Gene Cheung, Xianming Liu

National Institute of Informatics

11th July, 2016

Graph Signal Processing for Image Compression & Restoration (Part I)

.

Biography

2D video Communication (12 years)

- MS from **UC Berkeley** in EECS in 1998.
  - *Thesis*: Joint source / channel coding for wireless video.
- PhD from **UC Berkeley** in EECS in 2000.
  - *Thesis*: Computation / memory / distortion tradeoff in signal compression.
- Senior researcher in **HP Labs Japan** from 2000 ~ 2009.
  - *Topic 1*: 2D video coding & streaming (2000~2007).
  - *Topic 2*: Multiview video, w/ Prof. Ortega (2007~).

invent

3D video Communication (8 yrs)

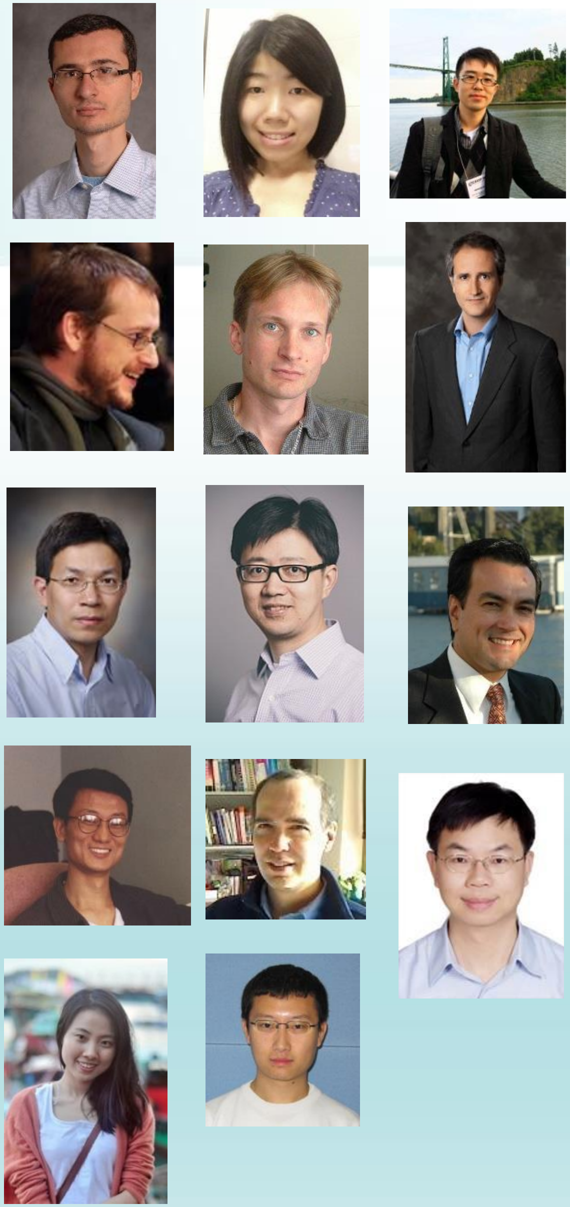
- Associated professor in **NII** from 11/2009 to now.
  - *Topic 1*: Image & video representation.
  - *Topic 2*: Immersive visual communication.
  - *Topic 3*: Graph signal processing.
- Adjunct associate professor in **HKUST** from 1/2015.

NII 大学共同利用機関法人 情報・システム研究機構
国立情報学研究所
National Institute of Informatics

Acknowledgement

Collaborators:

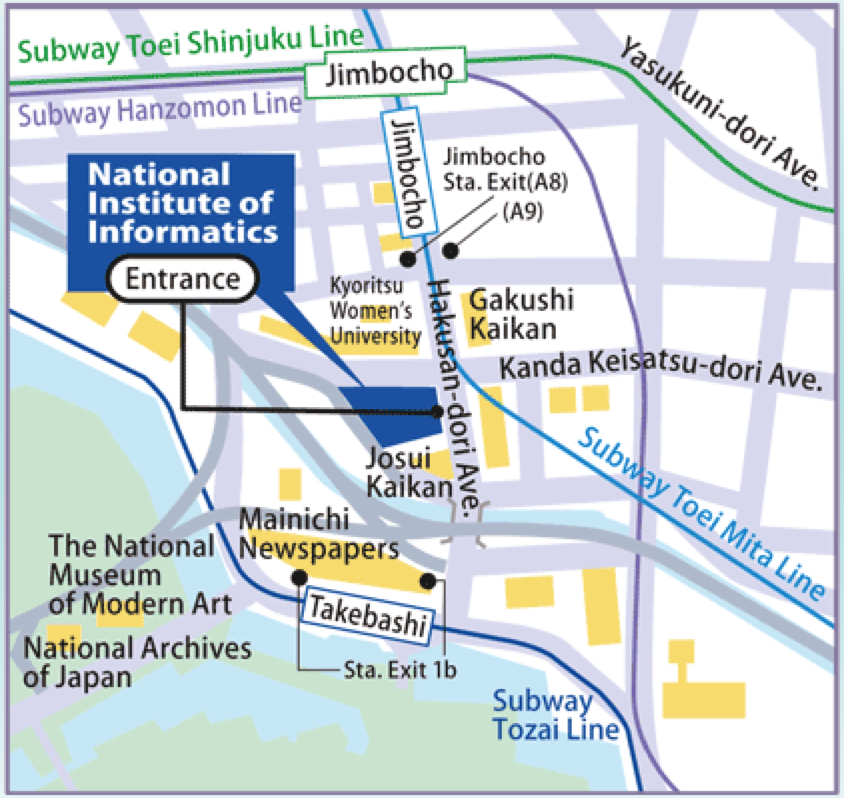
- B. Motz, Y. Mao, Y. Ji (NII, Japan)
- W. Hu, P. Wan, W. Dai, J. Pang, J. Zeng, A. Zheng, O. Au (HKUST, HK)
- Y.-H. Chao, A. Ortega (USC, USA)
- D. Florencio, C. Zhang, P. Chou (MSR, USA)
- Y. Gao, J. Liang (SFU, Canada)
- L. Toni, A. De Abreu, P. Frossard (EPFL, Switzerland)
- C. Yang, V. Stankovic (U of Strathclyde, UK)
- X. Wu (McMaster U, Canada)
- P. Le Callet (U of Nantes, France)
- H. Zheng, L. Fang (USTC, China)
- C.-W. Lin (National Tsing Hua University, Taiwan)



NII Overview

NII 大学共同利用機関法人 情報・システム研究機構
国立情報学研究所
National Institute of Informatics

- **National Institute of Informatics**
- Chiyoda-ku, Tokyo, Japan.
- Government-funded research lab.



- Offers graduate courses & degrees through **The Graduate University for Advanced Studies** (Sokendai).
- 60+ faculty in "**informatics**": quantum computing, discrete algorithms, database, machine learning, computer vision, speech & audio, image & video processing.

- **Get involved!**
  - 2-6 month Internships.
  - Short-term visits via MOU grant.
  - Lecture series, Sabbatical.

Outline (Part I)

- Fundamental of Graph Signal Processing (GSP)
  - Spectral Graph Theory
  - Graph Fourier Transform (GFT)
- Image Coding using GSP Tools
  - PWS Image Coding via Multi-resolution GFT
  - PWS Image Coding via Generalized GFT
  - Lifting Implementation of GFT

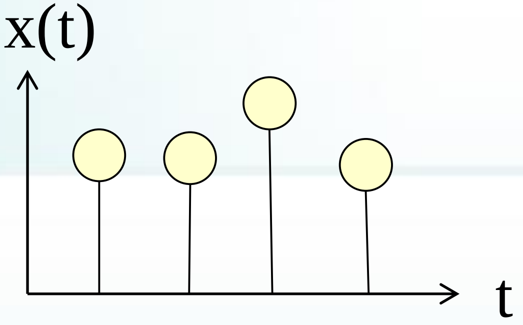
Outline (Part II)

- Image Restoration using GSP Tools
  - Image Denoising
  - Soft Decoding of JPEG Compressed Images
    - (Presented by Prof. Xianming Liu)

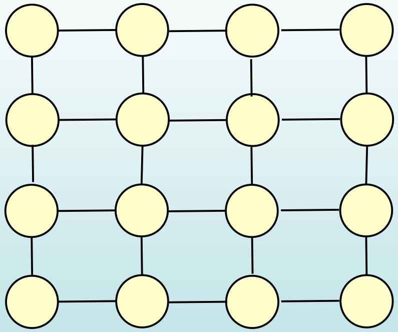
Outline (Part I)

- Fundamental of Graph Signal Processing (GSP)
  - Spectral Graph Theory
  - Graph Fourier Transform (GFT)
- Image Coding using GSP Tools
  - PWS Image Coding via Multi-resolution GFT
  - PWS Image Coding via Generalized GFT
  - Lifting Implementation of GFT

Traditional Signal Processing

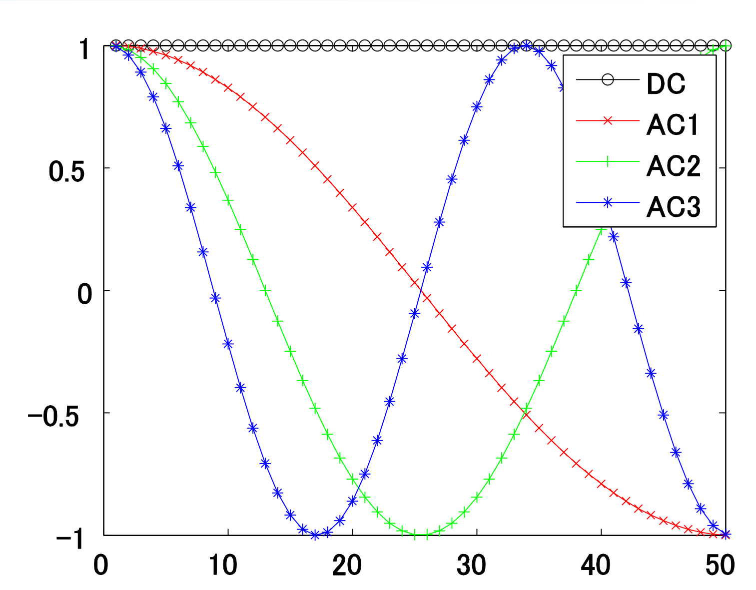


- Traditional discrete signals live on regular data kernels (**unstructured**).
  - Ex.1: audio / music / speech on regularly sampled timeline.
  - Ex.2: image on 2D grid.
  - Ex.3: video on 3D grid.



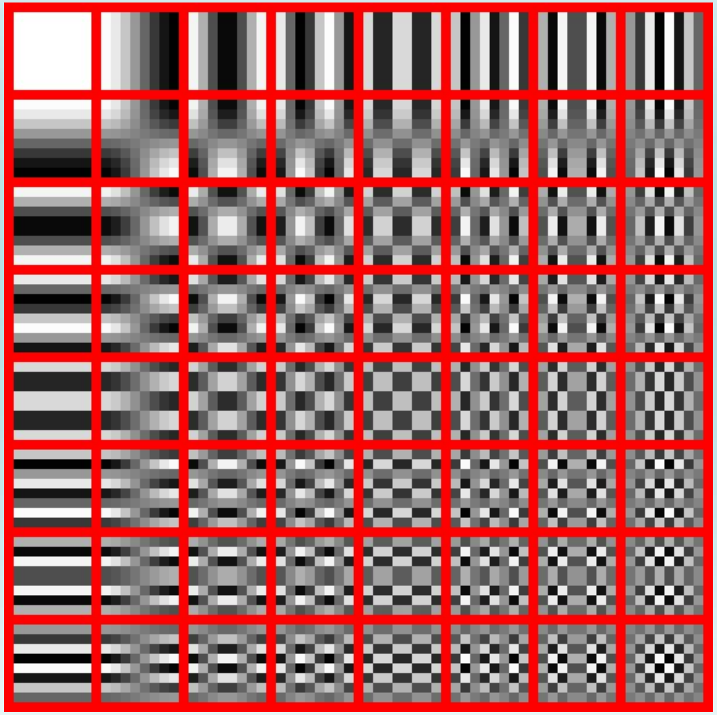
- Wealth of SP tools (transforms, wavelets, dictionaries, etc) for tasks such as:
  - compression, restoration, segmentation, classification.

Smoothness of Signals



- Many known signals are **smooth**.
- Notion of *frequency*, *band-limited*.
- Ex.1: DCT: $X_k = \sum_{n=0}^{N-1} x_n \cos\left(\frac{\pi}{N}\left(n+\frac{1}{2}\right)k\right)$

2D DCT basis is set of outer-product of 1D DCT basis in x- and y-dimension.



$$\mathrm{a} = \Phi\, \mathrm{x}$$

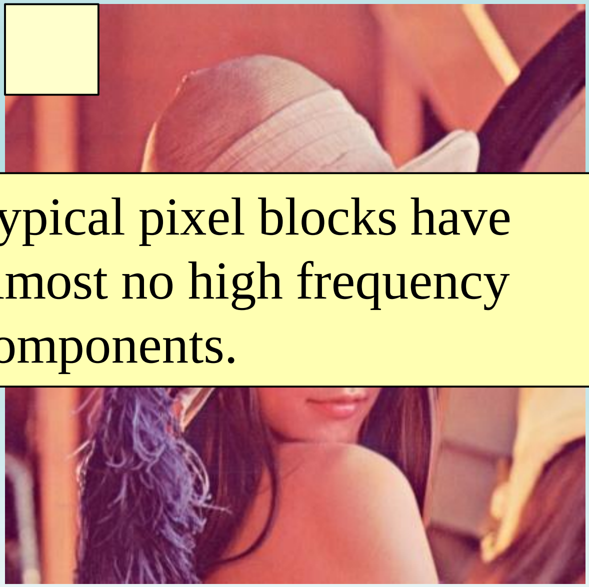
desired signal

transform

transform coeff.

$$\mathrm{a} = \begin{bmatrix} a_0 \\ a_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Compact signal representation



Typical pixel blocks have almost no high frequency components.

Sparsity of Signal Representations

- "Everything should be made as simple as possible, but no simpler." *paraphrase of Albert Einstein*
- "Among competing hypotheses, the hypothesis with the fewest assumptions should be selected (simplest explanation is usually the correct one)." *Occam's razor*

- Desirable signals are often **sparse**.

(over-complete) dictionary

desired signal

(sparse) code vector →

$$a = \Phi x$$

covariance matrix

$$E\left[xx^T\right] = C$$

- **KLT**: decorrelate input components.
  - Eigen-decomposition of covariance matrix.

eigen-matrix

diagonal matrix of eigen-values

$$CQ = Q\Lambda$$

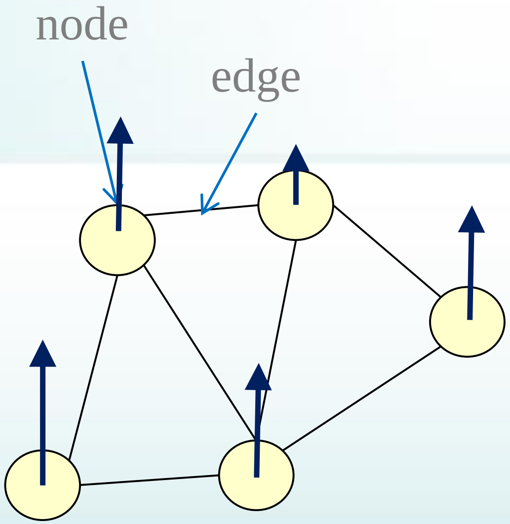
- **DCT** approximates KLT*.

Decorrelation leads to compact signal representation

---

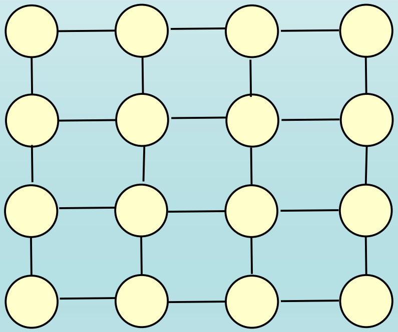
*Assuming simple AR process with same innovation variance.

Graph Signal Processing



- Signals live on graph.
  - Graph is a collection of nodes and edges.
  - Edges reveals *node-to-node relationships*.
- Data kernel itself is **structured**.

1. Data domain is naturally a graph.
   - Ex.1: posts on social networks.
   - Ex.2: temperatures on sensor networks.
2. **Embed signal structure in graph.**
   - Ex.1: images: 2D grid → structured graph.



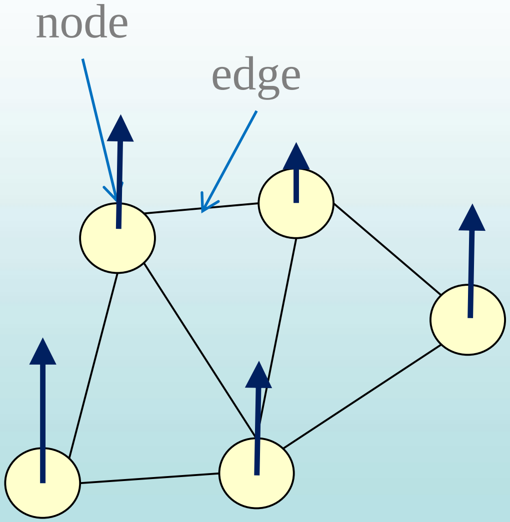
Graph Signal Processing (GSP) addresses the problem of processing signals that live on graphs.

[1] D. I. Shuman et al.,"**The Emerging Field of Signal Processing on Graphs: Extending High-dimensional Data Analysis to Networks and other Irregular Domains**," *IEEE Signal Processing Magazine*, vol.30, no.3, pp.83-98, 2013.

Graph Signal Processing

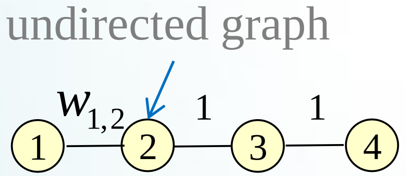
Research questions*:

- **Sampling**: how to efficiently acquire / sense a graph-signal?
  - Graph sampling theorems.
- **Representation**: Given graph-signal, how to compactly represent it?
  - Transforms, wavelets, dictionaries.
- **Signal restoration**: Given noisy and/or partial graph-signal, how to recover it?
  - Graph-signal priors.



* **Graph Signal Processing Workshop**, Philadelphia, US, May 25-27, 2016.
https://alliance.seas.upenn.edu/~gsp16/wiki/index.php?n=Main.Program

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*.

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

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

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

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

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

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

*https://en.wikipedia.org/wiki/Second_derivative

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:

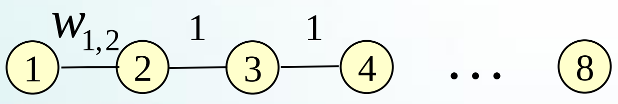
$$-\Delta\left(e^{2\pi i \xi t}\right) = \frac{\partial^2}{\partial t^2} e^{2\pi i \xi t} = (2\pi\xi)^2 e^{2\pi i \xi t}$$

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

eigenvector of graph Laplacian

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

Graph Fourier Transform (GFT)

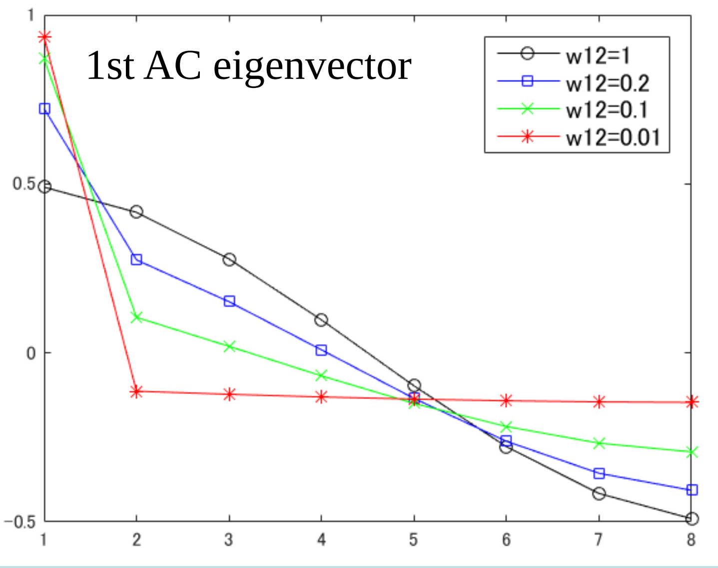


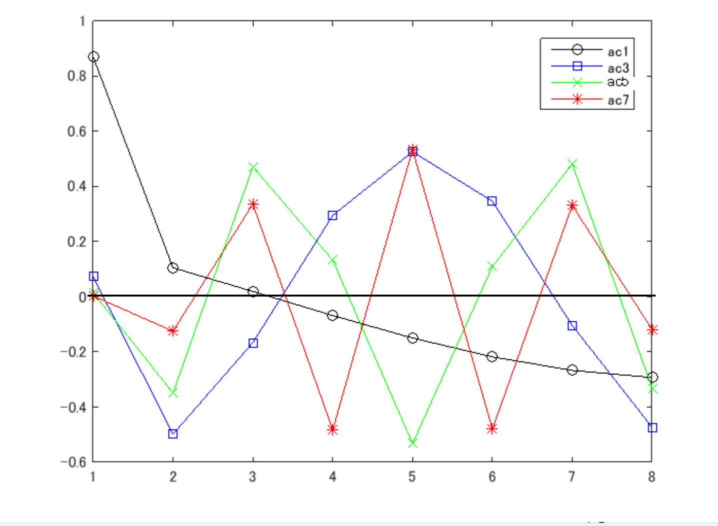
1. Sum of columns in L = $\mathbf{0}$ → constant eigenvector assoc. with $\lambda_0 = 0$
2. Edge weights affect shapes of eigenvectors.
3. Eigenvalues (≥ 0) as *graph frequencies*.
   - Constant eigenvector is DC component.
   - \# *zero-crossings* increases as λ increases.
4. GFT enables signal representation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

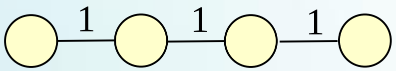
GFT

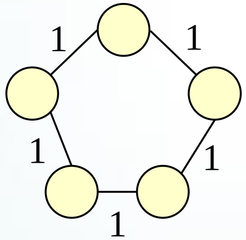
$$\alpha = \Psi \mathrm{x}$$

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









Facts of Graph Laplacian & GFT

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

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

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

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

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

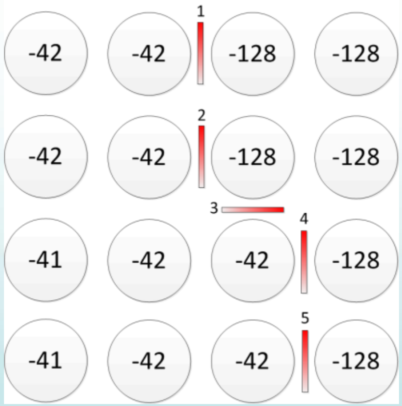
- GFT defaults to *DCT* for un-weighted connected line.
- GFT defaults to *DFT* for un-weighted connected circle.

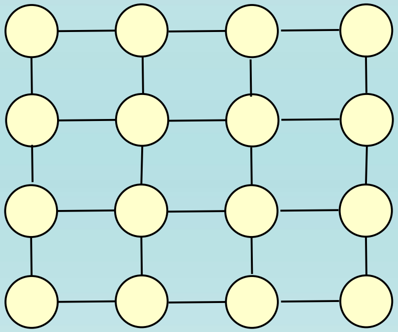
[2] D. Spielman, "**Spectral Graph Theory**," *Combinatorial Scientific Computing*, Chapman and Hall / CRC Press, 2012.

Outline (Part I)

- Fundamental of Graph Signal Processing (GSP)
  - Spectral Graph Theory
  - Graph Fourier Transform (GFT)
- Image Coding using GSP Tools
  - PWS Image Coding via Multi-resolution GFT
  - PWS Image Coding via Generalized GFT
  - Lifting Implementation of GFT

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.

$$\alpha = \Psi x$$

GFT

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

Shape-adaptive wavelets can also be done.

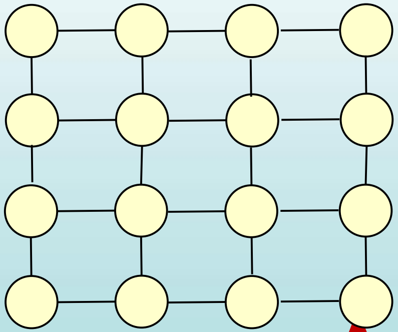
[3] G. Shen et al., "**Edge-adaptive Transforms for Efficient Depth Map Coding**," *IEEE Picture Coding Symposium*, Nagoya, Japan, December 2010.

[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?

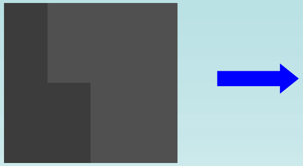
Ans 1: **Capture statistical structure of signal** in edge weights of graph.



- 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

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

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

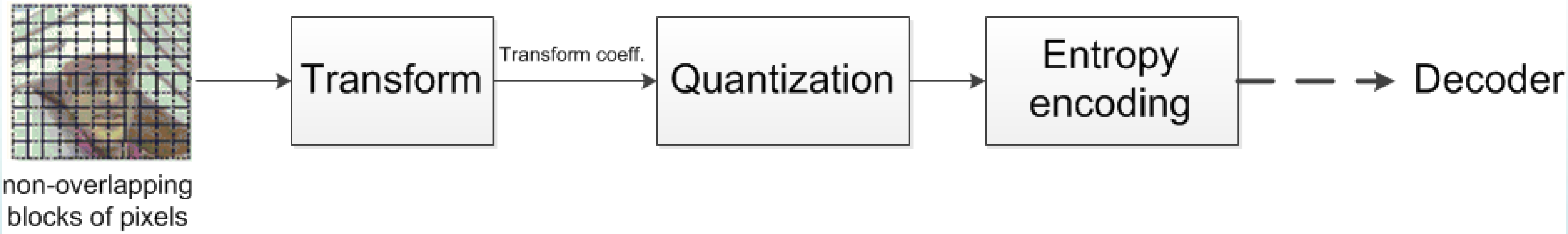
filtering operation

$$\alpha = \Psi x$$

[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*



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

[6] V. Goyal, "Theoretical foundations of transform coding," *in IEEE Signal Processing Magazine*, vol. 18, no.5, September 2001, pp. 9–21.

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	

GFT Comparison

HR-UGFT [3]	**MR-GFT**
unweighted graphs	unweighted & weighted graphs
no notion of optimality	define an optimality criterion
graphs are directly drawn from detected boundaries	propose efficient algorithms to search for the optimal GFT
Requires real-time eigen-decomposition	3 techniques to reduce computation complexity (multi-resolution, graph isomorphism, table lookup)

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

Search for Optimal GFT

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

- 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

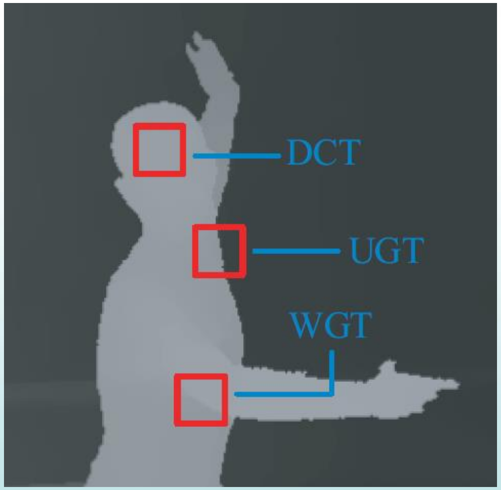
Rate of transform description T

[7] Gray and Neuhoff, "Quantization," *IEEE Transactions on Information Theory*, vol. 44, no. 6, Oct. 1998.

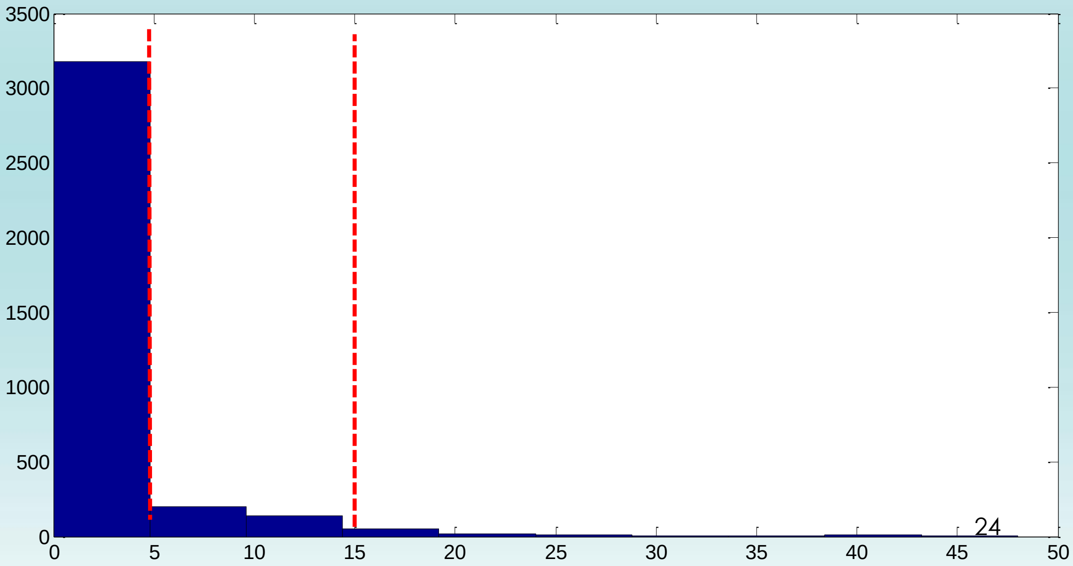
MR-GFT: Definition of the Search Space for Graph Fourier Transforms

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

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



- "1": *strong correlation* in smooth regions
- "0": *zero correlation* in sharp boundaries
- "c": *weak correlation* in slowly-varying parts



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

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

- 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

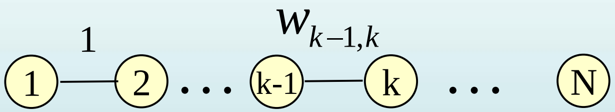
♦ What is the optimal c?

MR-GFT: Derivation of Optimal Edge Weights for Weak Correlation

- Assume a 1D first-order ***autoregressive (AR) process*** $\mathbf{x} = [x_1, ..., x_N]^T$ where,

$$x_k = \begin{cases} \eta, & k = 1 \\ x_{k-1} + e_k, & 1 < k \le N, \quad [k-1, k] \in \mathcal{S} \quad \longleftarrow \text{smooth} \\ x_{k-1} + g + e_k, & 1 < k \le N, \quad [k-1, k] \in \mathcal{P} \quad \longleftarrow \text{jump} \end{cases}$$

non-zero mean random var.



- Assuming the only weak correlation exists between x_{k-1} and x_k

$$\mathbf{F}\mathbf{x} = \mathbf{b},$$

$$\begin{aligned} x_1 &= \eta \\ x_2 - x_1 &= e_2 \\ &\cdots \\ x_k - x_{k-1} &= g + e_k \\ &\cdots \\ x_N - x_{N-1} &= e_N \end{aligned} \quad \Longrightarrow \quad \mathbf{F} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & \ddots & \ddots & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & \ddots & \ddots & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 0 \\ e_2 \\ \vdots \\ e_k \\ \vdots \\ e_N \end{bmatrix} + \begin{bmatrix} \eta \\ 0 \\ \vdots \\ g \\ \vdots \\ 0 \end{bmatrix}$$

$$\mu = \begin{bmatrix} 0 & \cdots & 0 & \underset{\text{k-th}}{m_g} & \cdots & m_g \end{bmatrix}^T$$

$$\boxed{\mathbf{x} = \mathbf{F}^{-1}\mathbf{b}}$$

MR-GFT: Derivation of Optimal Edge Weights for Weak Correlation (cont'd)

- Covariance matrix

$$\mathbf{x} = \mathbf{F}^{-1}\mathbf{b}$$

$$\begin{aligned}\mathbf{C} &= E[(\mathbf{x}-\mu)(\mathbf{x}-\mu)^T] \\ &= E[\mathbf{x}\mathbf{x}^T] - \mu\mu^T \\ &= E[\mathbf{F}^{-1}\mathbf{b}\mathbf{b}^T(\mathbf{F}^T)^{-1}] - \mu\mu^T \\ &= \mathbf{F}^{-1}E[\mathbf{b}\mathbf{b}^T](\mathbf{F}^T)^{-1} - \mu\mu^T\end{aligned}$$

$$\mu = \begin{bmatrix} 0 & \cdots & 0 & \overset{\text{k-th}}{m_g} & \cdots & m_g \end{bmatrix}^T$$

$$b = \begin{bmatrix} 0 \\ e_2 \\ \vdots \\ e_k \\ \vdots \\ e_N \end{bmatrix} + \begin{bmatrix} \eta \\ 0 \\ \vdots \\ g \\ \vdots \\ 0 \end{bmatrix}$$

$$\begin{array}{cl} E\left[\mathbf{b}\mathbf{b}^T\right] = \begin{bmatrix} \sigma_1^2 & 0 & & \cdots & 0 & \cdots & & 0 \\ 0 & 1 & & \cdots & 0 & \cdots & & 0 \\ & & \ddots & & & & & \\ 0 & 0 & \cdots & 1 & 0 & \cdots & & 0 \\ 0 & 0 & \cdots & 0 & \sigma_g^2 + m_g^2 + 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & 1 & \cdots & 0 \\ & & & & & & \ddots & \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 1 \end{bmatrix} & \begin{array}{l} \\ \\ \\ \\ \text{—— k-th row} \\ \\ \\ \\ \end{array} \end{array}$$

MR-GFT: Derivation of Optimal Edge Weights for Weak Correlation (cont'd)

- Precision matrix (tri-diagonal)

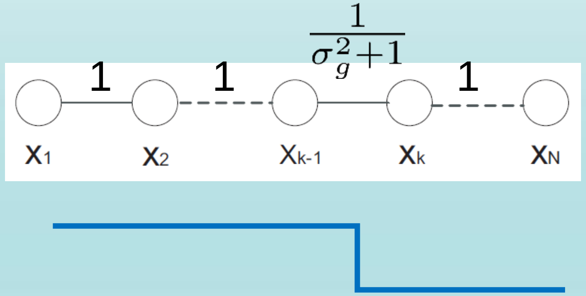
$$\mathbf{Q} = \mathbf{C}^{-1} = \begin{bmatrix}
1+\frac{1}{\sigma_1^2} & -1 & & & & & & & & \\
-1 & 2 & -1 & & & & & & & \\
& \ddots & \ddots & \ddots & & & & & & \\
& & -1 & 2 & -1 & & & & & \\
& & & -1 & \boxed{\frac{1}{\sigma_g^2+1}+1} & \overset{W_{k-1,k}}{\boxed{-\frac{1}{\sigma_g^2+1}}} & & & & \\
& & & & \underset{W_{k,k-1}}{\boxed{-\frac{1}{\sigma_g^2+1}}} & \boxed{\frac{1}{\sigma_g^2+1}+1} & -1 & & & \\
& & & & & -1 & 2 & -1 & & \\
& & & & & & \ddots & \ddots & \ddots & \\
& & & & & & & -1 & 2 & -1 \\
& & & & & & & & -1 & 1
\end{bmatrix}$$

(k-1)-th row

k-th row

$\approx \mathcal{L}$

Graph Laplacian matrix!



- the KLT basis = $\{\psi_l\}$ of the covariance matrix = $\{\psi_l\}$ of the precision matrix

$\simeq \{\psi'_l\}$ of the Laplacian matrix

$$\boxed{c = W_{k-1,k} = \frac{1}{\sigma_g^2+1}}$$

MR-GFT: Adaptive Selection of Graph Fourier Transforms

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

- Two sub-problems with two corresponding non-overlapping search spaces

1. Weighted GFT (WGFT): $\mathcal{C}_1 = \{1, c\}$ (weighted & connected graphs)
2. Unweighted GFT (UGFT): $\mathcal{C}_2 = \{1, 0\}$ (unweighted & disconnected graphs)

WGFT

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

- 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

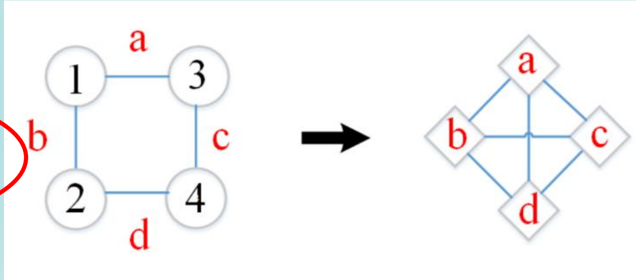
code only non-1’s

- Problem formulation for WGFT

deviation

separation

$$\begin{aligned}\min_{\mathbf{W}} \quad & \rho \sum_{e\in\mathcal{V}^d} \left[W_e \left(x_{v_1(e)} - x_{v_2(e)}\right)^2 + \gamma(1 - W_e)\right] + \sum_{(e,s)\in\mathcal{E}^d} |W_e - W_s| \\ \text{s.t.} \quad & W_e \in \{1, c\} \quad \forall\, e \in \mathcal{V}^d.\end{aligned}$$



orig graph

dual graph

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

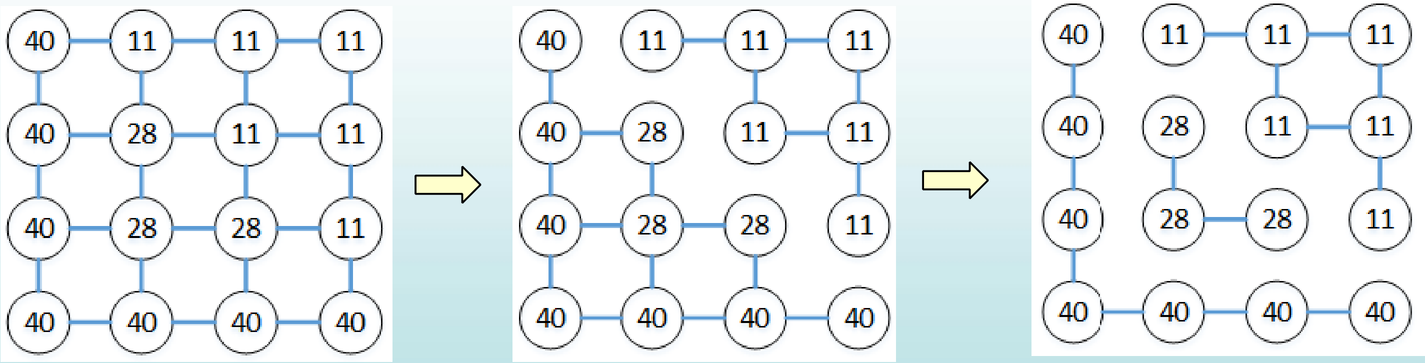
[8] D. S. Hochbaum, “**An Efficient and Effective Tool for Image Segmentation, Total Variations and Regularization,**” *SSVM’11 Proceedings of the Third International Conference on Scale Space and Variational Methods in Computer Vision*, 2011, pp.338-349.

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.

MR-GFT: Adaptive Selection of Graph Fourier Transforms

- Online eigen-decomposition: a hurdle to practical implementation

- Pre-compute and store GFTs in a table for simple lookups

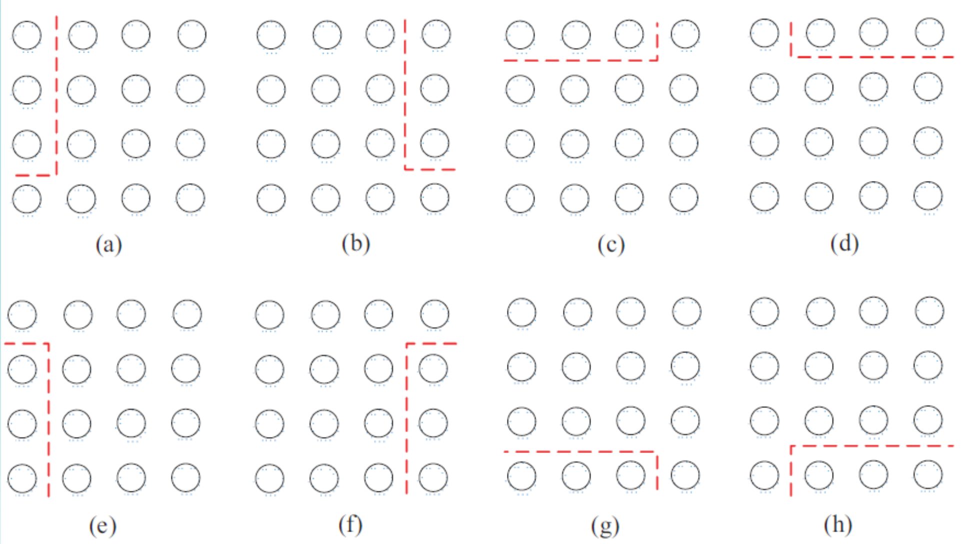
  - Perform GFT on a small block

  - Store the most frequently used GFTs

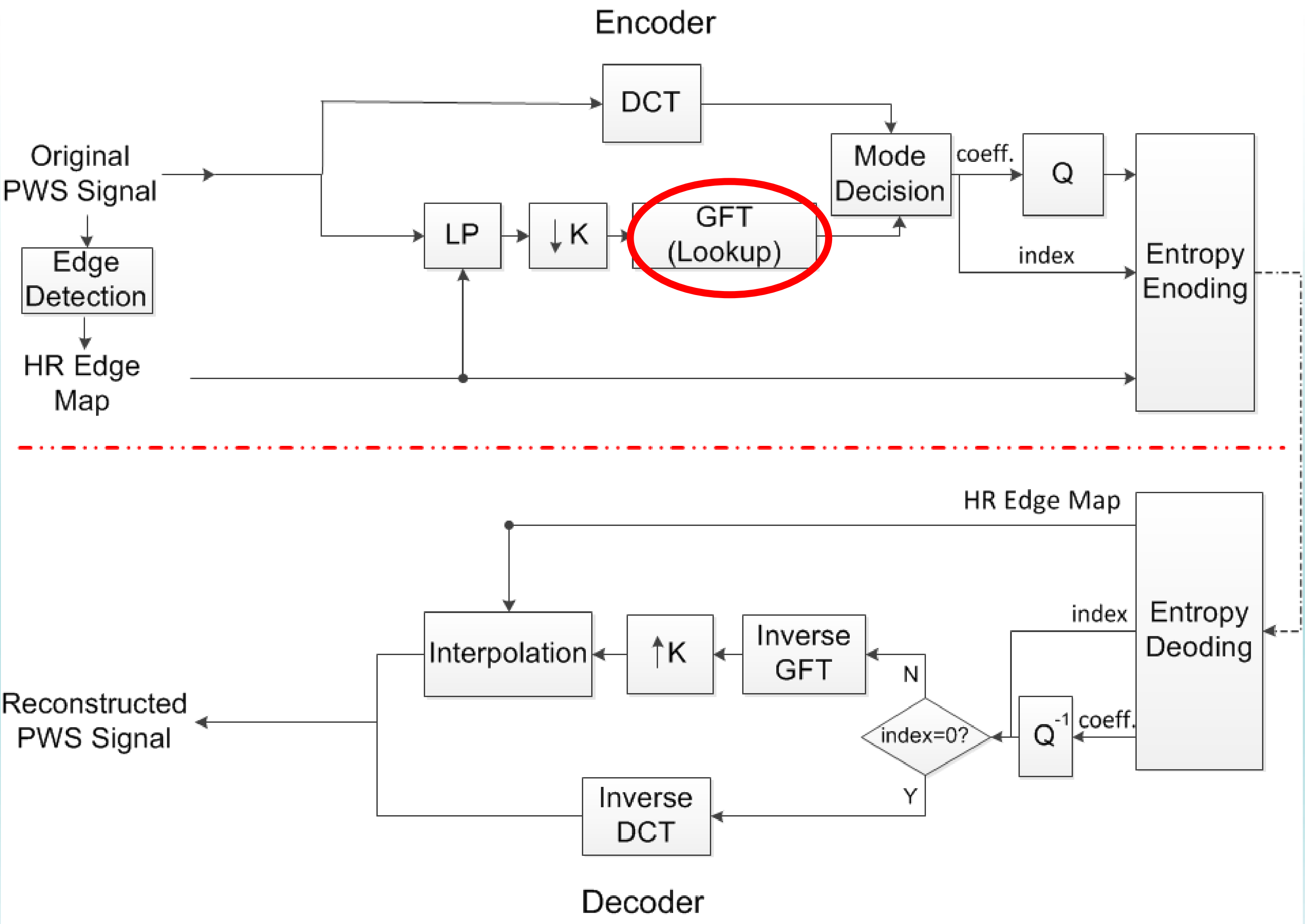
  - Exploit graph isomorphism

$$f : \mathcal{V}(\mathcal{G}) \rightarrow \mathcal{V}(\mathcal{H})$$

- Graph isomorphism



MR-GFT: Adaptive Selection of Graph Fourier Transforms



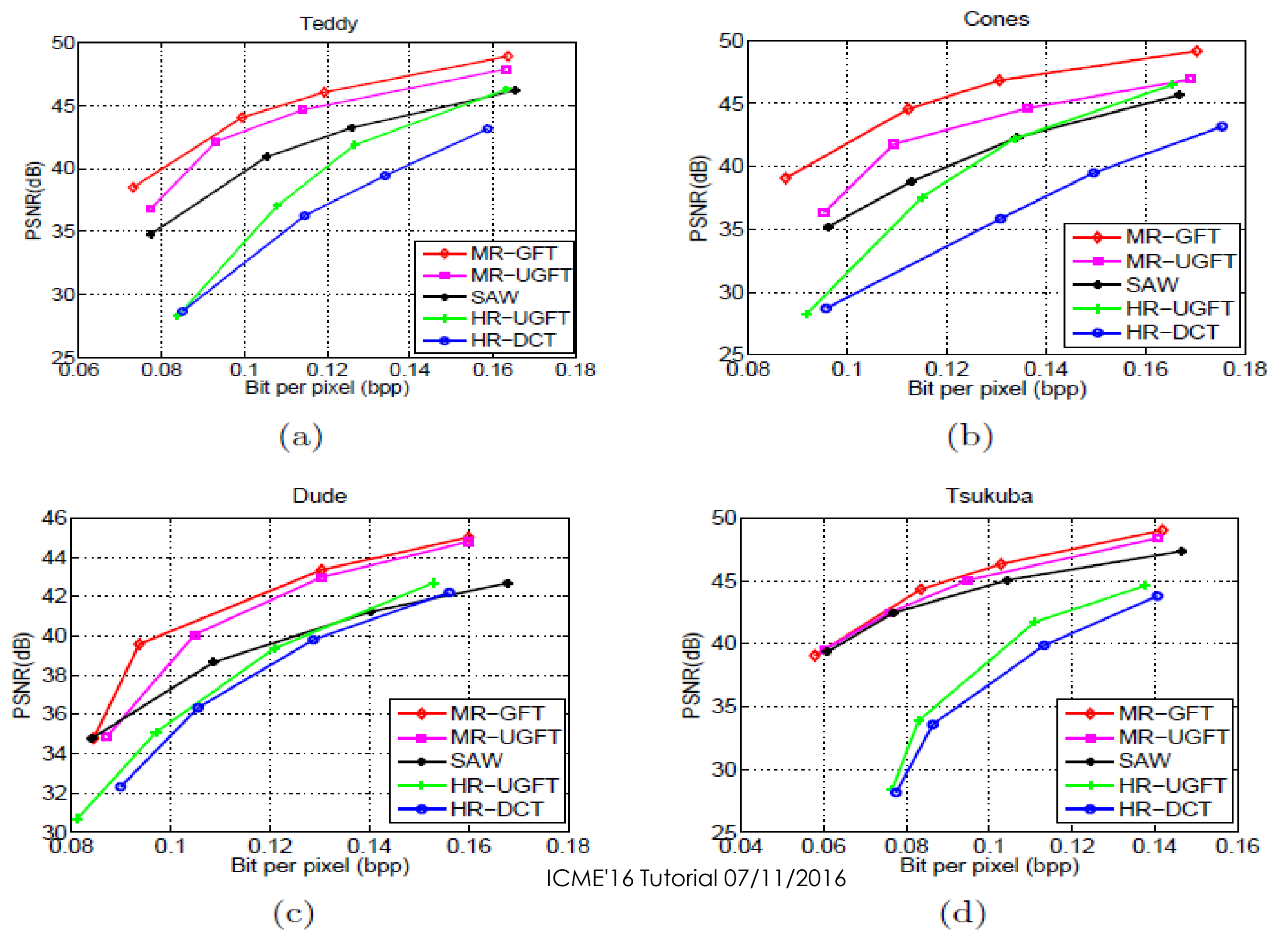
Experimentation

- Setup

  - Test images: depth maps of *Teddy* and *Cones*, and graphics images of *Dude* and *Tsukuba*.

  - Compare against: HR-DCT, HR-SGFT, SAW, MR-SGFT in H.264.

- Results

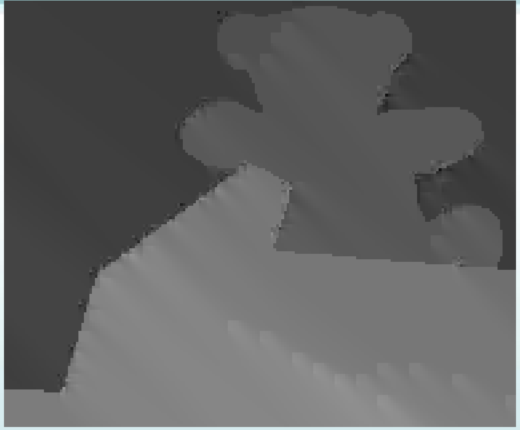


(a) (b)

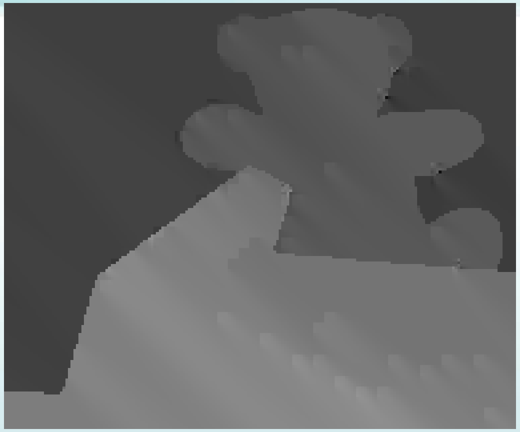
(c) (d)

HR-DCT:	6.8dB
HR-SGFT:	5.9dB
SAW:	2.5dB
MR-SGFT:	1.2dB

Subjective Results

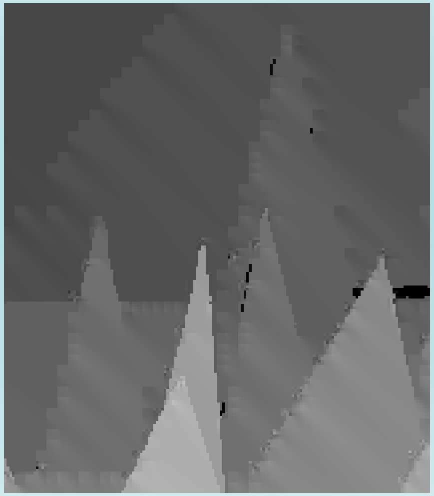


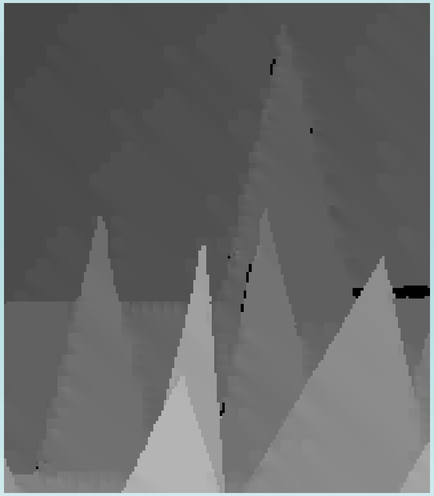
HR-DCT

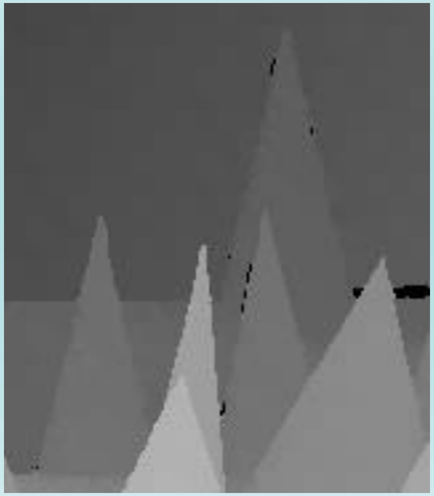


HR-SGFT

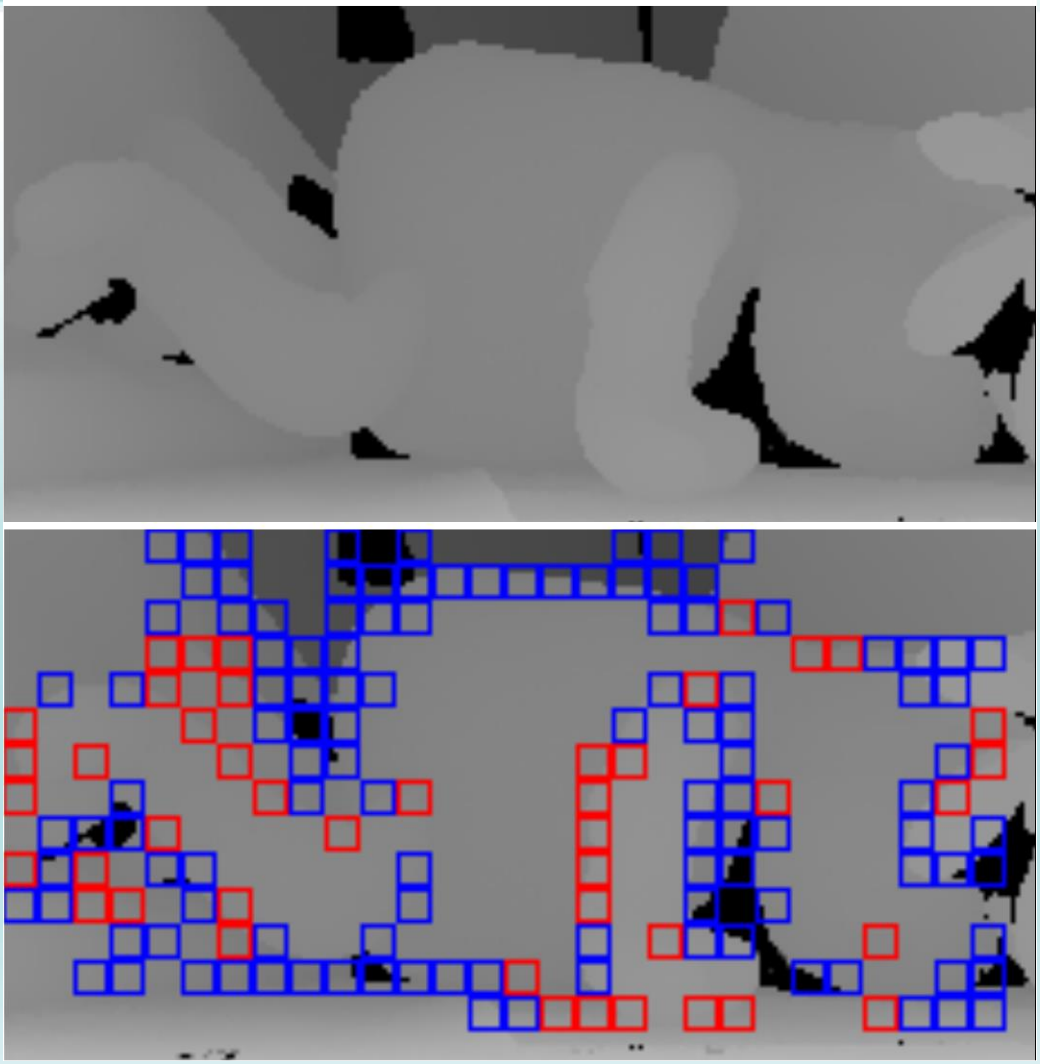
MR-GFT







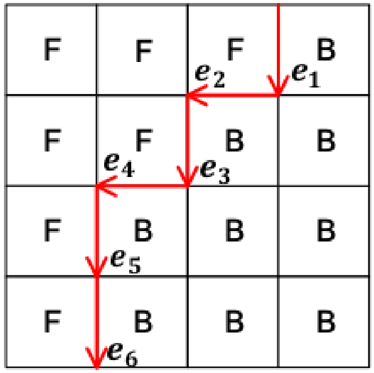
Mode Selection



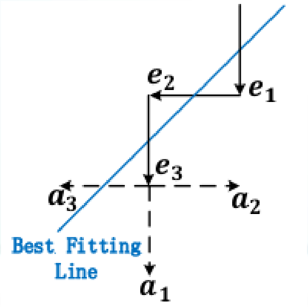
red: WGFT
blue: UGFT

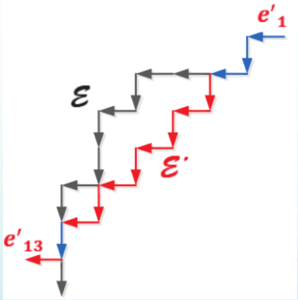
Edge Coding for PWS Image Compression

- **Arithmetic Edge Coding** [10]:
  - Coding of sequence of *between-pixel edges*, or chain code with symbols {L, S, R}.
  - Design a *context* to compute symbol probabilities for arithmetic coding.
  - **Extension**: better context based purely on symbol statistics analysis.

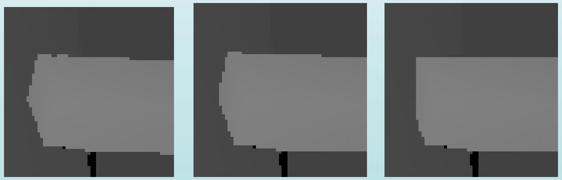


F	F	F	B
F	F	B	B
F	B	B	B
F	B	B	B

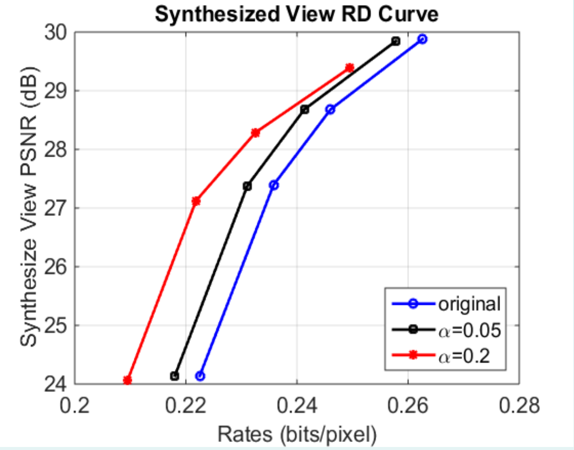




- **Contour Approximation & Depth Image Coding** [11]:
- Approximate contour while maintaining edge sharpness.
- Edge-adaptive blocked-based image coding (GFT).
- Average 1.68dB over GFT coding original contours at low-bitrate regions.



orig　Λ=800　Λ=3000



[10] I. Daribo, G. Cheung, D. Florencio, "**Arbitrarily Shaped Sub-block Motion Prediction in Depth Video Compression using Arithmetic Edge Coding**," *IEEE Trans on Image Processing*, Nov 2014.

[11] Y. Yuan G Cheung, P. Frossard, P. Le Callet, V. Zhao, "**Contour Approximation & Depth Image Coding for Virtual View Synthesis**," *IEEE MMSP*, October, 2015.

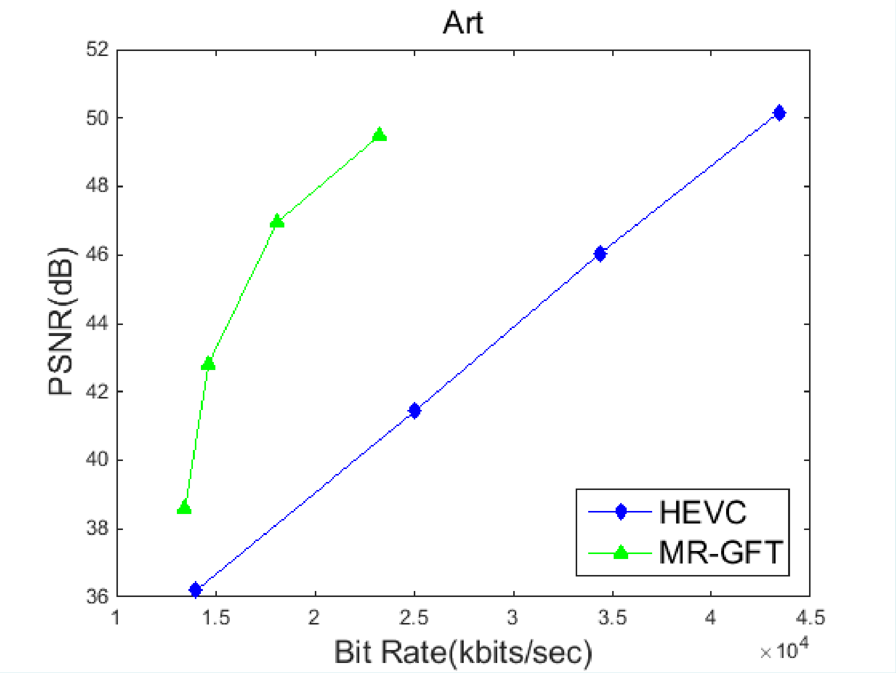
Experimentation

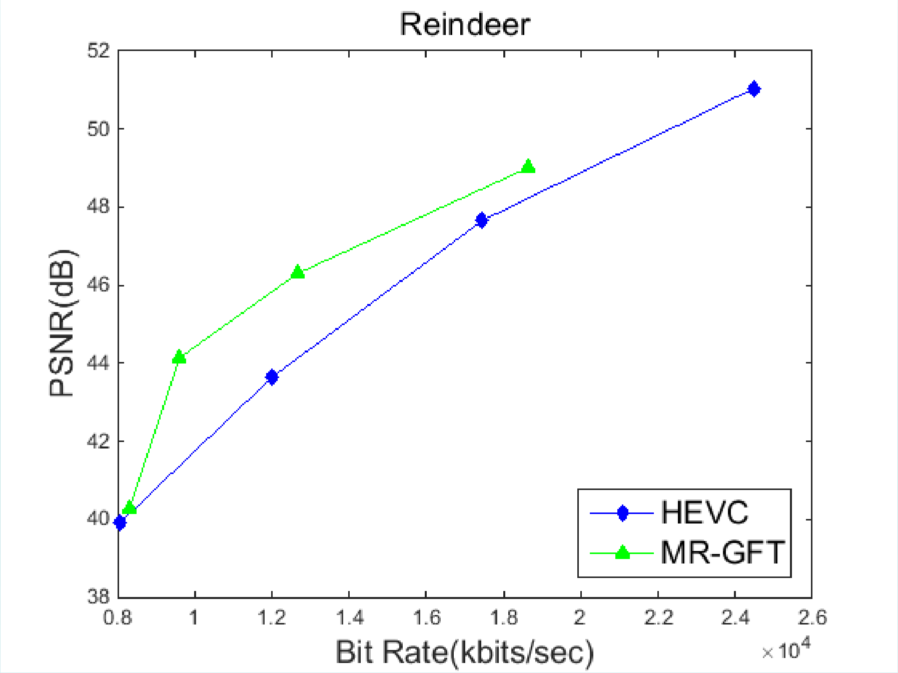
- Setup

  - AEC for contour coding plus MR-GFT.

  - Compare against: native HEVC

- Results





Summary for Multi-resolution Graph Fourier Transform

- A *multi-resolution (MR) graph Fourier transform (GFT)* coding scheme for compression of piecewise smooth images.

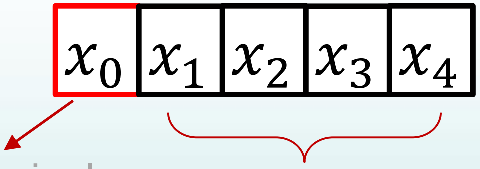
- Minimize transform representation cost + transform description cost given weight {0, 1, c}.
- Solve for optimal c, show optimality of GFT.
- **WGFT** {1, c}: formulate a separation-deviation (SD) problem.
- **UGFT** {1, 0}: greedy algorithm via spectral clustering.
- Practical implementation via multi-resolution, graph isomorphism and lookup tables.
- Excellent experimental results!

Outline (Part I)

- Fundamental of Graph Signal Processing (GSP)
  - Spectral Graph Theory
  - Graph Fourier Transform (GFT)
- Image Coding using GSP Tools
  - PWS Image Coding via Multi-resolution GFT
  - PWS Image Coding via Generalized GFT
  - Lifting Implementation of GFT

Motivation

- **Intra-prediction**

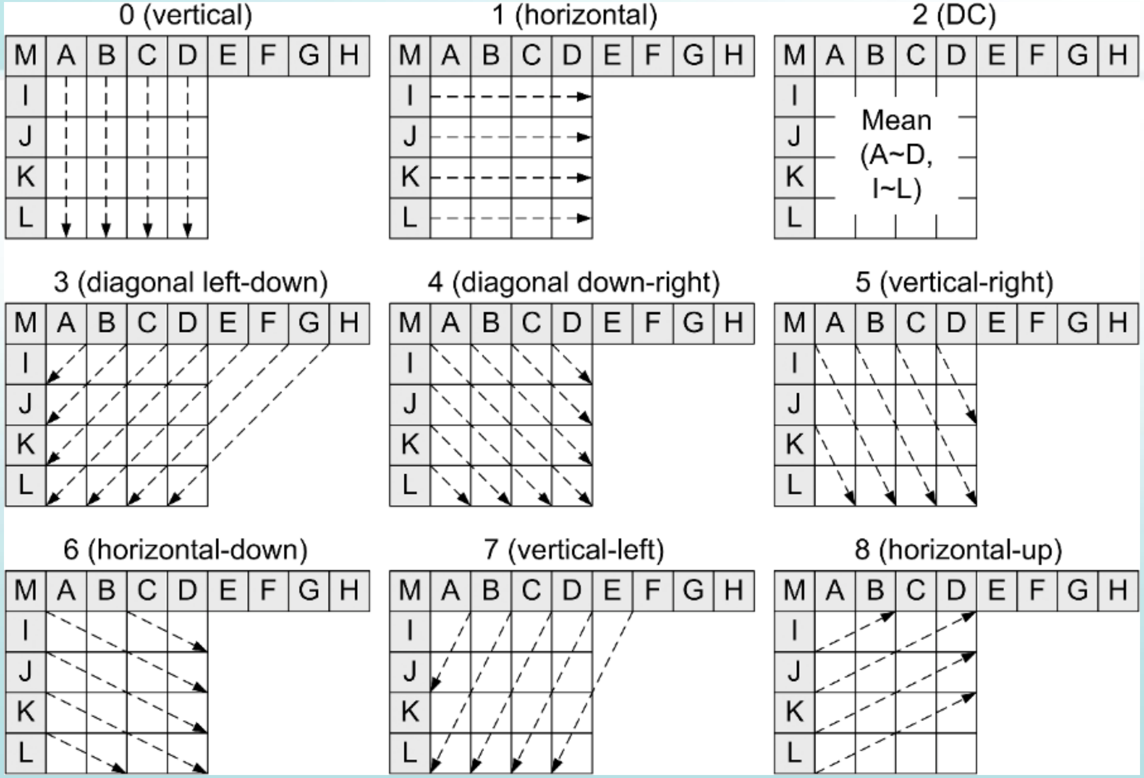


x_0 x_1 x_2 x_3 x_4

Boundary pixel (predictor)

Predicted pixels x_i

$x_i - x_0$: prediction residuals



Intra-prediction in H.264

- Discontinuities at block boundaries

  - intra-prediction will not be chosen or bad prediction

[12] W. Hu et al., "**Intra-Prediction and Generalized Graph Fourier Transform for Image Coding**," *IEEE Signal Processing Letters*, vol.22, no.11, pp. 1913-1917, November, 2015.

Contributions

- *Clustered-based* intra-prediction

  - **cluster** discontinuities at block boundaries

  - $\mu + x_0$: shift by cluster mean μ (side information)

- *Generalized* Graph Fourier Transform (GGFT)

  - **optimized** for intra-prediction residuals

  - generalized graph Laplacian: extra weight added at block boundaries

  - default to the DCT and ADST in some cases

Related Work

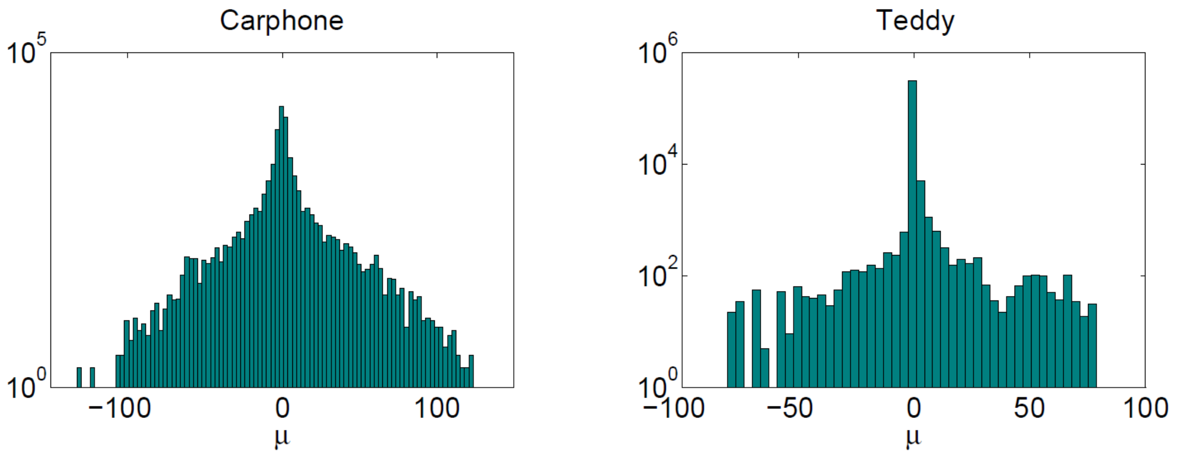
- Zhang et al, graph-based predictive transform coding, GMRF [5]
  - Assume given model. No discussion on how to derive model parameters.
- Wang et al, intra predictive graph transform coding [13]
  - Intra-prediction plus KLT, optional graph sparsification.
- Ye et al, MDDT [14]
  - Completely data-driven resulting in unstructured transform.

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

[13] Y. Wang, A. Ortega, and G. Cheung, "Intra predictive transform coding based on predictive graph transform," *ICIP*, September 2013.

[14] Y. Ye and M. Karczewicz, "Improved H.264 intra coding based on bidirectional intra prediction, directional transform, and adaptive coefficient scanning," *ICIP*, October 2008.

1D Signal Modeling

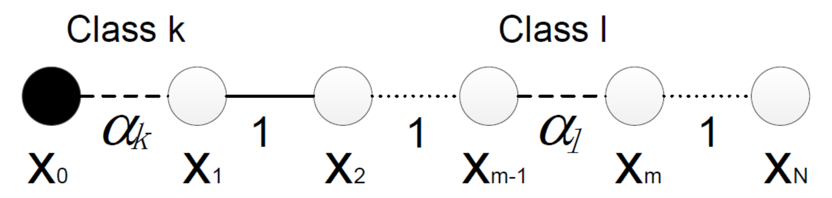


- Inter-pixel differences are concentrated around 0, occasionally large.

- Quantize inter-pixel differences into K bins, leveraging on *Lloyd algorithm*

$$x_n = x_{n-1} + \underbrace{\hat{\mu}_{i(\mu_n)}}_{\text{bin average}} + \underbrace{g_{i(\mu_n)}}_{\text{approximation error}}$$

Optimal 1D Intra prediction



$$\begin{bmatrix} x_0 + \hat{\mu}_a \\ \vdots \\ x_0 + \hat{\mu}_a \\ x_0 + \hat{\mu}_a + \hat{\mu}_b \\ \vdots \\ x_0 + \hat{\mu}_a + \hat{\mu}_b \end{bmatrix} \begin{array}{l} \rightarrow \text{Class k} \\ \\ \\ \rightarrow \text{Class l} \\ \\ \\ \end{array}$$

- Optimal in terms of resulting in a *zero-mean* prediction residual
- Default to conventional intra-prediction when $\hat{\mu}_a = \hat{\mu}_b = 0$, i.e.,

$$[x_0, ..., x_0]^T$$

Generalized Graph Fourier Transform

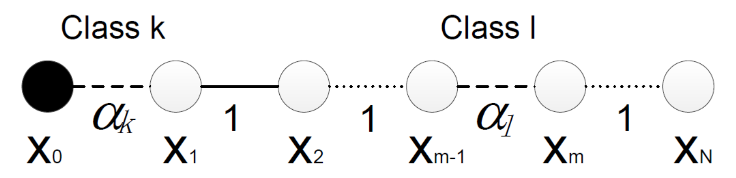
- The precision matrix of the prediction residual

$$\begin{bmatrix}
\boxed{\alpha_a}+1 & -1 & & & & & & \\
-1 & 2 & -1 & & & & & \\
& \ddots & \ddots & \ddots & & & & \\
& & -1 & 1+\boxed{\alpha_b} & -\alpha_b & & & \\
& & & -\alpha_b & \alpha_b+1 & -1 & & \\
& & & & \ddots & \ddots & \ddots & \\
& & & & & -1 & 2 & -1 \\
& & & & & & -1 & 1
\end{bmatrix}$$

$\alpha_a = \sigma^2_{g_0} / \sigma^2_{g_a}$ (inaccuracy of intra-prediction)

$\alpha_b = \sigma^2_{g_0} / \sigma^2_{g_b}$ (discontinuities within signal)

Variance of approximation error



Generalized Graph Fourier Transform

- The precision matrix of the prediction residual

$$\begin{bmatrix} \enclose{circle}{\alpha_a}+1 & -1 & & & & & & \\ -1 & 2 & -1 & & & & & \\ & \ddots & \ddots & \ddots & & & & \\ & & -1 & 1+\enclose{circle}{\alpha_b} & -\alpha_b & & & \\ & & & -\alpha_b & \alpha_b+1 & -1 & & \\ & & & & \ddots & \ddots & \ddots & \\ & & & & & -1 & 2 & -1 \\ & & & & & & -1 & 1 \end{bmatrix}$$

Generalized Laplacian

$$= \begin{bmatrix} 1 & -1 & & & & & & \\ -1 & 2 & -1 & & & & & \\ & \ddots & \ddots & \ddots & & & & \\ & & -1 & 1+\alpha_b & -\alpha_b & & & \\ & & & -\alpha_b & \alpha_b+1 & -1 & & \\ & & & & \ddots & \ddots & \ddots & \\ & & & & & -1 & 2 & -1 \\ & & & & & & -1 & 1 \end{bmatrix} + \begin{bmatrix} \alpha_a & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \end{bmatrix}$$

Combinatorial Laplacian

Degree matrix for boundary vertices

Generalized Graph Fourier Transform

- The precision matrix of the prediction residual

$$\begin{bmatrix} \enclose{circle}{\alpha_a}+1 & -1 & & & & & & & \\ -1 & 2 & -1 & & & & & & \\ & \ddots & \ddots & \ddots & & & & & \\ & & -1 & 1+\enclose{circle}{\alpha_b} & -\alpha_b & & & & \\ & & & -\alpha_b & \alpha_b+1 & -1 & & & \\ & & & & \ddots & \ddots & \ddots & & \\ & & & & & -1 & 2 & -1 \\ & & & & & & -1 & 1 \end{bmatrix}$$

  - Default to the DCT if $\alpha_a = 0$ and $\alpha_b = 1$
  - Default to the ADST [15] if $\alpha_a = 1$ and $\alpha_b = 1$

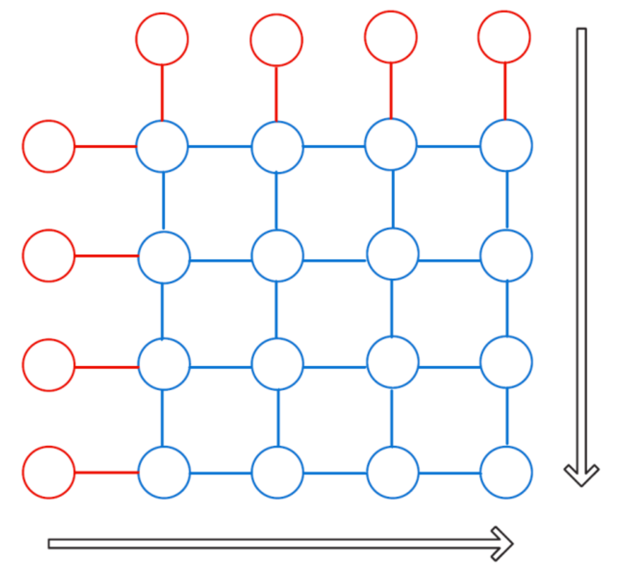
[15] J. Han et al., "**Jointly Optimized Spatial Prediction and Block Transform for video and Image Coding**," *IEEE Transactions on Image Processing*, vol.21, no.4, April 2012, pp.1874-1884.

Proposed Coding System

- Four clusters:
  - Strong correlation: $\hat{\mu}_0 = 0$
  - Weak correlations: $\hat{\mu}_{-1} < 0 < \hat{\mu}_1$
  - Zero correlation

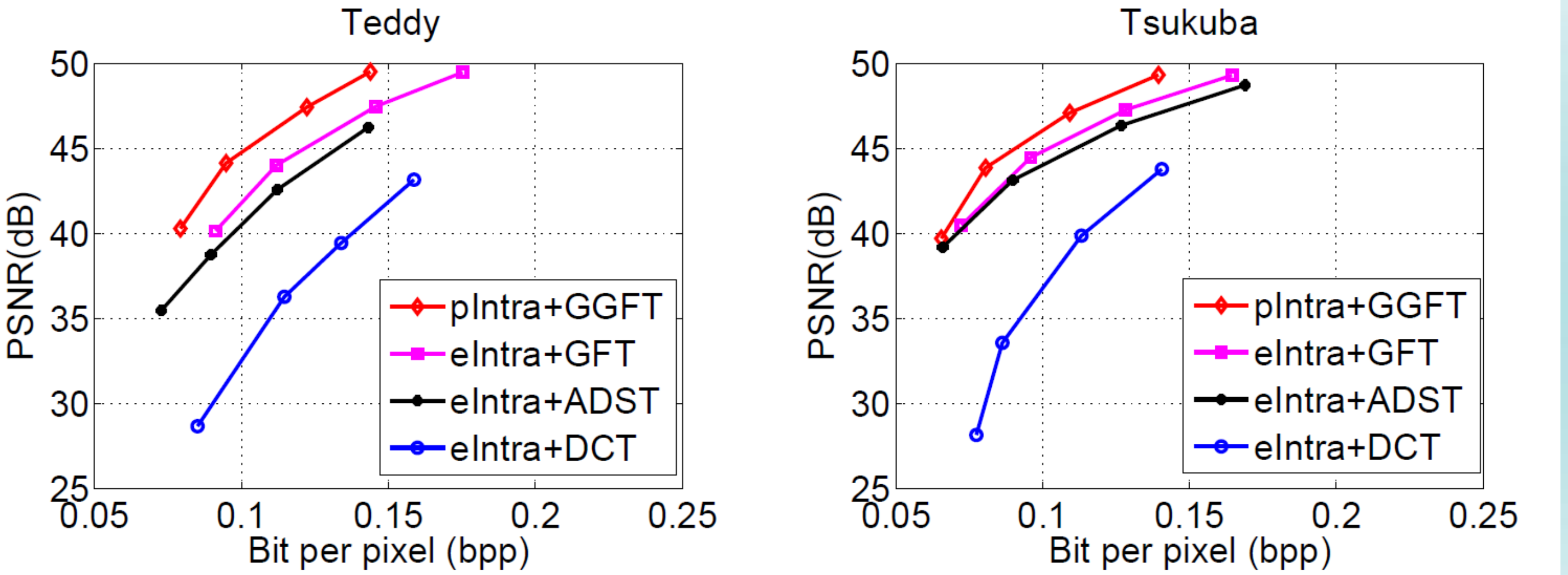
- Side information:
  - contours: arithmetic edge coding
  - cluster indicator: arithmetic coding

- 2D prediction and transform



Experimental Results

- Test images: PWS images and natural images
- Compare *proposed intra-prediction (pIntra) + GGFT* against:
  - edge-aware intra-prediction (eIntra) + DCT
  - eIntra + ADST
  - eIntra + GFT



(a) Teddy (b) Tsukuba

Experimental Results

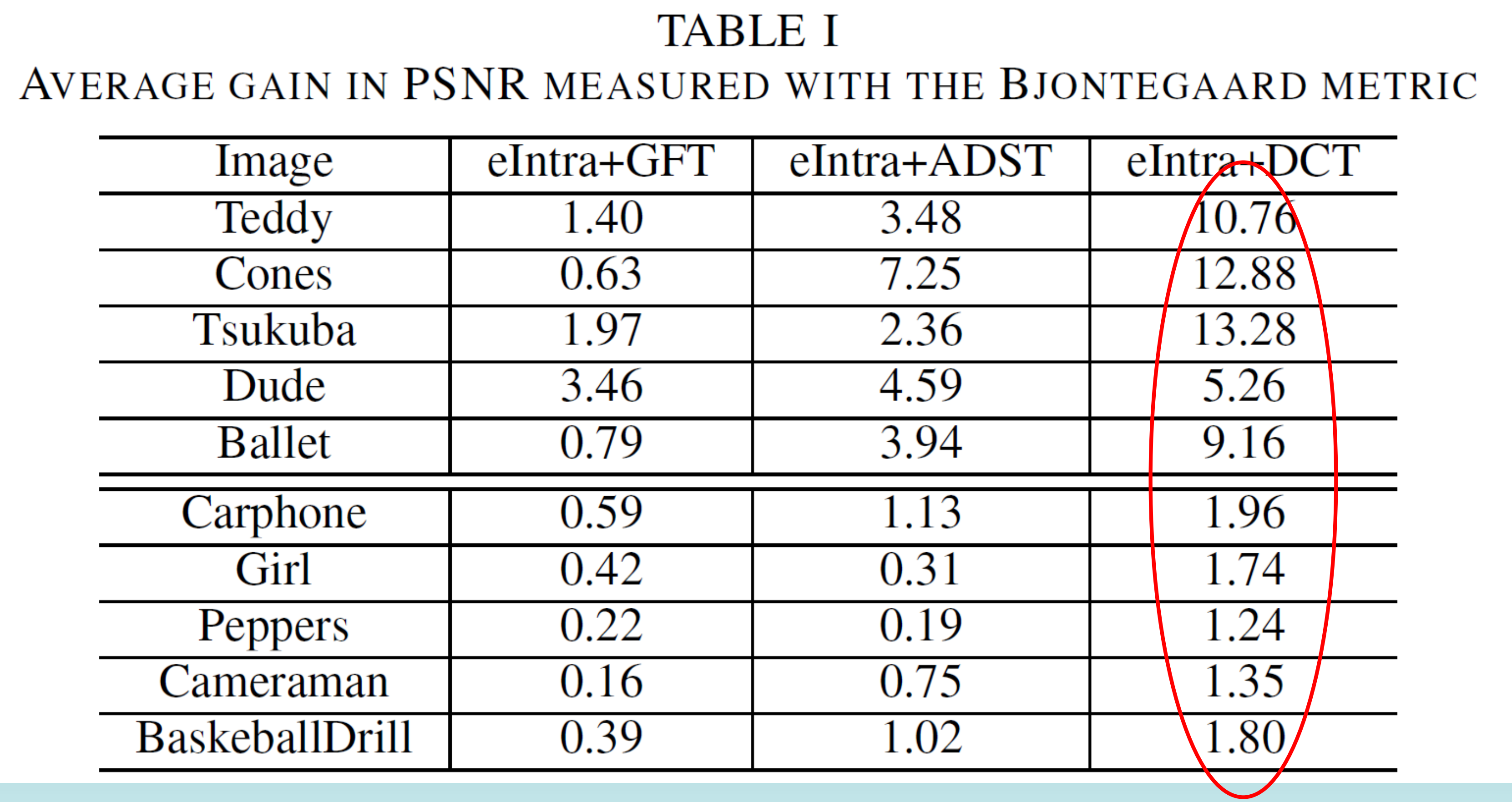
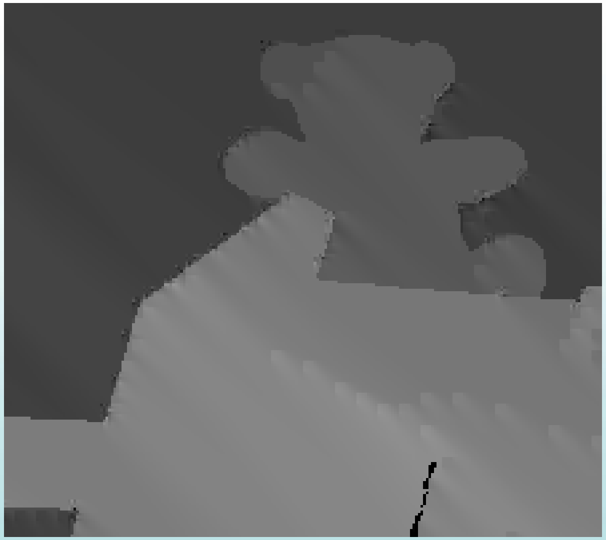


TABLE I

AVERAGE GAIN IN PSNR MEASURED WITH THE BJONTEGAARD METRIC

Image	eIntra+GFT	eIntra+ADST	eIntra+DCT
Teddy	1.40	3.48	10.76
Cones	0.63	7.25	12.88
Tsukuba	1.97	2.36	13.28
Dude	3.46	4.59	5.26
Ballet	0.79	3.94	9.16
Carphone	0.59	1.13	1.96
Girl	0.42	0.31	1.74
Peppers	0.22	0.19	1.24
Cameraman	0.16	0.75	1.35
BasketballDrill	0.39	1.02	1.80

Subjective Quality



eIntra + DCT

eIntra + GFT

pIntra + GGFT

Outline (Part I)

- Fundamental of Graph Signal Processing (GSP)
  - Spectral Graph Theory
  - Graph Fourier Transform (GFT)
- Image Coding using GSP Tools
  - PWS Image Coding via Multi-resolution GFT
  - PWS Image Coding via Generalized GFT
  - Lifting Implementation of GFT

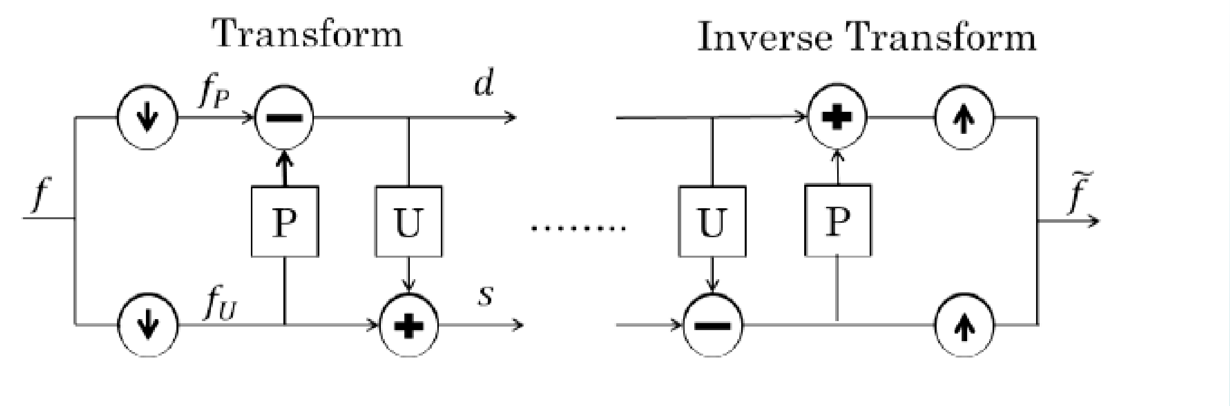
Motivation

- Complexity of GFT
  - Transform matrices are dense with floating point values
  - Complexity of transform operation: $O(N^2)$, N : number of nodes
- Contribution
  - Design low complexity transform w/ graph-based lifting $O(N \log N)$
    - Novel bipartition approach in the lifting transform
    - Novel multi-level lifting filterbanks design
  - Performance approximates the GFT for intra-predicted video coding

[16] Yung-Hsuan Chao, Antonio Ortega, and Sehoon Yea, "**Graph-based lifting transform for intra-predicted video coding**", *IEEE International Conference on Acoustics, Speech and Signal Processing* (ICASSP), 2016, pp. 1140-1144

Preliminaries - Lifting transform

- 1-level lifting transform

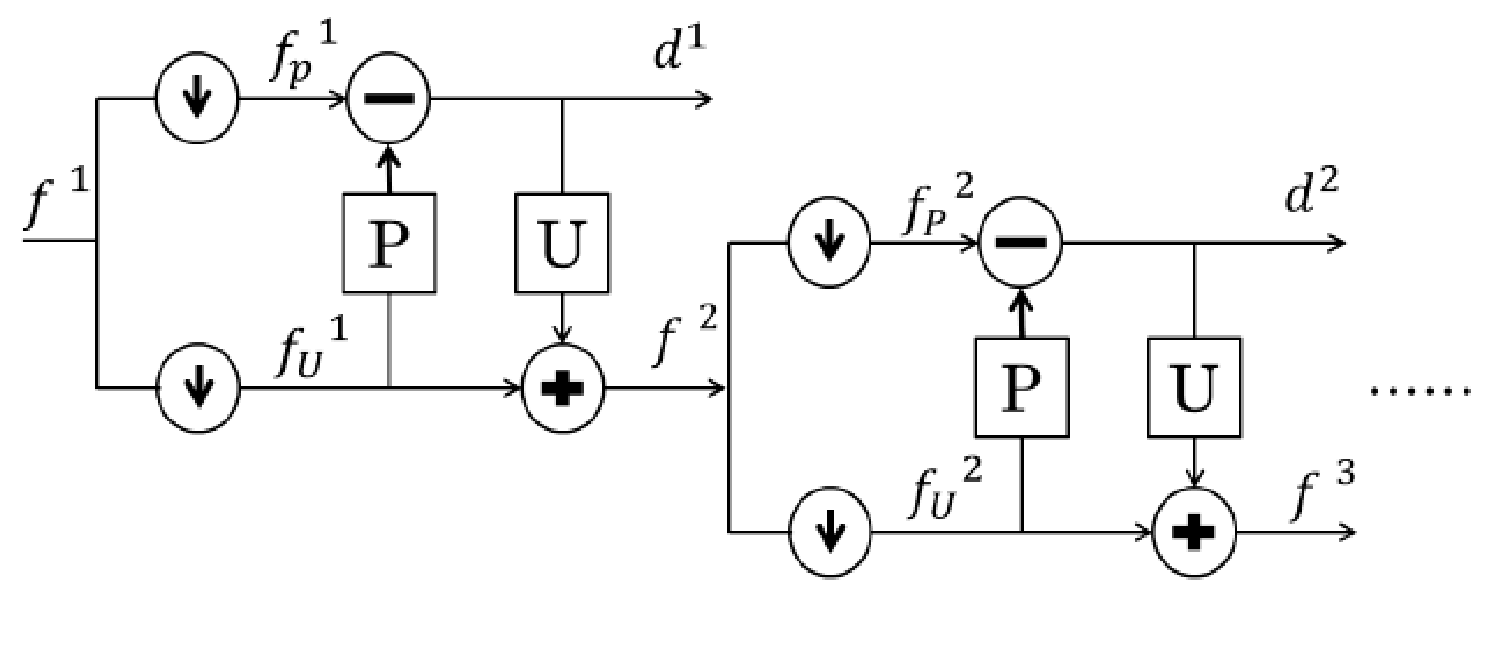


- The transform consists of 3 steps:
- (1) Bipartition into two disjoint sets S_U and S_P (signal f_U and f_P)
- (2) Predict signal f_P in S_P with nodes in S_U
- (3) Update signal f_U in S_U with the prediction residue stores in S_P

[17] W. Sweldens, “**The lifting scheme: A construction of second generation wavelets**,” Technical report 1995:6, Industrial Mathematics Initiative, Department of Mathematics, University of South Carolina.

Preliminaries - Lifting transform

- Multi-level lifting transform:
- Dyadic decomposition for S_U in each level



Bipartition Algorithm

- **Goal**: minimize energy of prediction residue (high frequency) stored in S_P → increase energy compaction
- Applied signal model
  - Gaussian Markov Random Field (GMRF): $\mathrm{f} \sim N\left(0, \Sigma = \mathrm{Q}^{-1}\right)$
  - Inverse covariance matrix $\mathrm{Q} = \sigma \mathrm{L} \rightarrow \sigma\left(\mathrm{L} + \delta \mathrm{L}\right)$
- Proposed bipartition $\mathrm{S}_U^* = \arg\min_{\mathrm{S}_U} E\left[\left\|\mathrm{f}_P - \mu_{P|U}\right\|^2\right]$
- Where $\mu_{P|U}$ is the MAP predictor of $\mathrm{f}_P \mid \mathrm{f}_U$
- NP hard → greedy approximation
- Initialization: $\mathrm{S}_U = 0$ and $\mathrm{S}_P = V$
- Select node $\mathrm{v}^* \in \mathrm{S}_P = \arg\min_{v} E\left[\left\|\mathrm{f}_P - \mu_{P|(U \cup \{v\})}\right\|^2\right]$ iteratively

Bipartite Graph Formulation

- Proposed method

1. Connecting nodes in S_P using Kron reduction* → bipartite graph A_{BP}

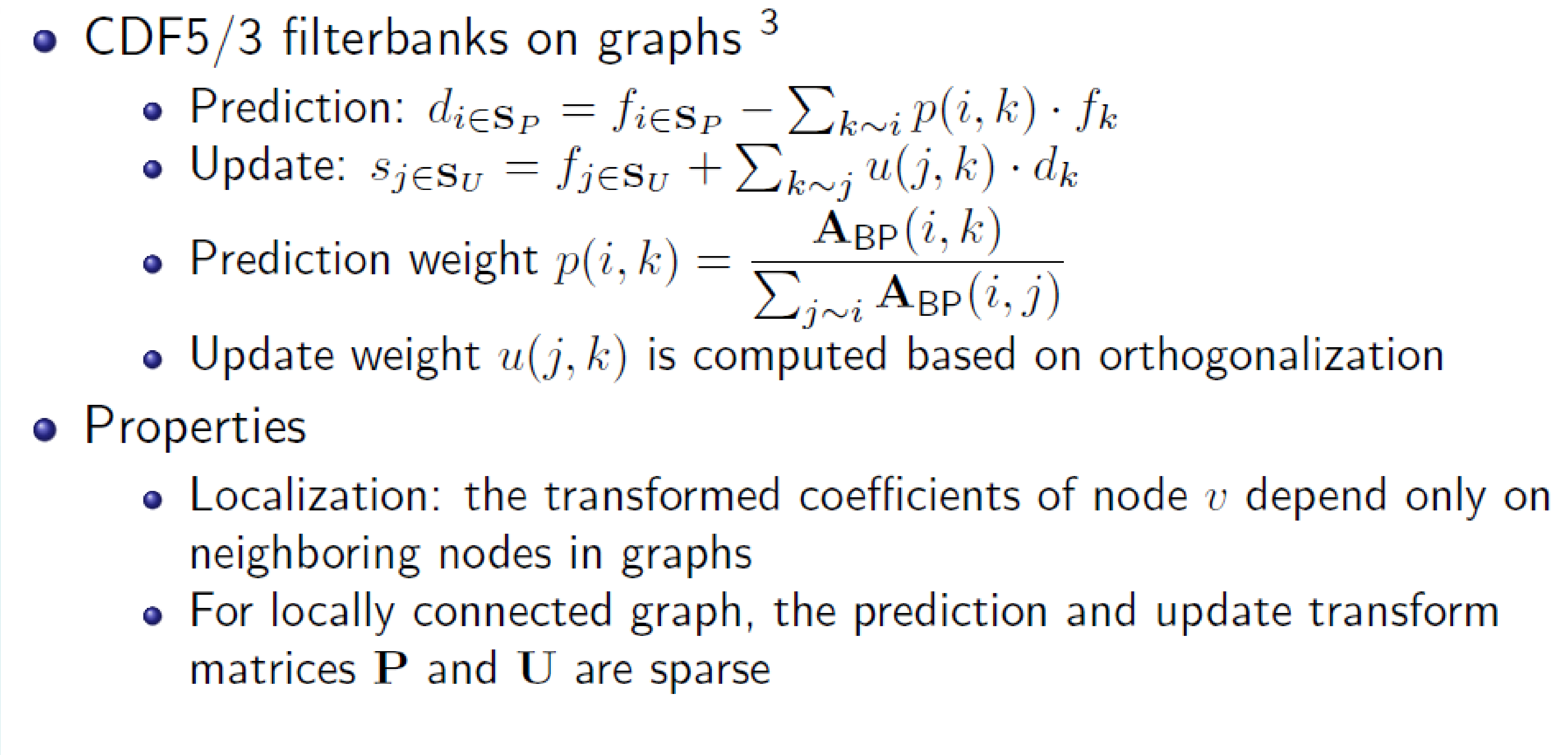
- Procedure

1. For each node v in S_P
2. Remove nodes in $S_P / \{v\}$
3. Apply Kron reduction on the resulting subgraph
    1. derive connection from v to S_U
4. Prune the links of v by keeping the 4 links with largest weights
5. Repeat the same procedure for all the nodes in S_P

---

* Dorfler, Florian and Bullo, Francesco, "Kron reduction of graphs with applications to electrical networks," IEEE Transactions on Circuits and Systems I 2013

Filterbanks Design

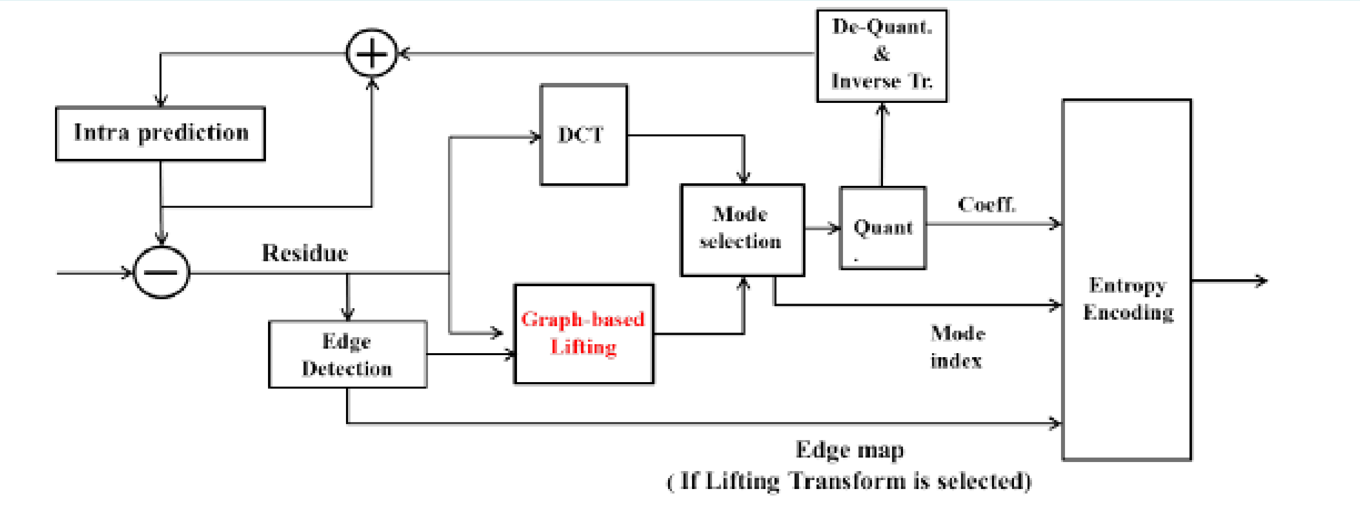


- CDF5/3 filterbanks on graphs [3]
  - Prediction: $d_{i \in \mathbf{S}_P} = f_{i \in \mathbf{S}_P} - \sum_{k \sim i} p(i,k) \cdot f_k$
  - Update: $s_{j \in \mathbf{S}_U} = f_{j \in \mathbf{S}_U} + \sum_{k \sim j} u(j,k) \cdot d_k$
  - Prediction weight $p(i,k) = \dfrac{\mathbf{A}_{\text{BP}}(i,k)}{\sum_{j \sim i} \mathbf{A}_{\text{BP}}(i,j)}$
  - Update weight $u(j,k)$ is computed based on orthogonalization
- Properties
  - Localization: the transformed coefficients of node v depend only on neighboring nodes in graphs
  - For locally connected graph, the prediction and update transform matrices $\mathbf{P}$ and $\mathbf{U}$ are sparse

* Shen, Godwin, and Antonio Ortega. Tree-based wavelets for image coding: Orthogonalization and tree selection. PCS 2009

Experimental Results - Simulation Setup Test

- Test Sequences: *Foreman*, *Mobile*, *Silent*, *Deadline*
- Extract intra-predicted residual blocks from HEVC (HM-14) encoder
- Fixed block size as 8 x 8
- Transform coefficients are uniformly quantized and then encoded
- using a symbol grouping-based arithmetic entropy encoder (AGP*)
- Edge Map is encoded using arithmetic edge coding (AEC**)

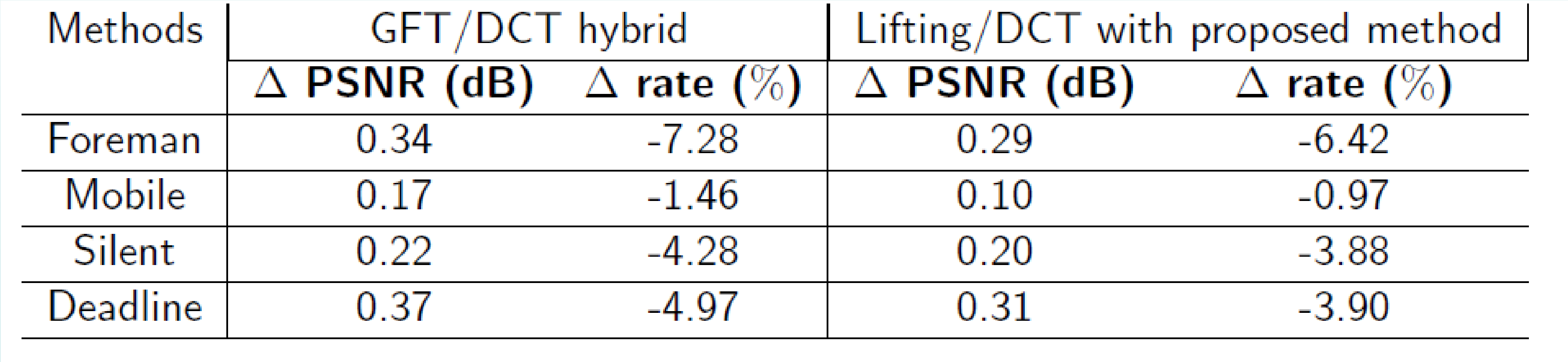


*A. Said and W. A. Pearlman, Low-complexity waveform coding via alphabet and sample-set partitioning, SPIE Visual Communications and Image Processing (1997)

** Daribo, Ismael and Cheung, Gene and Florencio, Dinei, Arithmetic edge coding for arbitrarily shaped sub-block motion prediction in depth video compression, ICIP 2012

Experimental Results

- Hybrid mode between proposed lifting and DCT:
- based on RD cost *SSE* + λ *Bitrate*
- Bjontegaard Distortion-Rate against DCT



Methods	GFT/DCT hybrid		Lifting/DCT with proposed method	
	Δ PSNR (dB)	**Δ rate (%)**	**Δ PSNR (dB)**	**Δ rate (%)**
Foreman	0.34	-7.28	0.29	-6.42
Mobile	0.17	-1.46	0.10	-0.97
Silent	0.22	-4.28	0.20	-3.88
Deadline	0.37	-4.97	0.31	-3.90

Summary (Part I)

- GFT is signal-adaptive transform.
- Adaptive transform → energy compaction.
- Adaptive transform → signal overhead.
- GFT approx. KLT assuming 1D AR model.
- GGFT approx. KLT for prediction residual.
- PWS signal, signaling can be coded contour via AEC.
- Lifting Implementation of GFT.

Open Questions

- Coding of Natural images:
  - What is optimal clustering?
  - What is optimal graph, optimal transform?
  - Joint intra-prediction / transform design?
- Coding of 3D data:
  - How to use GSP tools to code dynamic 3D geometry?
  - How to use GSP tools to code light field data?

Q&A

- Email: cheung@nii.ac.jp
- Homepage: http://research.nii.ac.jp/~cheung/