D^{-1/2} = I - D^{-1/2} A D^{-1/2}$$

  - **Random walk** graph Laplacian:

$$L_{rw} = D^{-1} L = I - D^{-1} A$$

  - **Generalized** graph Laplacian [1]:

$$L_g = L + D^*$$

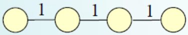
Characteristics:

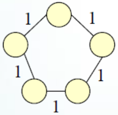
- Normalized.
- Symmetric.
- No DC component.

- Normalized.
- Asymmetric.
- Eigenvectors not orthog.

- Symmetric.
- **L** plus self loops.
- Defaults to DST, ADST.

[1] Wei Hu, Gene Cheung, Antonio Ortega, "**Intra-Prediction and Generalized Graph Fourier Transform for Image Coding**," *IEEE*





Facts of Graph Laplacian & GFT

- $\mathrm{x}^T \mathrm{L} \mathrm{x}$ (graph Laplacian quadratic form) [2]) is one measure of variation in signal → graph-signal smoothness prior.

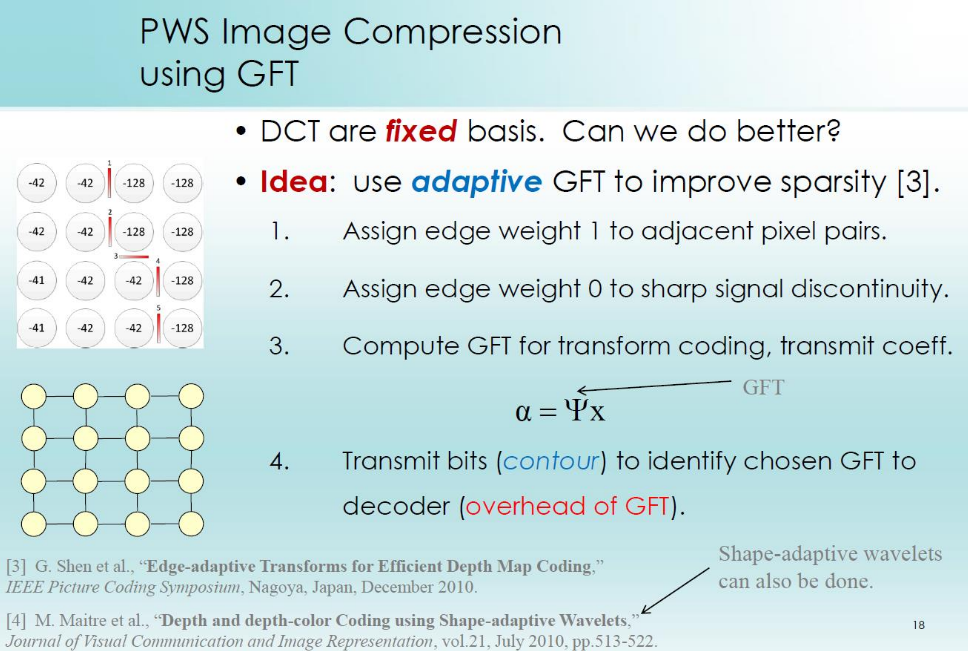
$$\mathrm{x}^T \mathrm{L} \mathrm{x} = \frac{1}{2} \sum_{i,j} w_{i,j} \left(x_i - x_j \right)^2 = \sum_i \lambda_i \alpha_i^2$$

- Eigenvalues can be defined iteratively via Rayleigh quotient (Courant-Fischer Theorem):

$$\lambda_0 = \min_{f \in \mathrm{R}^N, \|f\|_2 = 1} \left\{ f^T L f \right\}$$

$$\lambda_n = \min_{f \in \mathrm{R}^N, \|f\|_2 = 1, \, f \perp span\{u_0, \ldots, u_{n-1}\}} \left\{ f^T L f \right\} \quad n = 1, 2, \ldots, N-1$$

- GFT defaults to *DCT* for un-weighted connected line.
- GFT defaults to *DFT* for un-weighted connected circle.



PWS Image Compression using GFT

- DCT are ***fixed*** basis. Can we do better?
- **Idea**: use ***adaptive*** GFT to improve sparsity [3].
  1. Assign edge weight 1 to adjacent pixel pairs.
  2. Assign edge weight 0 to sharp signal discontinuity.
  3. Compute GFT for transform coding, transmit coeff.

$$\alpha = \Psi x$$

  4. Transmit bits (*contour*) to identify chosen GFT to decoder (overhead of GFT).

Shape-adaptive wavelets can also be done.

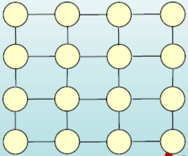
[3] G. Shen et al., "**Edge-adaptive Transforms for Efficient Depth Map Coding**," *IEEE Picture Coding Symposium*, Nagoya, Japan, December 2010.

[4] M. Maitre et al., "**Depth and depth-color Coding using Shape-adaptive Wavelets**," *Journal of Visual Communication and Image Representation*, vol.21, July 2010, pp.513-522.

PWS Image Compression

Q: Why GFT leads to sparseness?

Ans 1: **Capture statistical structure of signal** in edge weights of graph.



- Adjacent pixel correlation 0 or 1 for ***piecewise smooth*** (PWS) signal.
- Can be shown GFT approximates KLT given ***Gaussian Random Markov Field*** *(GRMF)* model [5].

Ans 2: **Avoid filtering across sharp edges.**

- Low-freq GFT basis are PWS for PWS signals (discussed later).

a 4x4 block

$\rightarrow$ GFT $\alpha_1 = \begin{bmatrix} 237 & 0 & 0 & 0 \\ 163 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

DCT $\alpha_2 = \begin{bmatrix} 285 & -29 & -5 & -4 \\ 16 & 1 & -16 & -4 \\ -5 & 3 & 5 & -7 \\ -1 & -4 & 1 & 9 \end{bmatrix}$

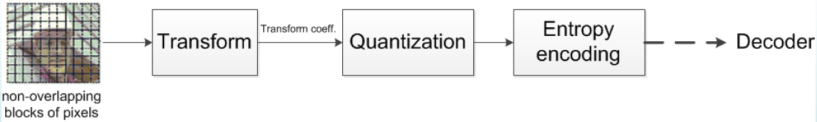
filtering operation

$$\alpha = \Psi x$$

[5] C. Zhang and D. Florencio, "**Analyzing the optimality of predictive transform coding using graph-based models**," *IEEE Signal Processing Letters*, vol. 20, NO. 1, January 2013, pp. 106–109.

Graph Fourier Transform (GFT) for Block-based Image Coding

- Block-based Transform coding of images*



non-overlapping blocks of pixels → Transform → Transform coeff. → Quantization → Entropy encoding → Decoder

Two things to transmit for adaptive transforms:

- transform coefficients → the cost of transform representation
- adaptive transform itself → the cost of transform description

- What's a *good* transform?
  - minimize the cost of transform representation & the cost of transform description

Transform Comparison

	Transform Representation	**Transform Description**
Karhunen-Loeve Transform (KLT)	"Sparsest" signal representation given available data / statistical model	Can be expensive (if poorly structured)
Discrete Cosine Transform (DCT)	*non-sparse signal representation* across sharp boundaries	little (fixed transform)
Graph Fourier Transform (GFT)	minimizes the total rate of signal's transform representation & transform description	

Search for Optimal GFT

- Rate-distortion performance: $D + \lambda R$

D: distortion R: bit-rate

- Assumption: high bit rate, uniform quantization

Distortion does not change when considering different transforms! [6]

Consider Rate only!

- For a given image block $\mathbf{x} \in \mathbb{R}^N$ under fixed uniform quantization at high rate, the optimal GFT is the one that minimizes the total rate:

$$\min_{\mathbf{W}} \quad R_\alpha(\mathbf{x}, \mathbf{W}) + R_T(\mathbf{W})$$

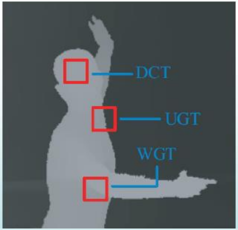
Rate of transform coefficient vector α

Rate of transform description T

MR-GFT: Definition of the Search Space for Graph Fourier Transforms

$$\min_{\mathbf{W}} \quad R_\alpha(\mathbf{x}, \mathbf{W}) + R_T(\mathbf{W})$$

- In general, weights could be any number in [0,1]
- To limit the description cost R_T
  - Restrict weights to a small discrete set $\mathcal{C} = \{1, 0, c\}$



- For ease of computation, divide the optimization into two sub-problems

1. Weighted GFT (WGFT): $\mathcal{C}_1 = \{1, c\}$
2. Unweighted GFT (UGFT): $\mathcal{C}_2 = \{1, 0\}$

Strong correlation only? Default to the DCT

WGFT

$$\min_{\mathbf{W}} \quad R_\alpha(\mathbf{x},\mathbf{W}) + R_T(\mathbf{W})$$

$$\text{s.t.} \quad W_{i,j} \in \{1, c\} \quad \forall\, i,j \in \mathcal{V}$$

- Cost function of transform coefficients

$$\hat{R}_\alpha(\mathrm{x},\mathrm{W}) \approx \mathrm{x}^T\mathrm{L}\mathrm{x} = \sum_{(i,j)\in \mathrm{E}} W_{i,j}\left(x_i - x_j\right)^2 = \sum_k \lambda_k \alpha_k^2$$

GFT coeff

graph freq.

- Cost function of transform description

$$\hat{R}_T(\mathbf{W}) = \sum_{(e,s)\in\mathcal{E}^d} |W_e - W_s| + \sum_{e\in\mathcal{V}^d} \gamma\rho(1 - W_e)$$

costly if many weight changes

code only non-1's

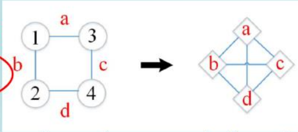
- Problem formulation for WGFT

deviation

separation

$$\min_{\mathbf{W}} \; \rho \sum_{e\in\mathcal{V}^d} \left[W_e\left(x_{v_1(e)} - x_{v_2(e)}\right)^2 + \gamma(1 - W_e)\right] + \sum_{(e,s)\in\mathcal{E}^d} |W_e - W_s|$$

$$\text{s.t.} \quad W_e \in \{1, c\} \quad \forall\, e \in \mathcal{V}^d.$$



orig graph

dual graph

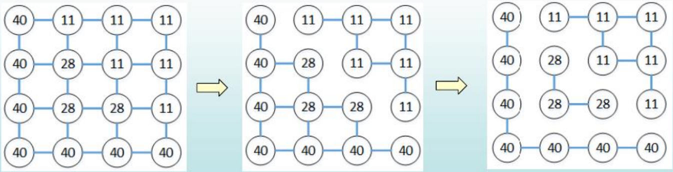
- ***Separation-Deviation*** (SD) problem, solvable in polynomial time [8].

[8] D. S. Hochbaum, "**An Efficient and Effective Tool for Image Segmentation, Total Variations and Regularization,**" *SSVM'11 Proceedings of the Third International Conference on Scale Space and Variational Methods in Computer Vision*, 2011, pp.338-349.

UGFT

$$\min_{\mathbf{W}} \quad R_\alpha(\mathbf{x}, \mathbf{W}) + R_T(\mathbf{W})$$
$$\text{s.t.} \quad W_{i,j} \in \{1, 0\} \quad \forall\, i, j \in \mathcal{V}$$

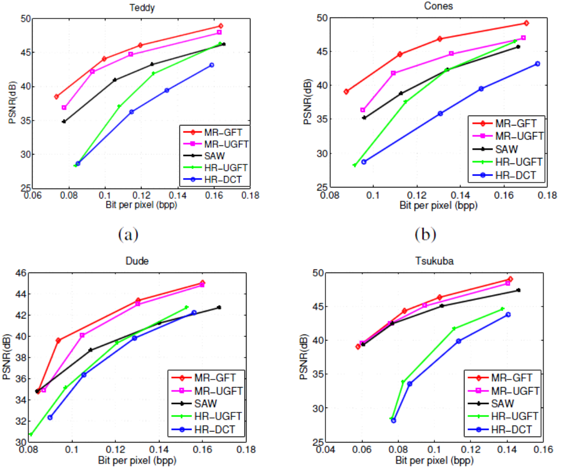
- A greedy algorithm



- Divide graph into disconnected sub-graphs via spectral clustering [9].
- Check objective function, further sub-divide if cost decreases.

[9] J. Shi and J. Malik, "**Normalized Cuts and Image Segmentation,**" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, August 2000.

Gene Cheung



(a)

(b)

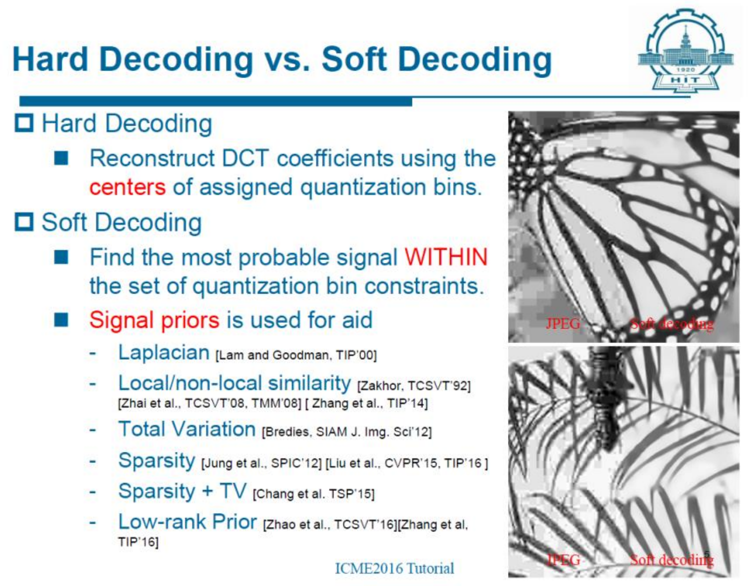
Wei Hu, Gene Cheung, Antonio Ortega, Oscar Au, "**Multiresolution Graph Fourier Transform for Compression of Piecewise Smooth Images**," *IEEE Transactions on Image Processing*, vol.24, no.1, pp.419-433, January 2015.

Gene Cheung

Skipped part

1. PWS image coding : Generalized GFT with intra-coding of H.264
2. Lifting implementation : lowering complexity from $O(N^2)$ to $O(N\log(N))$
3. Image denoising with sparsity and smoothness prior

Xianming Liu : Soft decoding of JPEG images



Hard Decoding vs. Soft Decoding

- ☐ Hard Decoding
  - Reconstruct DCT coefficients using the centers of assigned quantization bins.
- ☐ Soft Decoding
  - Find the most probable signal WITHIN the set of quantization bin constraints.
  - Signal priors is used for aid
    - Laplacian [Lam and Goodman, TIP'00]
    - Local/non-local similarity [Zakhor, TCSVT'92] [Zhai et al., TCSVT'08, TMM'08] [Zhang et al., TIP'14]
    - Total Variation [Bredies, SIAM J. Img. Sci'12]
    - Sparsity [Jung et al., SPIC'12] [Liu et al., CVPR'15, TIP'16]
    - Sparsity + TV [Chang et al. TSP'15]
    - Low-rank Prior [Zhao et al., TCSVT'16][Zhang et al, TIP'16]

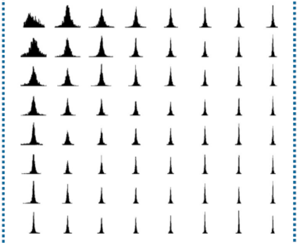
ICME2016 Tutorial

Laplacian Prior

- Q-bins: constrain the search space of individual DCT coefficients
- Laplacian Prior: states the probability density function of individual DCT coefficients

$$P_L(Y_i) = \frac{\mu_i}{2} \exp(-\mu_i \left|Y_i\right|)$$

[Lam and Goodman, TIP'00]



- MMSE Formulation

$$Y_i^* = \arg\min_{Y_i^o} \int_{q_iQ_i}^{(q_i+1)Q_i} (Y_i^o - Y_i)^2 \, P_L(Y_i) \; dY_i.$$

- Closed-form Solution

$$Y_i^* = \frac{(q_iQ_i + \mu_i)\, e^{\left\{\frac{-q_iQ_i}{\mu_i}\right\}} - ((q_i + 1)Q_i + \mu_i)\, e^{\left\{\frac{-(q_i+1)Q_i}{\mu_i}\right\}}}{e^{\left\{\frac{-q_iQ_i}{\mu_i}\right\}} - e^{\left\{-\frac{(q_i+1)Q_i}{\mu_i}\right\}}}$$

For higher frequencies, the Laplacian parameter is larger; i.e., the distribution is sharper and more skewed to 0.

Sparsity Prior

- Sparse Signal Model

$$\mathbf{x} = \underbrace{\boldsymbol{\Phi}}_{\text{over-complete dictionary}}\underbrace{\boldsymbol{\alpha}}_{\text{sparse code}} + \boldsymbol{\xi}$$

- Sparse Coding

$$\boldsymbol{\alpha}^* = \arg\min_{\boldsymbol{\alpha}} \|\mathbf{x} - \boldsymbol{\Phi}\boldsymbol{\alpha}\|_2^2 + \lambda \|\boldsymbol{\alpha}\|_0,$$

  - orthogonal matching pursuit (OMP) [Cai and Wang, TIT'11]
  - computational complexity is linear with the size of dictionary

- Sparsity Prior

$$P_S(\mathbf{x}) \propto \exp(-\lambda \|\boldsymbol{\alpha}\|_0).$$

Sparsity-based Soft Decoding

$$\min_{\{\mathbf{x},\boldsymbol{\alpha}\}} \|\mathbf{x}-\boldsymbol{\Phi}\boldsymbol{\alpha}\|_2^2 + \lambda \|\boldsymbol{\alpha}\|_0 ,$$
$$\text{s.t. } \mathbf{qQ} \preceq \mathbf{TMx} \prec (\mathbf{q}+1)\mathbf{Q}$$

- ☐ *Step 1–Initial Estimation:* The Laplacian prior is used to get an initial estimation of **x**.
- ☐ *Step 2–Sparse Decomposition:*

$$\boldsymbol{\alpha}^{(t)} = \arg\min_{\boldsymbol{\alpha}} \left\|\mathbf{x}^{(t)}-\boldsymbol{\Phi}\boldsymbol{\alpha}\right\|_2^2 + \lambda \|\boldsymbol{\alpha}\|_0 ,$$

- ☐ *Step 3–Quantization Constraint:*

$$\mathbf{x}^{(t+1)} = \arg\min_{\mathbf{x}} \left\|\mathbf{x}-\boldsymbol{\Phi}\boldsymbol{\alpha}^{(t)}\right\|_2^2 ,$$
$$\text{s.t. } \mathbf{qQ} \preceq \mathbf{TMx} \prec (\mathbf{q}+1)\mathbf{Q}$$

Lemma 1: The sparsity-based soft decoding algorithm converges to a local minimum.

Limitation of Small KSVD Dictionary

- Complexity linearly increases with the size of dictionary.
- In practice, a just reasonable over-complete dictionary is used.
- KSVD Dictionary Training

$$\min_{\boldsymbol{\Phi},\{\boldsymbol{\alpha}_i\}} \sum_{i=1}^{N} \|\mathbf{x}_i - \boldsymbol{\Phi}\boldsymbol{\alpha}_i\|_2^2 + \lambda\|\boldsymbol{\alpha}_i\|_0,$$

Training pixel patch

DCT patch $\mathbf{X}_i = \mathbf{T}'\mathbf{x}_i$

Parsavel's theorem

pre-set sparsity limit

$$\min_{\boldsymbol{\Phi},\{\boldsymbol{\alpha}_i\}} \sum_{i=1}^{N} \|\mathbf{X}_i - \mathbf{T}'\boldsymbol{\Phi}\boldsymbol{\alpha}_i\|_2^2, \quad \text{s.t.,} \quad \|\boldsymbol{\alpha}_i\|_0 \leq K$$

We analyze the behavior of dictionary learning in frequency domain

Limitation of Small KSVD Dictionary

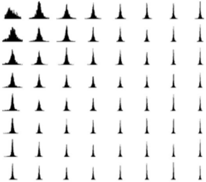
When K = 1, dictionary learning becomes vector quantization (VQ) design problem

- Selecting *M* atoms is analogous to designing *M* partitions

$$\mathbf{R} = \cup_{m=1}^{M} \mathbf{R}_m \qquad \mathbf{R}_i \cap \mathbf{R}_j = \emptyset, \forall i \neq j$$

- When *N* tends to infinite:

$$\min_{\{\phi_m\}} \sum_{m=1}^{M} \int_{\mathbf{R}_m} \underbrace{\|\mathbf{X} - \mathbf{T}'\phi_m\|_2^2}_{\text{Expected square error}} \underline{P(\mathbf{X})} d\mathbf{X}$$



a product of Laplacian distributions for individual DCT frequencies
- **low frequencies: decay slowly**
- **high frequencies: more skewed and concentrated around zero**

Limitation of Small KSVD Dictionary

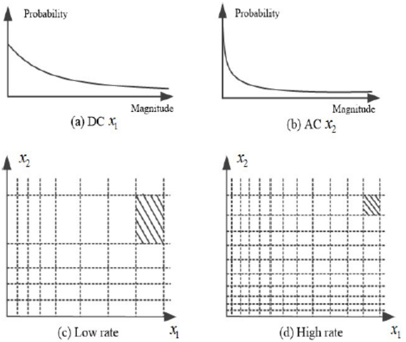


Illustration of product VQ for DC and AC frequencies

- ☐ When the number of atoms is small
  - quantization is coarser for large magnitude in AC than DC

When the dictionary Φ is small, the sparsity prior is difficult to recover large magnitude of high DCT frequencies.

- ☐ When the dictionary is large enough
  - quantization for large magnitude in high frequency is sufficiently fine.

When the dictionary Φ is large enough, the sparsity prior can recover large magnitude of high DCT frequencies well.

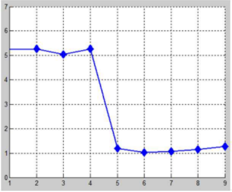
Graph-signal Smoothness Prior

- ☐ Graph Laplacian Regularizer

$$\mathbf{x}^T\mathbf{L}\mathbf{x} = \frac{1}{2}\sum_{(i,j)\in\mathcal{E}}(x_i - x_j)^2 W_{i,j} \quad\Longrightarrow\quad P_G(\mathbf{x}) \propto \exp\left(-\lambda_2 \mathbf{x}^T\mathbf{L}\mathbf{x}\right)$$

- ☐ Graph Frequency Interpretation
  - Eigen decomposition: $\mathbf{L} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T$
    - eigenvalues carry the notion of frequency
  - Graph Fourier transform: $\mathbf{F} = \mathbf{U}^T. \rightarrow \boldsymbol{\alpha} = \mathbf{F}\mathbf{x}$
  - We get

$$\mathbf{x}^T\mathbf{L}\mathbf{x} = \boldsymbol{\alpha}^T\boldsymbol{\Lambda}\boldsymbol{\alpha} = \sum_k \eta_k\,\alpha_k^2.$$



- ☐ Minimizing $\mathbf{x}^T\mathbf{L}\mathbf{x}$ will suppress high graph frequencies and preserve low graph frequencies.
  - **x** is smoothened with respect to the graph
  - PWS signals can be well approximated by low graph frequencies for appropriately constructed graphs. [Hu et al., MMSP'14, ICIP'14]

Interpretation from the Perspective of Spectral Clustering

Rayleigh quotient with respect to $\mathcal{L}_n$

$$\min_{\mathbf{v}} \boxed{\frac{\mathbf{v}^T \mathcal{L}_n \mathbf{v}}{\mathbf{v}^T \mathbf{v}}}, \quad \text{s.t.} \quad \mathbf{v}^T \mathbf{v}_1 = 0$$

- v is orthogonal to v_1, according to Rayleigh quotient, the solution is the second eigenvector of $\mathcal{L}_n$

Ln:Normalized Laplacian See p.56

The second eigenvector v_2 of L_n is a relaxed solution to the Ncut problem, which is PWS;
if the solution becomes exact, then v_2 is PWC.

- Matrix similarity transformation[1]

$$\mathcal{L}_r := \mathbf{D}^{-1/2} \mathcal{L}_n \mathbf{D}^{1/2} = \mathbf{D}^{-1} \mathbf{L}$$

Random walk graph Laplacian!

- $\mathcal{L}_r$ has the left eigenvectors $\mathbf{V}^T \mathbf{D}^{1/2}$

$$\mathbf{V}^T \mathbf{D}^{1/2} \mathcal{L}_r = \mathbf{\Lambda} \mathbf{V}^T \mathbf{D}^{1/2} \qquad \mathcal{L}_n = \mathbf{V} \mathbf{\Lambda} \mathbf{V}^T$$

- GFT using the left eigenvectors

$$\boldsymbol{\beta} = \mathbf{V}^T \mathbf{D}^{1/2} \mathbf{x}$$

Random Walk Graph Laplacian

- We use $\mathcal{L}_r^T\mathcal{L}_r$ instead, and can derive:

$$\begin{tikzcd}
\boxed{\mathbf{x}^T\mathcal{L}_r^T\mathcal{L}_r\mathbf{x} = (\mathbf{x}^T\mathbf{D}^{1/2}\mathcal{L}_n)\mathbf{D}^{-1}(\mathcal{L}_n\mathbf{D}^{1/2}\mathbf{x})} \arrow[d, "\boldsymbol{\gamma} = \mathcal{L}_n\mathbf{D}^{1/2}\mathbf{x}"] & \\
\boxed{\mathbf{x}^T\mathcal{L}_r^T\mathcal{L}_r\mathbf{x} = \boldsymbol{\gamma}^T\mathbf{D}^{-1}\boldsymbol{\gamma}} \arrow[d] & \\
\boxed{\frac{\boldsymbol{\gamma}^T\boldsymbol{\gamma}}{d_{\max}} \leq \boldsymbol{\gamma}^T\mathbf{D}^{-1}\boldsymbol{\gamma} \leq \frac{\boldsymbol{\gamma}^T\boldsymbol{\gamma}}{d_{\min}}} \arrow[r] & \boxed{(d_{\min}^{-1})\boldsymbol{\gamma}^T\boldsymbol{\gamma}}
\end{tikzcd}$$

Random Walk Graph Laplacian

$$\gamma^T\gamma = \mathbf{x}^T\mathbf{D}^{1/2}\mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^T\mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^T\mathbf{D}^{1/2}\mathbf{x}$$
$$= \boldsymbol{\beta}^T\boldsymbol{\Lambda}^2\boldsymbol{\beta} = \sum_k \tilde{\eta}_k^2\,\beta_k^2.$$

☐ We have a graph frequency interpretation of our Left Eigenvector Random-walk Graph Laplacian (LERaG) $(d_{\min}^{-1})\gamma^T\gamma$:

high frequencies of random walk graph Laplacian are suppressed to restore smooth signal x

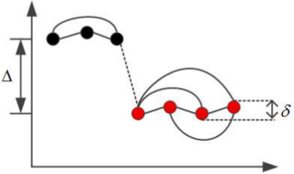
☐ The proposed regularizer can be efficiently computed as:

$$\boxed{(d_{\min}^{-1})\gamma^T\gamma = \mathbf{x}^T(d_{\min}^{-1})\mathbf{L}\mathbf{D}^{-1}\mathbf{L}\mathbf{x}}$$

Only adjacency matrix is involved, no need to compute other matrix

Analysis of Piecewise Smooth Signals

- 1D piecewise smooth (PWS) signal:
- A full-connected graph is built



- The normalized graph Laplacian $\mathcal{L}_n$ is still block-diagonal
- The second eigenvector $\mathbf{v}_2$

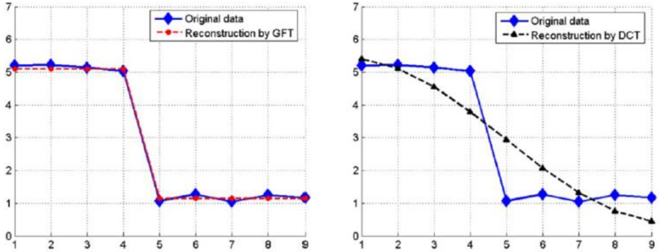
1st eigenvector for e-value = 0: $\mathbf{v}_1 = \mathbf{D}^{1/2}\mathbf{1}$

$$v_{2,i} = \begin{cases} \frac{D_{i,i}^{1/2}}{\sum_{j=1}^{l} D_{j,j}} & \text{if } 1 \le i \le l \\ -\frac{D_{i,i}^{1/2}}{\sum_{j=l+1}^{n} D_{j,j}} & \text{if } 1 < i \le n \end{cases}$$

⇨ **Roughly PWS**

- $\mathbf{D}^{1/2}\mathbf{x}$ is also roughly PWS: $\mathbf{D}^{1/2}\mathbf{x} \approx a_1\mathbf{v}_1 + a_2\mathbf{v}_2$
- There is a small penalty of LERaG.

Analysis of Ideal Piecewise Smooth Signals



(a)

(b)

Soft Decoding via Priors Mixture

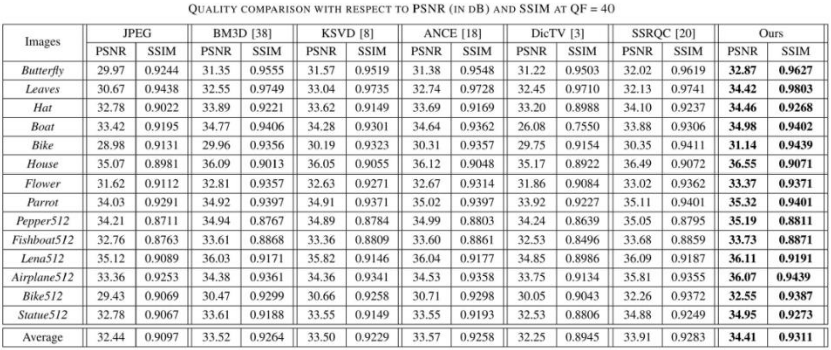
- The objective function

$$\arg\min_{\{\mathbf{x},\boldsymbol{\alpha}\}} \|\mathbf{x}-\boldsymbol{\Phi}\boldsymbol{\alpha}\|_2^2 + \lambda_1\|\boldsymbol{\alpha}\|_0 + \lambda_2\mathbf{x}^T(d_{\min}^{-1})\mathbf{L}\mathbf{D}^{-1}\mathbf{L}\mathbf{x},$$

$$\text{s.t. } \mathbf{q}\mathbf{Q} \preceq \mathbf{T}\mathbf{M}\mathbf{x} \prec (\mathbf{q}+1)\mathbf{Q}$$

  - λ1 is fixed
  - We adaptively increase λ2 if q-bin indices q indicate the presence of high DCT frequencies in target **x**.
- Optimization
  - Laplacian prior provides an initial estimation;
  - Fix **x** and estimate **α**;
  - Fix **α** and estimate **x**.

PSNR and SSIM Evaluation



QUALITY COMPARISON WITH RESPECT TO PSNR (IN DB) AND SSIM AT QF = 40

Images	JPEG		BM3D [38]		KSVD [8]		ANCE [18]		DicTV [3]		SSRQC [20]		Ours	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Butterfly	29.97	0.9244	31.35	0.9555	31.57	0.9519	31.38	0.9548	31.22	0.9503	32.02	0.9619	**32.87**	**0.9627**
Leaves	30.67	0.9438	32.55	0.9749	33.04	0.9735	32.74	0.9728	32.45	0.9710	32.13	0.9741	**34.42**	**0.9803**
Hat	32.78	0.9022	33.89	0.9221	33.62	0.9149	33.69	0.9169	33.20	0.8988	34.10	0.9237	**34.46**	**0.9268**
Boat	33.42	0.9195	34.77	0.9406	34.28	0.9301	34.64	0.9362	26.08	0.7550	33.88	0.9306	**34.98**	**0.9402**
Bike	28.98	0.9131	29.96	0.9356	30.19	0.9323	30.31	0.9357	29.75	0.9154	30.35	0.9411	**31.14**	**0.9439**
House	35.07	0.8981	36.09	0.9013	36.05	0.9055	36.12	0.9048	35.17	0.8922	36.49	0.9072	**36.55**	**0.9071**
Flower	31.62	0.9112	32.81	0.9357	32.63	0.9271	32.67	0.9314	31.86	0.9084	33.02	0.9362	**33.37**	**0.9371**
Parrot	34.03	0.9291	34.92	0.9397	34.91	0.9371	35.02	0.9397	33.92	0.9227	35.11	0.9401	**35.32**	**0.9401**
Pepper512	34.21	0.8711	34.94	0.8767	34.89	0.8784	34.99	0.8803	34.24	0.8639	35.05	0.8795	**35.19**	**0.8811**
Fishboat512	32.76	0.8763	33.61	0.8868	33.36	0.8809	33.60	0.8861	32.53	0.8496	33.68	0.8859	**33.73**	**0.8871**
Lena512	35.12	0.9089	36.03	0.9171	35.82	0.9146	36.04	0.9177	34.85	0.8986	36.09	0.9187	**36.11**	**0.9191**
Airplane512	33.36	0.9253	34.38	0.9361	34.36	0.9341	34.53	0.9358	33.75	0.9134	35.81	0.9355	**36.07**	**0.9439**
Bike512	29.43	0.9069	30.47	0.9299	30.66	0.9258	30.71	0.9298	30.05	0.9043	32.26	0.9372	**32.55**	**0.9387**
Statue512	32.78	0.9067	33.61	0.9188	33.55	0.9149	33.55	0.9193	32.53	0.8806	34.88	0.9249	**34.95**	**0.9273**
Average	32.44	0.9097	33.52	0.9264	33.50	0.9229	33.57	0.9258	32.25	0.8945	33.91	0.9283	**34.41**	**0.9311**