Gene Cheung
National Institute of Informatics
6th July, 2015

Depth Image Coding & Processing Part 1: Introduction

.

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~).

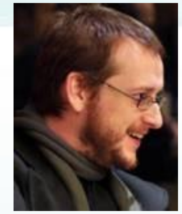
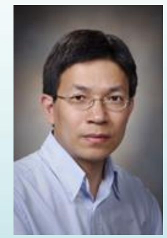
3D video Communication (8 yrs)

- Faculty 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.

Acknowledgement

Collaborators:

- Y. Mao, X. Liu, 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)
- T. Maugey, L. Toni, B. Motz, 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)
- B. Machiavello, C. Dorea, M. Hung (University of Brasilia, Brazil)





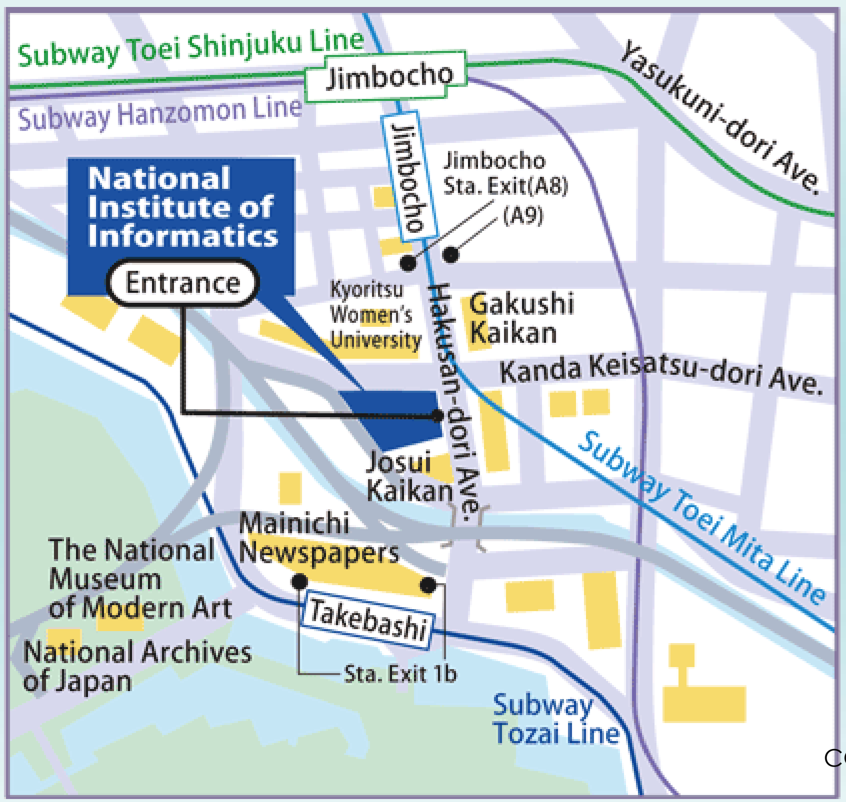







NII Overview

- **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

- Introduction
  - What is 3D imaging?
  - Key problems: 3D video coding & streaming
  - Other problems: denoising, interpolation, bit-depth enhancement, de-quantization
  - Applications: joint gaze correction / face beautification, sleep monitoring, heart rate estimation
- Fundamental of Graph Signal Processing (GSP)
  - Spectral Graph Theory
  - Graph-based Transform

Outline

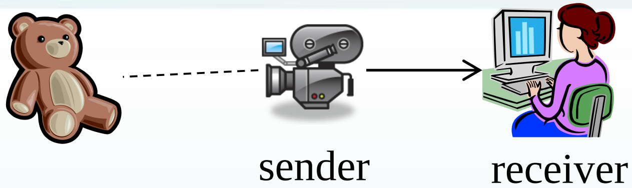
- Introduction
  - What is 3D imaging?
  - Key problems: 3D video coding & streaming
  - Other problems: denoising, interpolation, bit-depth enhancement, de-quantization
  - Applications: joint gaze correction / face beautification, sleep monitoring, heart rate estimation
- Fundamental of Graph Signal Processing (GSP)
  - Spectral Graph Theory
  - Graph-based Transform

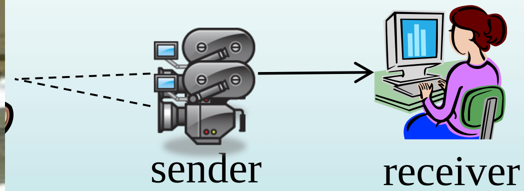
Video Communication: 2D to 2.5D to 3D

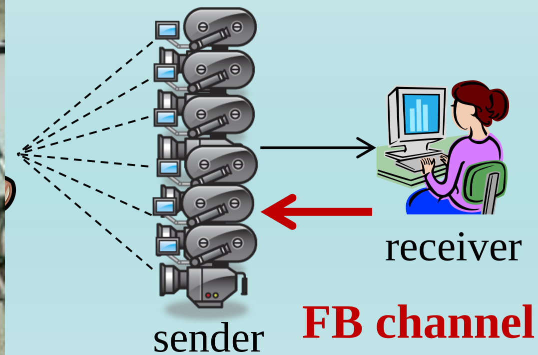
Microsoft Kinect

Mesa's SwissRanger

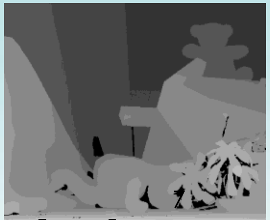
- **2D Video**
  - 1 capturing camera @ sender.
  - 1 2D display @ receiver (non-interactive).







texture map



depth map

Multiview Video Streaming

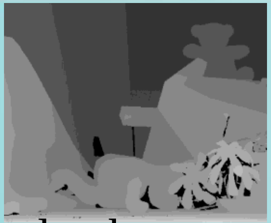
- Interactive view-switches among captured camera viewpoints.

Free Viewpoint Video Streaming

- Interactive view-switches to ***any*** virtual camera viewpoints.

*Courtesy of KDDI Laboratories, Japan

texture map

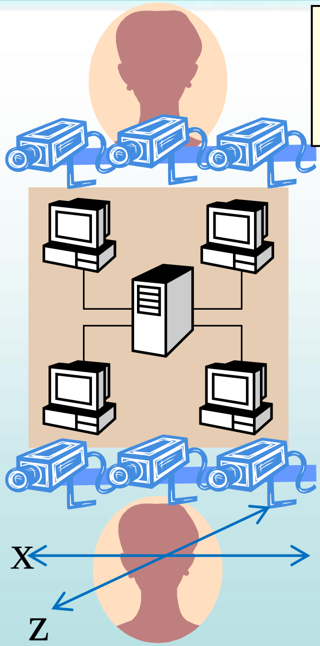


depth map

3D Imaging: Overview

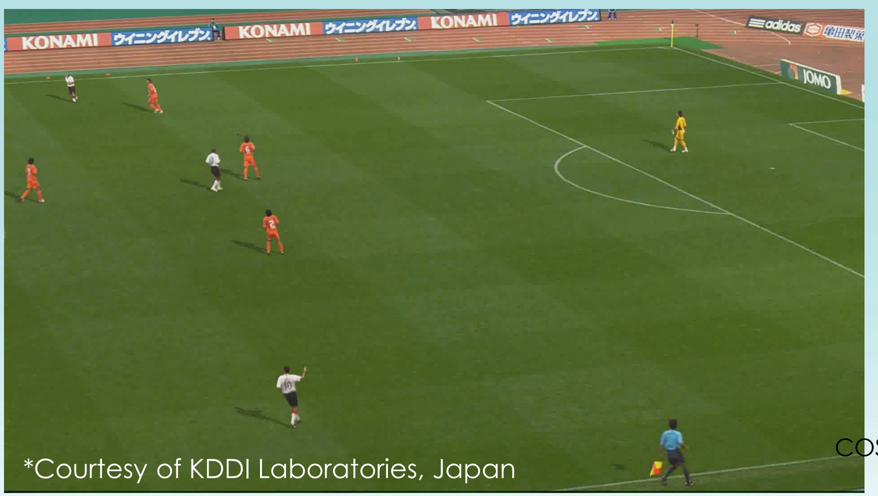
Microsoft Kinect

Mesa's SwissRanger



- **3D Imaging?** *Given depth sensors capturing texture & geometry per pixel, process multi-d image data for 3D apps.*

- **App 1**: ***Immersive visual communication***:
  - video conferencing (gaze correction, motion parallax)
  - free viewpoint TV (user view selection)
  - mixed reality (virtual objs + real view)

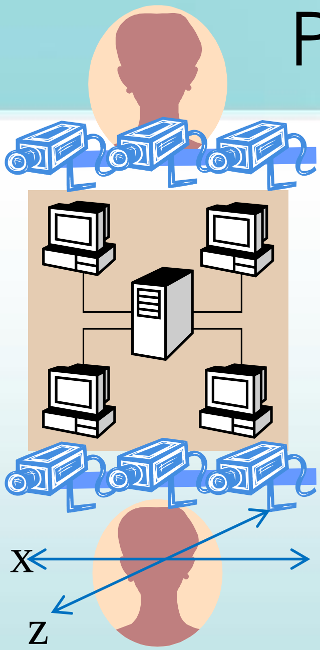


*Courtesy of KDDI Laboratories, Japan

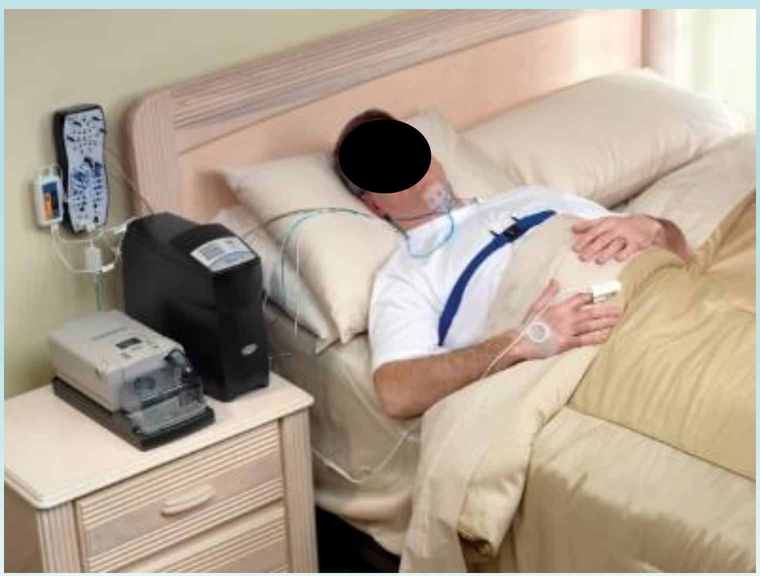
App 2: **3D data sensing & analysis**:

- 3D event detection
- Non-intrusive health / sleep monitoring

Potential Impact



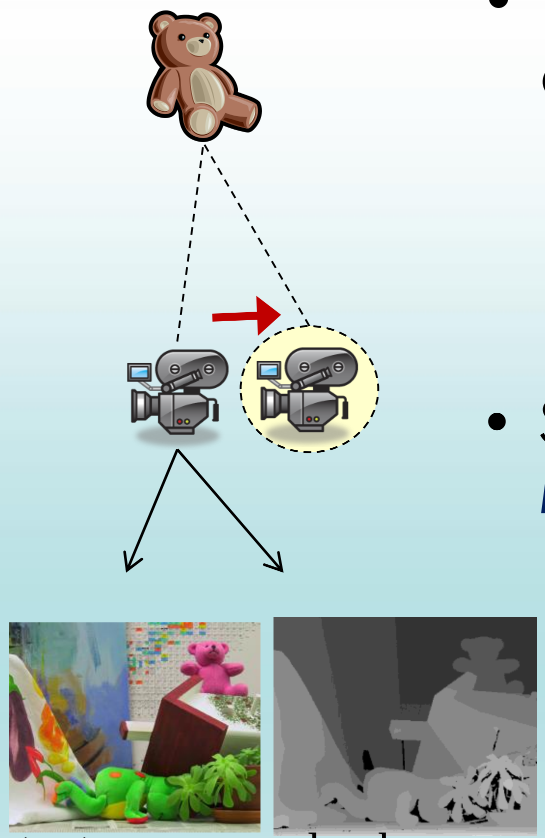
- Commercial Immersive Communication
  - Immersive entertainment (Disney, Samsung).
  - Immersive conferencing (KDDI, Cisco).
  - 3D visualization (Microsoft, Google).
- Enhance Virtual Reality is 1 of 14 grand challenges chosen by ***National Academy of Engineering*** for 21st century.



- Growing Market for Health Monitoring
  - 1 / 4.5 Japanese have sleep disorders.
  - Sleep clinic is 7 billion $ US market.
  - *Imaging means "**Non-intrusiveness**".*

NATIONAL ACADEMY OF ENGINEERING
OF THE NATIONAL ACADEMIES

3D Video Representation



texture map depth map

- Texture + depth maps from 1 or more camera viewpoints [1].
  - **Texture map**: color image like RGB.
  - **Depth map**: per-pixel distance bt'n captured objects in 3D scene & capturing camera.
- Synthesis of intermediate views via *depth-image-based rendering* (DIBR) [2].
  - Computation-efficient.
  - Unlike model-based approach, complexity not scene-dependent.

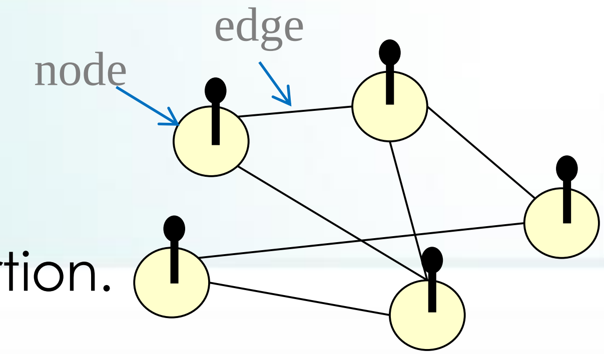
[1] P. Merkle, et al., "**Multi-view Video Plus Depth Representation and Coding**," *IEEE International Conference on Image Processing*, San Antonio, TX, October 2007,.

[2] D. Tian, et al., "**View synthesis techniques for 3D Video**," *Applications of Digital Image Processing XXXII, Proceedings of the SPIE*, vol. 7443, (2009), San Diego, CA, February 2009,.

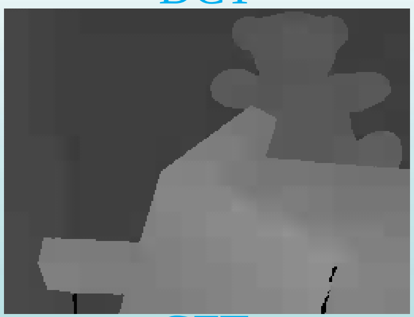
Outline

- Introduction
  - What is 3D imaging?
  - Key problems: 3D video coding & streaming
  - Other problems: denoising, interpolation, bit-depth enhancement, de-quantization
  - Applications: joint gaze correction / face beautification, sleep monitoring, heart rate estimation
- Fundamental of Graph Signal Processing (GSP)
  - Spectral Graph Theory
  - Graph-based Transform

Depth Image Compression



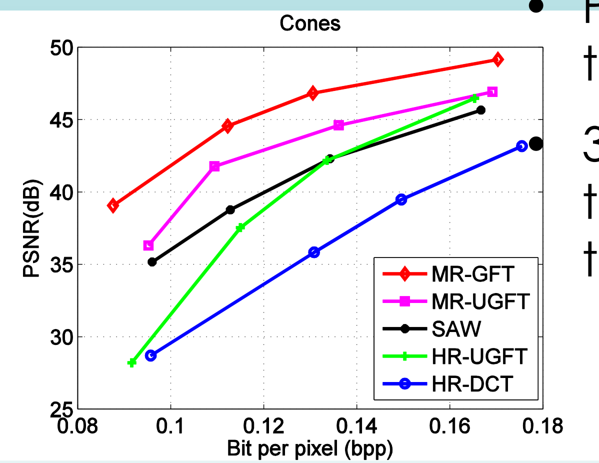
DCT

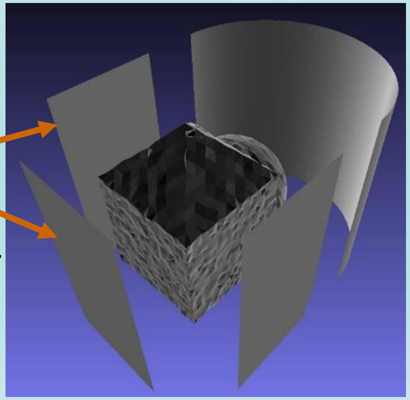


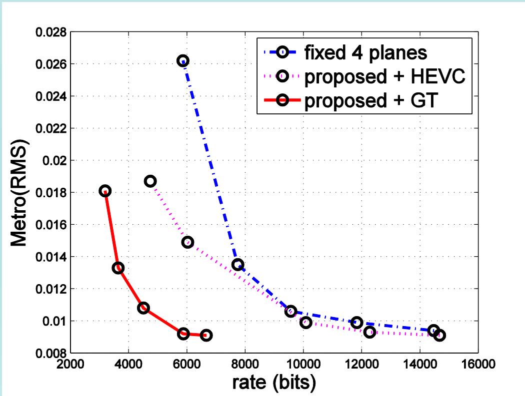
GFT

- **Goal:** reduce coding rate w/o distortion.
- **Graph Signal Processing**:
  - Graphs describe correlations bet’n neighboring pixels.
  - Transforms defined on graphs replace DCT.
- **Depth Image Compression**:
  - ***Graph Fourier Transform*** compactly describes PWS signals.
  - up to 68% bitrate reduction compared to HR-DCT [3].
- **Geometry Coding**:
  - Project 3D geometry to 2D image tiles.

35% rate saving for opt tiles w/ GFT over naïve tiles w/ DCT [4].





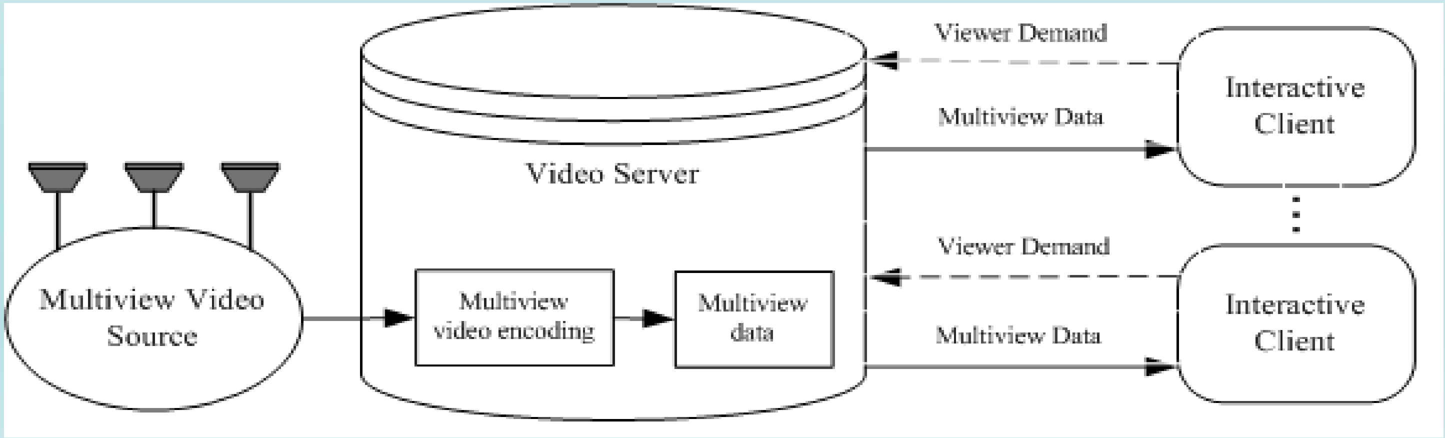


[3] W. Hu, G. Cheung, A. Ortega, O. Au, "**Multi-resolution Graph Fourier Transform for Compression of Piecewise Smooth Images**," *IEEE Trans on Image Proc*, Jan 2015.
[4] Y. Gao, G. Cheung, T. Maugey, P. Frossard, J. Liang, “**3D Geometry Representation using Multiview Coding of Image Tiles**,” *IEEE ICASSP*, May, 2014.

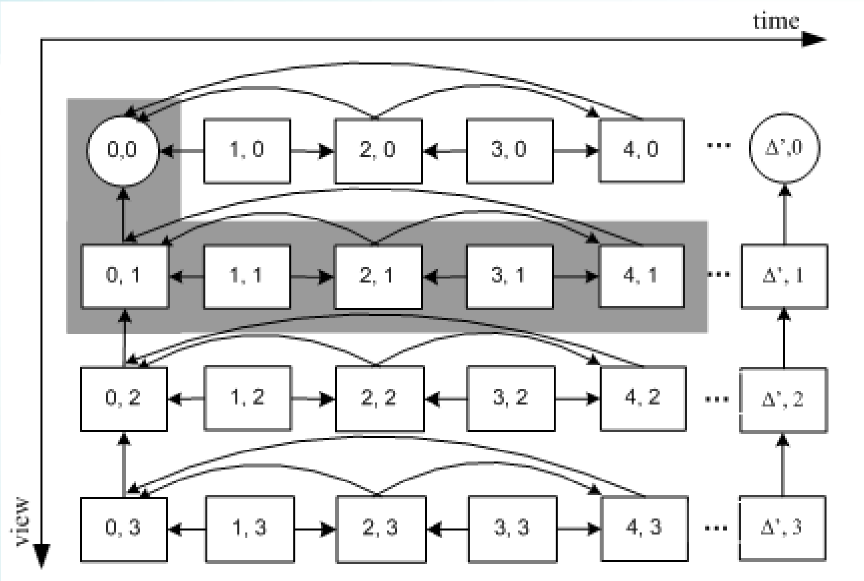
Microsoft Research

Background to Interactive Multiview Video Streaming

- **Interactive Multiview Video Streaming (IMVS)**
  - A client can periodically request one of many captured views, as video is played back in time.
  - To reduce transmission BW, ***transmit only views interactively selected by client***.
  - The encoding is done once at the server for a possibly large group of clients.



Background to Interactive Multiview Video Streaming



Multiview Video Coding (MVC) [5]

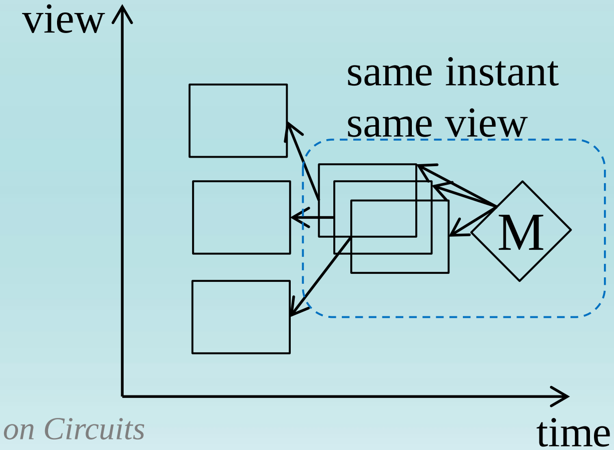
- Code frames of all views in RD manner.

Disadvantage: MVC not suitable for IMVS

- *Insufficient decoding flexibility* for interactive view-switching of 1 view.

Redundant frame structure for IMVS [6]

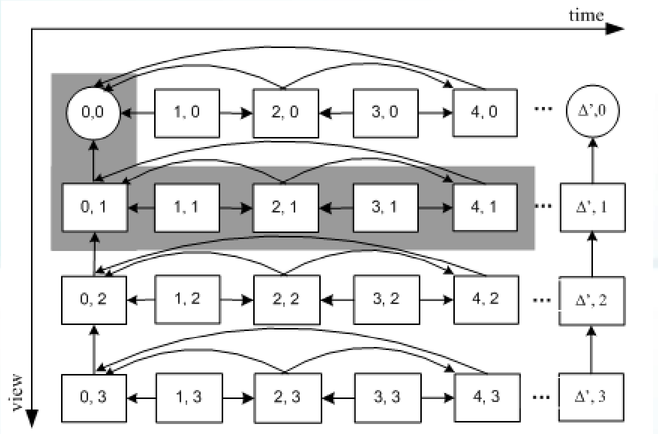
- *Multiple decoding paths to facilitate view-switching.*
- Distributed source coding (DSC) frames to merge decoding paths.



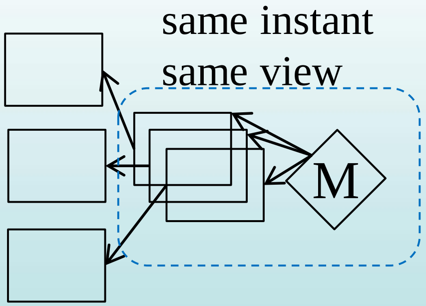
[5] P. Merkle et al., "**Efficient Prediction Structures for Multiview Video Coding**," *IEEE Transactions on Circuits and System for Video Technology*, vol.17, no.11, pp.1461-1473, November 2007.

[6] Cheung et al., "**Interactive Streaming of Stored Multiview Video using Redundant Frame Structures**," *IEEE Transactions on Image Processing*, vol.20, no.3, pp.744-761, March 2011.

Advanced Interactive 3D Video Streaming

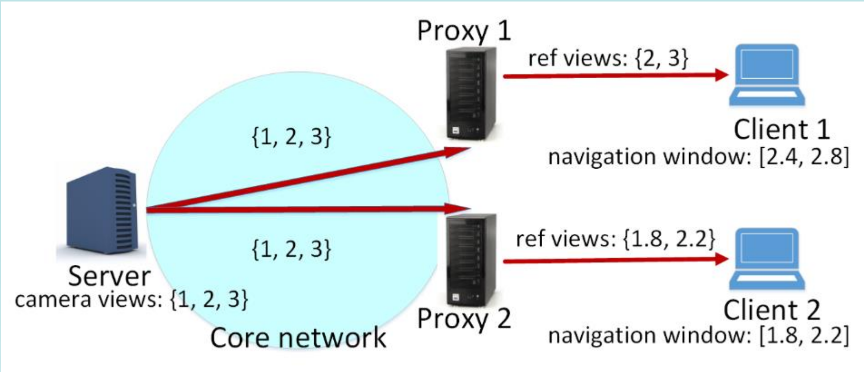


- **3D Video Compression with Flexible Decoding**:
    - **Prob**: diff. coding in video has dependency.
    - **Solution**: Interactive view-switching using *Merge frame* built from piecewise constant func.
    - Distributed Source Coding w/o channel codes! [7]



[7] W. Dai et al., "**Rate-distortion Optimized Merge Frame using Piecewise Constant Functions**," *IEEE International Conference on Image Processing*, Melbourne, Australia, September, 2013. (**Best student paper award**)

- **Interactive Navigation of High-dimen. Data**:
    1. View selection for observer group [8].
    2. Synthesize new reference views at proxies for delay-free navigation at client [9].



[8] D. Ren et al., "**Anchor View Allocation for Collaborative Free Viewpoint Video Streaming**," *IEEE Transactions on Multimedia*, vol.17, no.3, pp.307-322, March, 2015.

[9] L. Toni et al., "**In-Network View Re-Sampling for Interactive Free Viewpoint Video Streaming**," accepted to *IEEE International Conference on Image Processing*, Quebec City, September, 2015.

Outline

- Introduction
  - What is 3D imaging?
  - Key problems: 3D video coding & streaming
  - Other problems: denoising, interpolation, bit-depth enhancement, de-quantization
  - Applications: joint gaze correction / face beautification, sleep monitoring, heart rate estimation
- Fundamental of Graph Signal Processing (GSP)
  - Spectral Graph Theory
  - Graph-based Transform

Video Enhancement for Depth Camera

- **Problem**: Depth images from ToF camera are low-resolution, blurred, noisy

➢ **Setting**: Given a noisy, low-resolution depth map D_L and a registered noise-free, high-solution color image I

➔ Estimate D_H

*Courtesy of Prof. M. Do, UIUC, USA

Proposed Method: Weighted Mode Filtering

- Generating joint histogram [10]:
  - *g(p)*: color value at pixel p
  - *f(p)*: depth value at pixel p
  - $f_G(p)$: enhanced depth value at pixel p
  - G_I, G_S, G_r: Gaussian function

pixel *p*

color Gaussian

spatial Gaussian

err Gaussian

$$H_G(p,d) = \sum_{q \in N(p)} G_I(g(p)-g(q))G_S(p-q)G_r(d-f(q))$$

*d*th bin

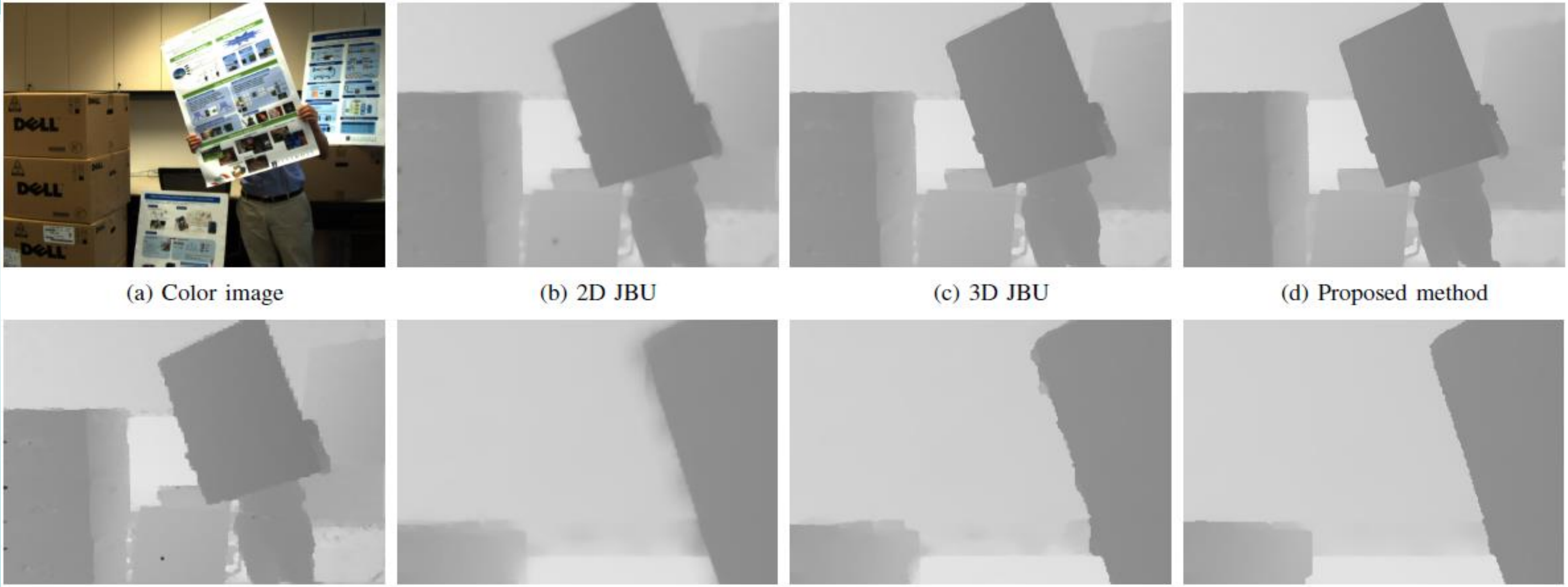
neighbors of pixel *p*

$$f_G(p) = \arg\max_d H_G(p,d)$$

[10] D. Min, J. Lu, and M. N. Do, "**Depth video enhancement based on weighted mode filtering**," *IEEE Trans. on Image Processing*, 2012.

*Courtesy of Prof. M. Do, UIUC, USA

Result Comparison

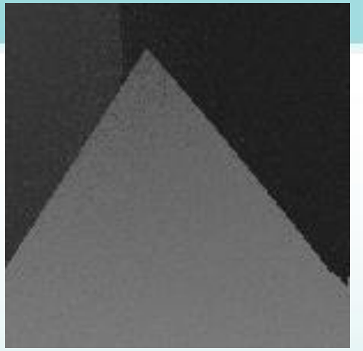


(a) Color image
(b) 2D JBU
(c) 3D JBU
(d) Proposed method
(e) Initial depth map
(f) Cropped image of (b)
(g) Cropped image of (c)
(h) Cropped image of (d)

Up-sampling results for low-quality depth image (from 'Mesa Imaging SR4000', 176x144) with corresponding color image (from 'Point Grey Flea', 1024x768).

*Courtesy of Prof. M. Do, UIUC, USA

Depth Image Denoising (w/o color)



bilateral filter

nonlocal GBT

		σ				
Image	Method	10	15	20	25	30
Cones	NLGBT	42.84	39.18	36.53	34.43	32.97
	BM3D	40.56	37.49	35.28	33.81	32.75
	NLM	39.42	35.84	34.64	32.95	31.62
	BF	33.34	30.53	27.96	26.03	24.21
Teddy	NLGBT	42.29	39.38	36.71	34.62	33.42
	BM3D	41.36	38.33	36.12	34.45	33.25
	NLM	39.57	36.24	35.17	33.49	32.22
	BF	34.49	31.25	28.87	26.50	23.70
Sawtooth	NLGBT	48.41	45.30	43.22	41.71	40.01
	BM3D	46.04	43.51	41.84	40.16	39.13
	NLM	41.14	37.56	38.28	36.54	35.01
	BF	36.36	30.99	27.62	25.38	23.61

- Denoising problem is ill-defined:

desired signal

observation

$$\mathbf{y} = \mathbf{x} + \mathbf{v}$$

noise

- Interpret graph-signal as instance of MRF, then $\|\Phi \mathbf{x}\|_0$ is ***graph sparsity prior*** [11]:

GFT

L0-norm

$$\min_{\Phi,\mathbf{x}} \|\mathbf{y} - \mathbf{x}\|_2 + \lambda \|\Phi \mathbf{x}\|_0$$

fidelity term

sparsity prior

- Interpret graph-signal as samples on high-dimen. manifold, then $x^T L x$ is ***graph-signal smoothness prior*** [12]:

$$\min_{\mathbf{L},\mathbf{x}} \|\mathbf{y} - \mathbf{x}\|_2 + \lambda \mathbf{x}^T \mathbf{L} \mathbf{x}$$

L = D - A

fidelity term

smoothness prior

Joint denoising / compression of multiview depth images.

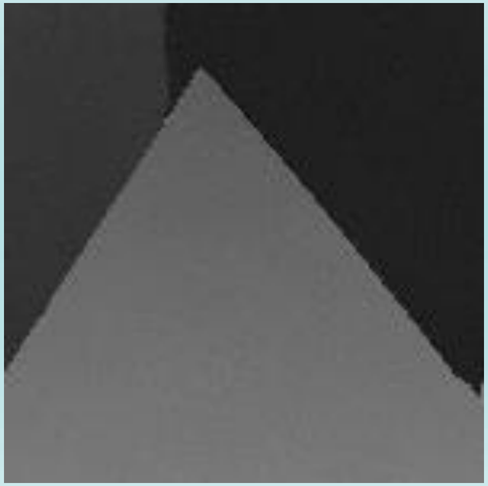
[13] W. Sun, G. Cheung, P. Chou, D. Florencio, C. Zhang, O. Au, "**Rate-constrained 3D Surface Estimation from Noise-corrupted Multiview Depth Videos**," *IEEE Transactions on Image Processing*, vol.23, no.7, pp.3138-3151, July 2014.

Depth Image Denoising (w/o color)

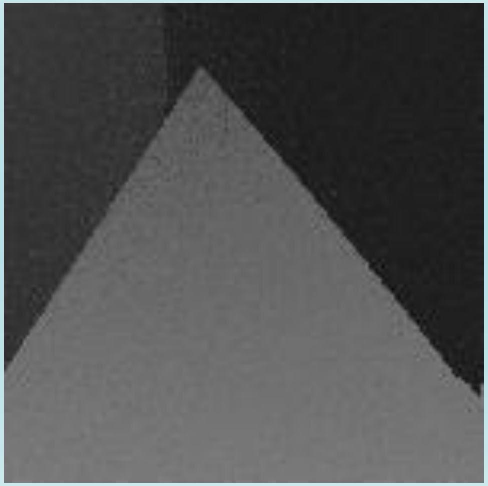
- **Experimental Setup:**
  - Test Middleburry depth maps: Sawtooth
  - Additive White Gaussian Noise (AWGN)
  - Compare to: Bilateral Filtering (BF), Non-Local Means Denoising, (NLM), Block-Matching 3D (BM3D).
- **Results**: 2.28dB improvement over BM3D.

NLGBT

BM3D

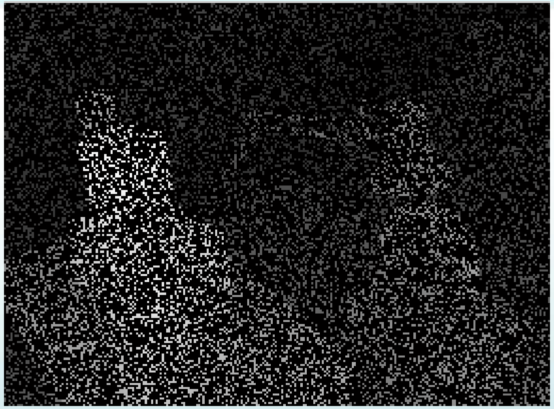


NLM

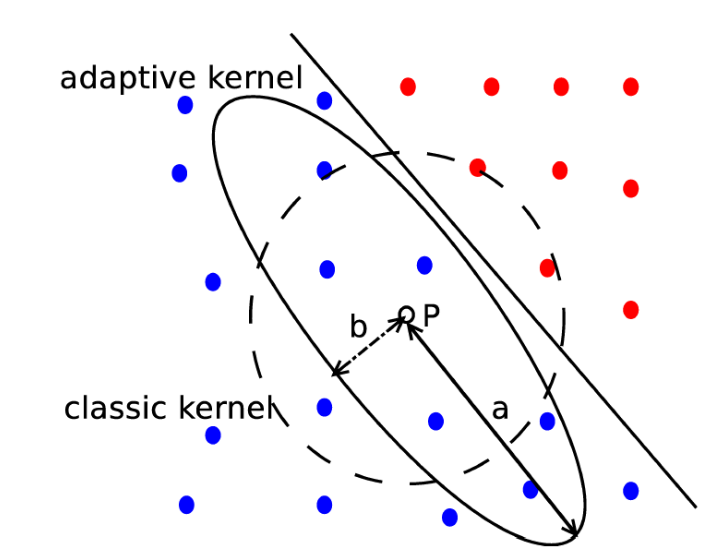


BF

Depth Image Interpolation (local)



Depth image missing 85% pixels



- **Problem**:
  - Fill holes in sparsely sampled depth images.
- **Idea**:*
  1. Find right graph for missing pixels.
     - Adaptive kernel
  2. Compute edge weights using initial values.

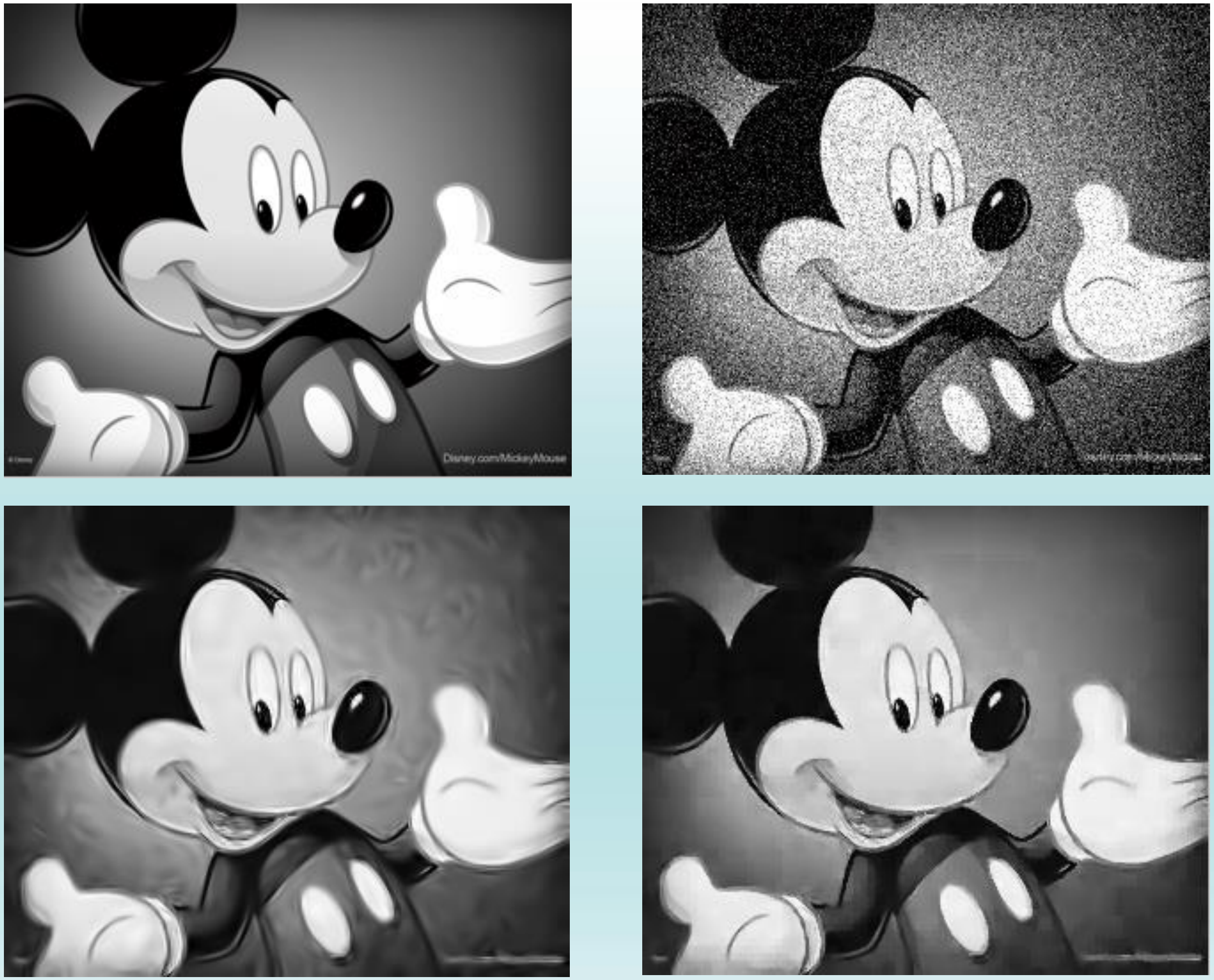
$$\mathrm{w}_{\mathrm{i,j}} = \exp\left\{ -\frac{\left| y_i - y_j \right|^2}{\sigma^2} \right\}$$

  3. Find smooth graph-signal given observations.

$$\min_{\mathrm{x}} \underbrace{\sum_i \left\| \mathrm{u}_i^t \mathrm{x} - y_i \right\|_2^2}_{\text{fidelity term}} + \underbrace{\lambda \mathrm{x}^t \mathrm{L} \mathrm{x}}_{\text{smoothness prior}}$$

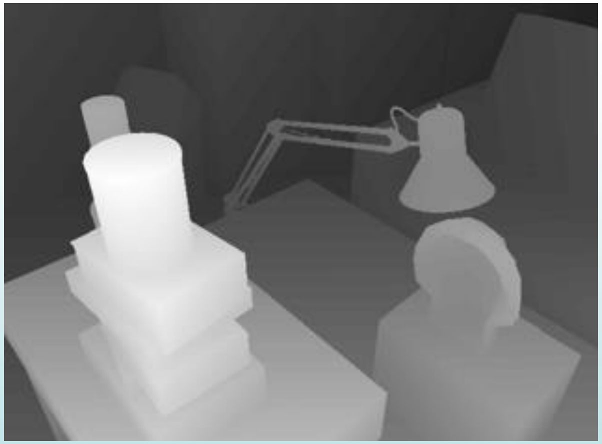
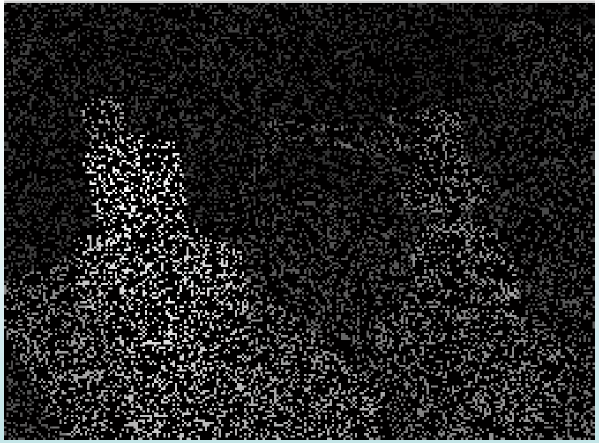
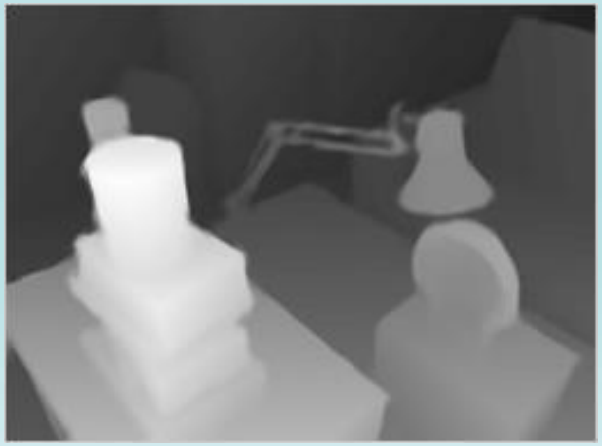
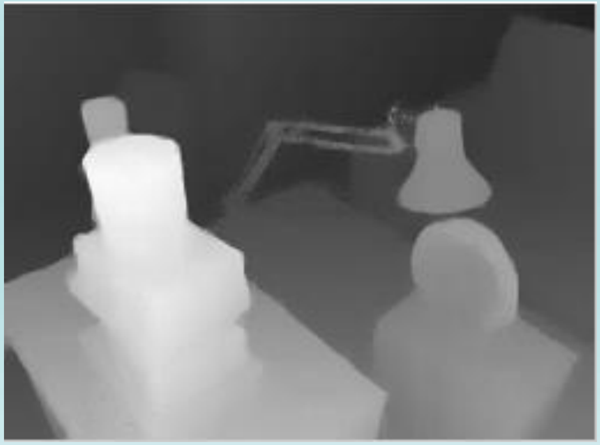
[14] Yu Mao, Gene Cheung, Yusheng Ji, "**Image Interpolation During DIBR View Synthesis Using Graph Fourier Transform**," *3DTV-Conference 2014*, Budapest, Hungary, July, 2014.

Denoising Results



Original	Noised (theta = 20)
LARK (26.99dB)	Our Proposal (27.36dB)

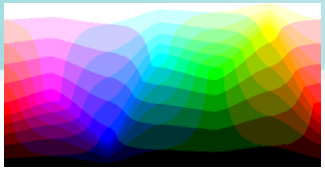
Interpolation Results

Original	Partial sample
LARK (34.82dB)	Our Proposal (35.31dB)

Image Bit-depth Enhancement

Problem:



$\mathbf{y}$ → Bit-depth Enhancement → $\hat{\mathbf{x}}$

low bit-depth (LBD) image **y**—a *quantized* version of underlying HBD image $\mathbf{x}$

an estimate of the original HBD image

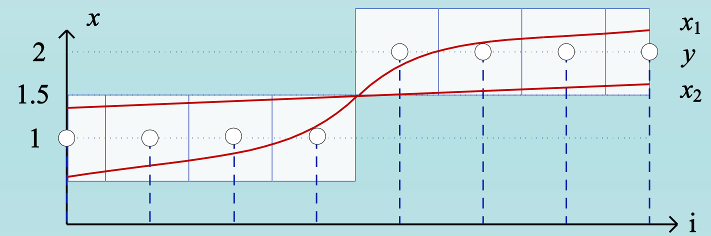
Objective: find $\hat{\mathrm{x}}$ that minimizes mean-squared-error (MSE),

$$\hat{\mathrm{x}}^{MMSE} = \arg\min_{\hat{\mathrm{x}}} \int \left\| \hat{\mathrm{x}} - \mathrm{x} \right\|_2^2 \; f\left(\mathrm{x} \mid \mathrm{y}\right) dx$$

squared err

posterior prob of HBD signal x given LBD signal y

Posterior: $f(\mathbf{x}|\mathbf{y}) \propto f(\mathbf{y}|\mathbf{x})\; f(\mathbf{x})$



Likelihood: equals to 1 iff x_i quantizes to y_i

$$f(\mathbf{y}|\mathbf{x}) = \begin{cases} 1, & \text{if } \mathrm{quant}(x_i) = y_i,\ \forall i \\ 0, & \text{otherwise} \end{cases}$$

Smoothness prior: HBD signal is smooth

Conventional smoothness (*e.g.*, total Variation* are signal-independent → ***over-smoothing***

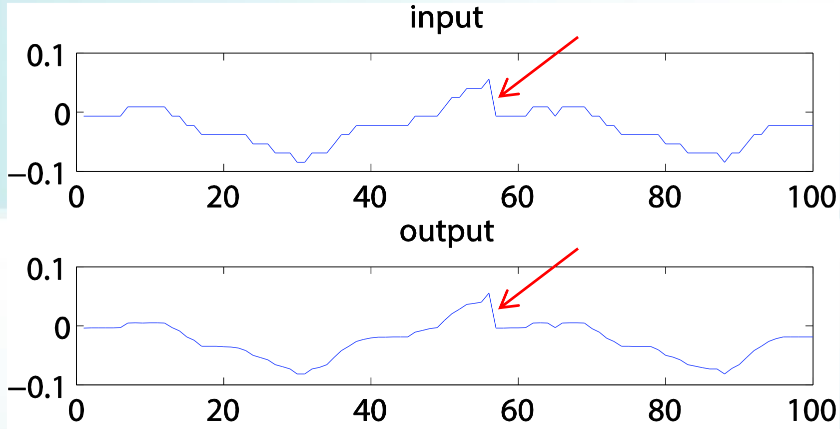
[15] Rudin et al, “Nonlinear total variation based noise removal algorithms”, Elsevier, 1992

Image Bit-depth Enhancement

Graph-signal smoothness prior

$$f(\mathbf{x}) = \frac{1}{K} \exp\left\{-\sigma \mathbf{x}^{\intercal} \mathbf{L} \mathbf{x}\right\}$$

L is the graph Laplacian matrix
describing inter-pixel similarities*



Reconstruct smooth signal without blurring edges

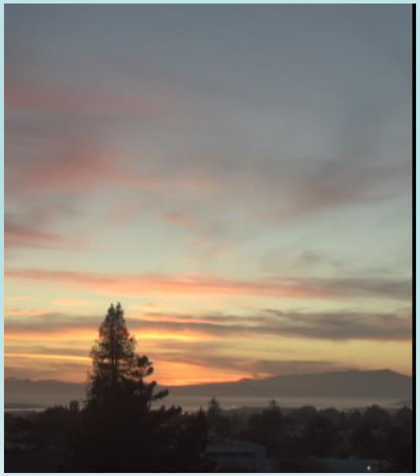
Proposed ACDC Algorithm:

- Compute edge weights from quantized signal.
- Compute MAP solution of AC signal.
- Compute MMSE solution of DC signal given AC signal.

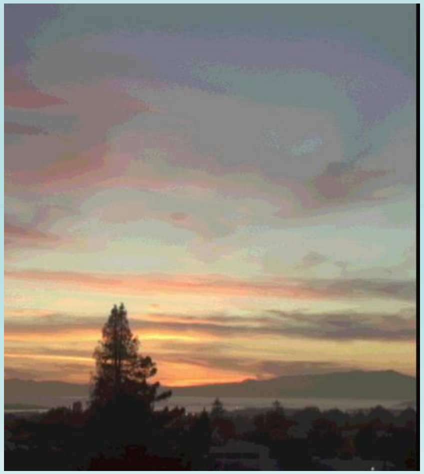
Input bit-depth

Table 1: Average PSNR results

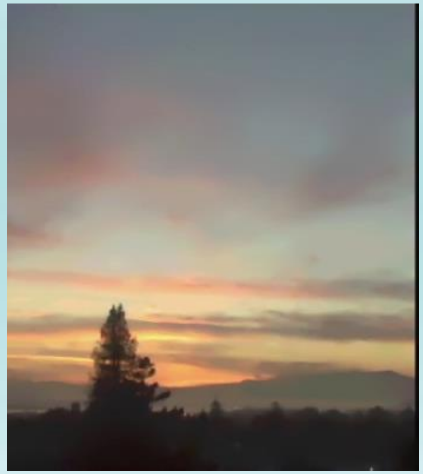
	ANC	SMOOTH	DECONT	INTERP	DMAP	ACDC
$b = 4$	34.80	35.71	35.17	35.25	35.29	**37.46**
$b = 6$	46.93	47.36	46.70	48.48	46.99	**49.50**
$b = 8$	58.92	57.30	57.64	59.43	58.55	**60.35**



Ground-truth



Input **y**



Our output

[16] P. Wan et al., "**Image Bit-depth Enhancement via Maximum-a-Posteriori Estimation of Graph AC Component**," *IEEE International Conference on Image Processing*, Paris, France, October, 2014. (Top 10% accepted paper recognition)

Soft-Decoding of Compressed Images

Problem

Encoder: quantization of DCT coefficients.

DCT transform

8x8 pixel block

$$q_i = round(\mathrm{Y}_i / \mathrm{Q}_i), \quad \mathbf{Y} = \mathbf{Ty}$$

DCT coefficients

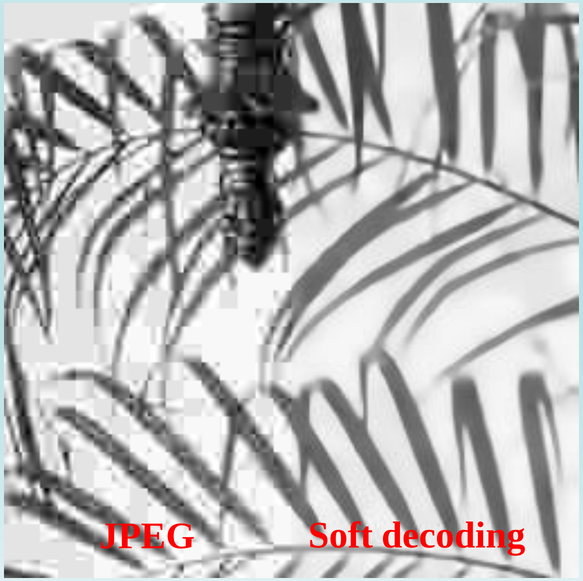
quantization parameter

Decoder: 64 quantization bin constraints / block.

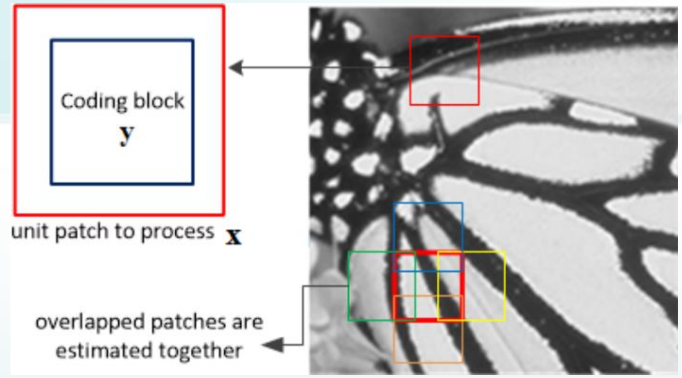
$$q_i \, \mathrm{Q}_i \leq \mathrm{Y}_i \leq (q_i + 1)\, \mathrm{Q}_i$$

Soft Decoding

- Find most probable signal within quantization bins with help of pre-determined signal priors.



Priors for Soft-Decoding of Compressed Images



Sparsity prior

$$\mathbf{x} = \boldsymbol{\Phi}\boldsymbol{\alpha} + \varepsilon,$$

noise term

sparse vector

dictionary

- Sparse linear combination of atoms approx. signal.
- Help restore high frequency content (image texture).

Graph-signal smoothness prior $f(\mathbf{x}) = \frac{1}{K}\exp\{-\sigma \mathbf{x}^{\intercal}\mathbf{L}\mathbf{x}\}$

- Promote signal-dependent smoothing (avoid over-smoothing of edges).
- Help restore sharp discontinuities (image structure).

Sparsity prior

Objective function: combine two priors with inter-patch consistency:

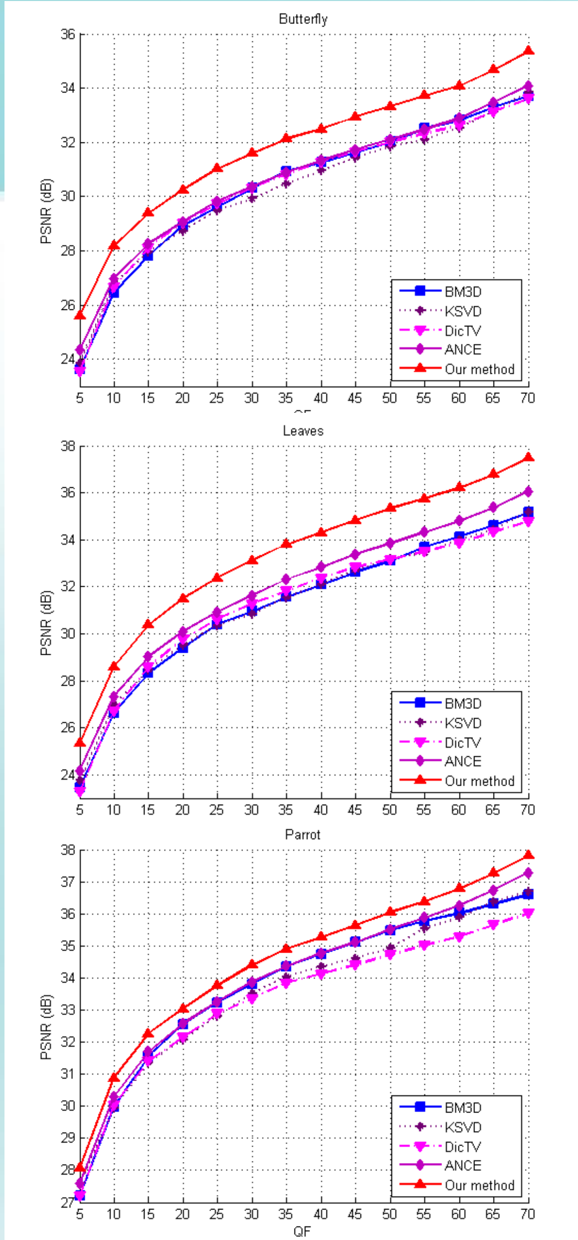
Graph-signal smoothness prior

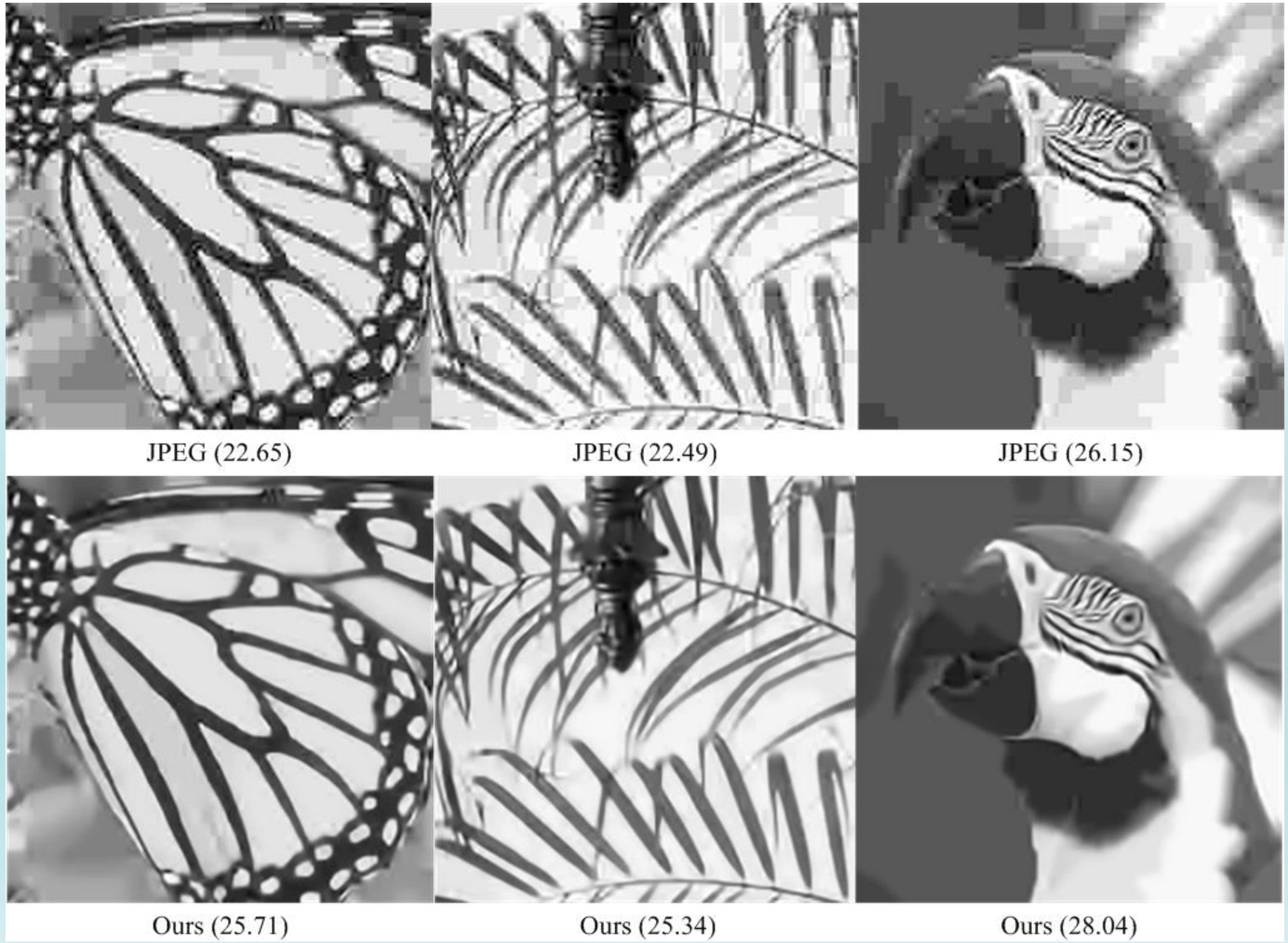
$$\arg\min_{\{\mathbf{x}_i,\,\boldsymbol{\alpha}_i\}} \sum_i \left\| \mathbf{x}_i - \boldsymbol{\Phi}\,\boldsymbol{\alpha}_i \right\|_2^2 + \lambda_1 \left\| \boldsymbol{\alpha}_i \right\|_1 + \lambda_2 \mathbf{x}_i^T \mathbf{L}_i \mathbf{x}_i$$

$$\text{s.t.,} q_{k,i} Q_k \le \mathbf{1}(k)^T \mathbf{TM}\mathbf{x}_i < (q_{k,i}+1) Q_k, \forall k$$ Quantization constraint

$$\sum_{j \in \mathrm{N}(i)} \left\| R_{i,j}\mathbf{x}_i - R_{j,i}\mathbf{x}_j \right\|_2^2 \le \tau \qquad \forall i.$$ Inter-patch consistency

Experimentation





JPEG (22.65) JPEG (22.49) JPEG (26.15)

Ours (25.71) Ours (25.34) Ours (28.04)

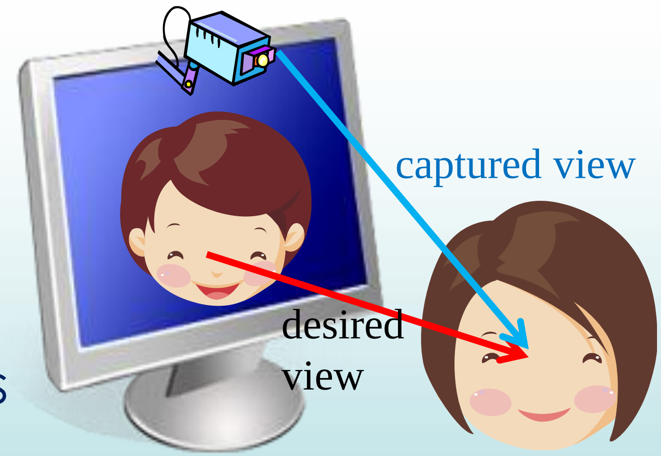
[17] Xianming Liu et al., "**Inter-Block Soft Decoding of JPEG Images with Sparsity and Graph-Signal Smoothness Priors**," accepted to *IEEE International Conference on Image Processing*, Quebec City, September, 2015.

Outline

- Introduction
  - What is 3D imaging?
  - Key problems: 3D video coding & streaming
  - Other problems: denoising, interpolation, bit-depth enhancement, de-quantization
  - Applications: joint gaze correction / face beautification, sleep monitoring, heart rate estimation
- Fundamental of Graph Signal Processing (GSP)
  - Spectral Graph Theory
  - Graph-based Transform

Gaze Correction / Face Beautification in Conference Video

Gaze mismatch problem: video conference subjects cannot see eye-to-eye.



Solution

1. Learn offline frontal faces of **beautiful people** from large databases of images (**big data**).
2. During conference, re-draw gaze-corrected faces & beautify facial components (eye, eyebrows).

- Handles “blinking”.

Enhanced Reality: looks real, but overcomes limitations of reality.

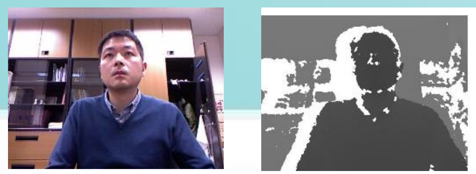
Results: natural looking facial video, w/o explicit 3D geometric info.

Note: 1 joint patent application w/ KDDI (submitted).

Designing The Future KDDI

*In collaboration with KDDI Research Labs, Saitama, Japan, and Harbin Institute of Technology, China.

System Overview

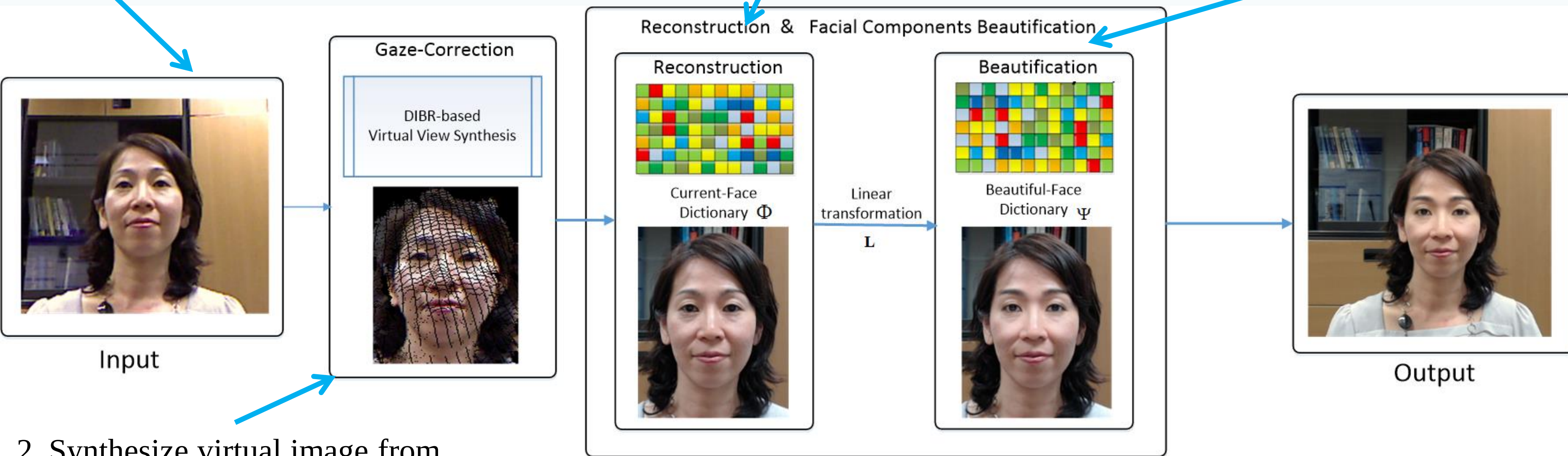


Microsoft Kinect

1. Capture texture & depth images.

3a. Repair face using 1st *patch*-based dictionary.

3b. Beautify facial components using 2nd *component*-based dictionary (eyes, eyebrows).



2. Synthesize virtual image from screen center viewpoint.

Dual sparse coding formulation

$$\min_{\alpha,\beta} \left\| x - \Phi\alpha \right\|_2^2 + \lambda_1 \left\| \alpha \right\|_0 + \mu \left\| S(\Phi\alpha) - S\big(W(\Phi\alpha, \Psi\beta)\big) \right\|_2^2 + \lambda_2 \left\| \beta \right\|_0^w$$

[18] X. Liu, G. Cheung, D. Zhai, D. Zhao, H. Sankoh, S. Naito, "**Joint Gaze-Correction and Beautification of DIBR-Synthesized Human Face via Dual Sparse Coding**," *IEEE International Conference on Image Processing*, Paris, France, October, 2014.

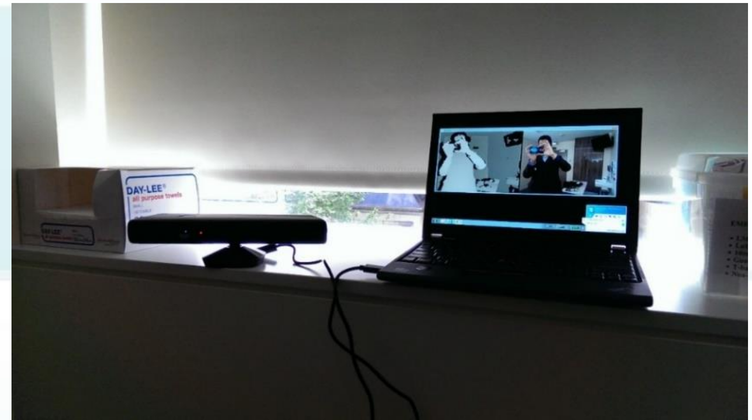
Gaze Correction / Face Beautification Image Results: no expression

Gaze Correction / Face Beautification Image Results: with expression

Gaze Correction / Face Beautification Video Results: talking gene

[19] X. Liu, G. Cheung, D. Zhai, D. Zhao, "**Sparsity-based Joing Gaze Correction & Face Beautification for Conference Video**," submitted to *IEEE International Conference on Visual Communication and Image Processing (VCIP)*, Singapore, December, 2015.

Sleep Monitoring via Depth Video



Problem: Detect ***sleep apnea*** non-intrusively.

Solution: Sleep monitoring via depth video analysis

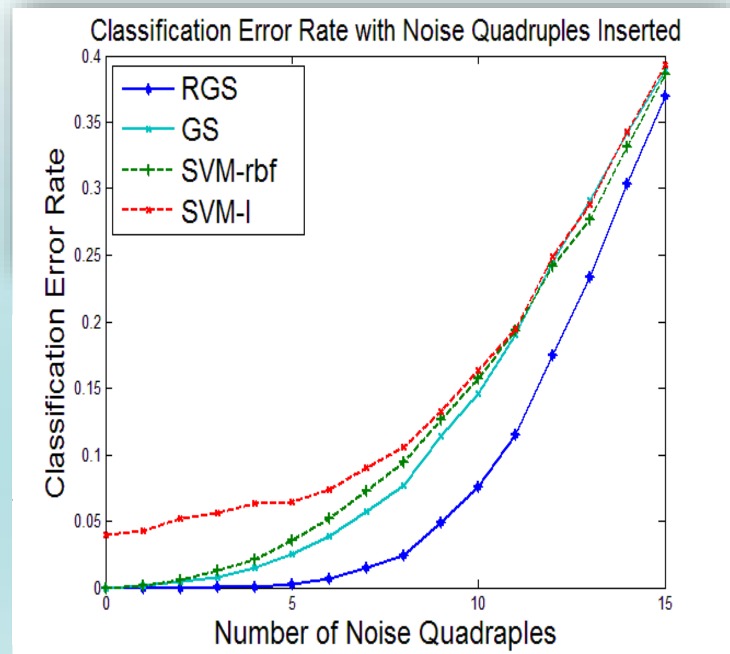
1. Efficient 8-bit depth video coding via H.264.
2. ***Temp***
   graph
3. ***Sleep***
   1. De
      mo
   2. ***Cl***
      gr

Results:

- 100%
  (same
- More

Our system: inexpensive, accurate, non-intrusive

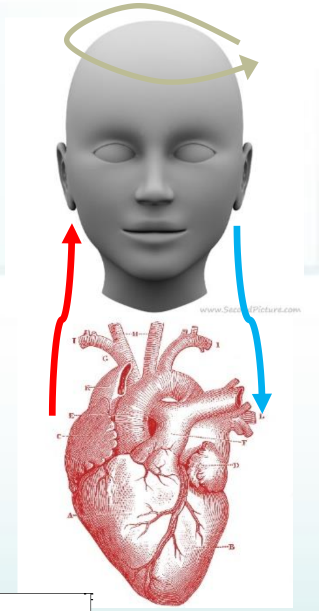
Philips PDx: accurate, intrusive



30 training quadruples

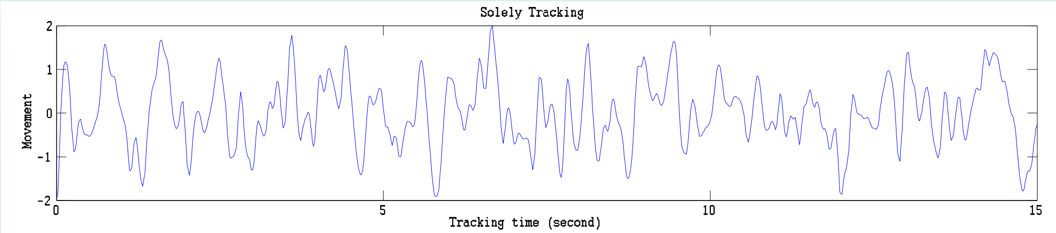
[20] C. Yang, Y. Mao, G. Cheung, V. Stankovic, K. Chan, "**Graph-based Depth Video Denoising and Event Detection for Sleep Monitoring**," *IEEE International Workshop on Multimedia Signal Processing*, Jakarta, Indonesia, September, 2014.

Heart Rate Detection from Depth Video

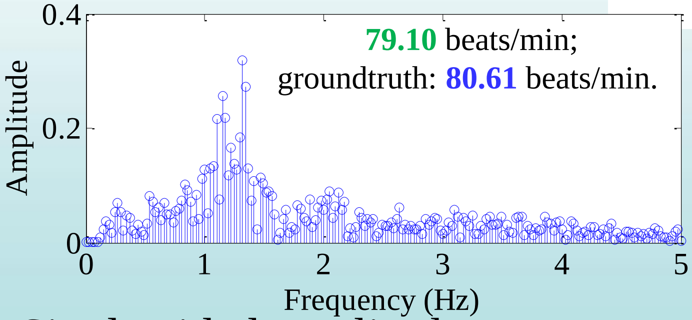


Heart rate detection from depth video:

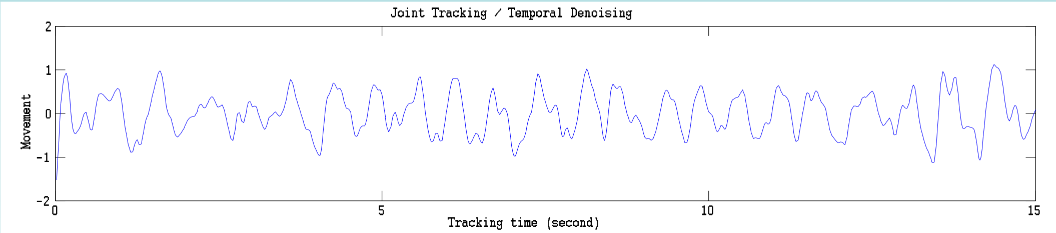
- Head oscillates slightly as blood circulates.
- **Methodology**: joint bit-depth enhancement, denoising via GSP, motion tracking, PCA, identify largest frequency.



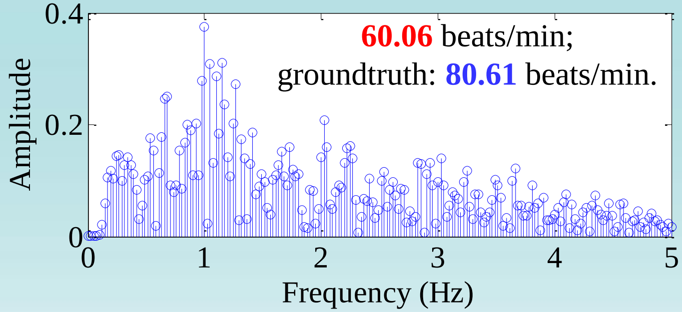
noisy temporal data



Single-sided amplitude spectrum of (a).



denoised temporal data



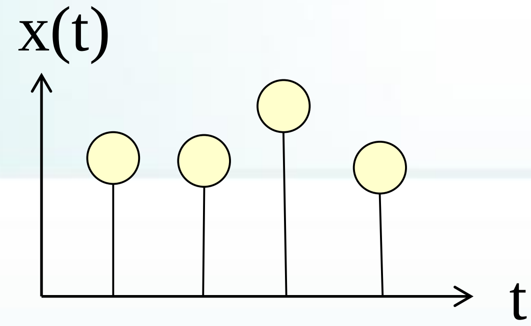
Single-sided amplitude spectrum of (b).

[21] C. Yang, G. Cheung, V. Stankovic, "**Estimating Heart Rate via Depth Video Motion Tracking**," *IEEE International Conference on Multimedia and Expo*, Torino, Italy, June, 2015 (**best paper candidate**).

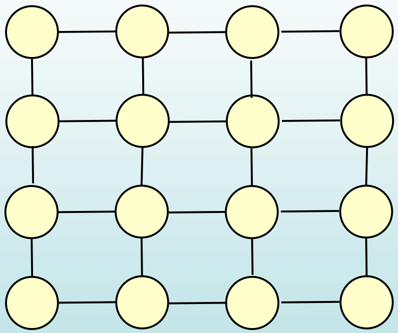
Outline

- Introduction
  - What is 3D imaging?
  - Key problems: 3D video coding & streaming
  - Other problems: denoising, interpolation, bit-depth enhancement, de-quantization
  - Applications: joint gaze correction / face beautification, sleep monitoring, heart rate estimation
- Fundamental of Graph Signal Processing (GSP)
  - Spectral Graph Theory
  - Graph-based Transform

Traditional Signal Processing



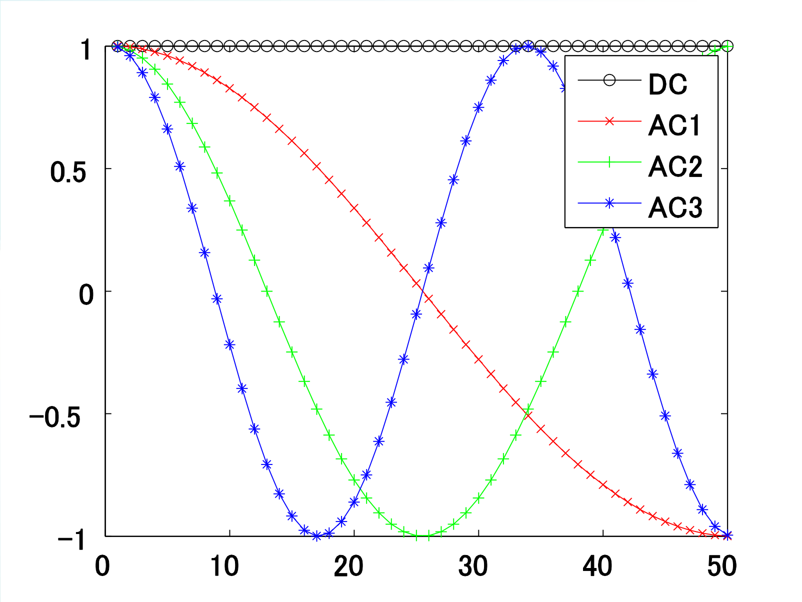
- Traditional discrete signals live on regular data kernels (**unstructured**).
  - Ex.1: audio / music / speech on regularly sampled timeline.
  - Ex.2: image on 2D grid.
  - Ex.3: video on 3D grid.



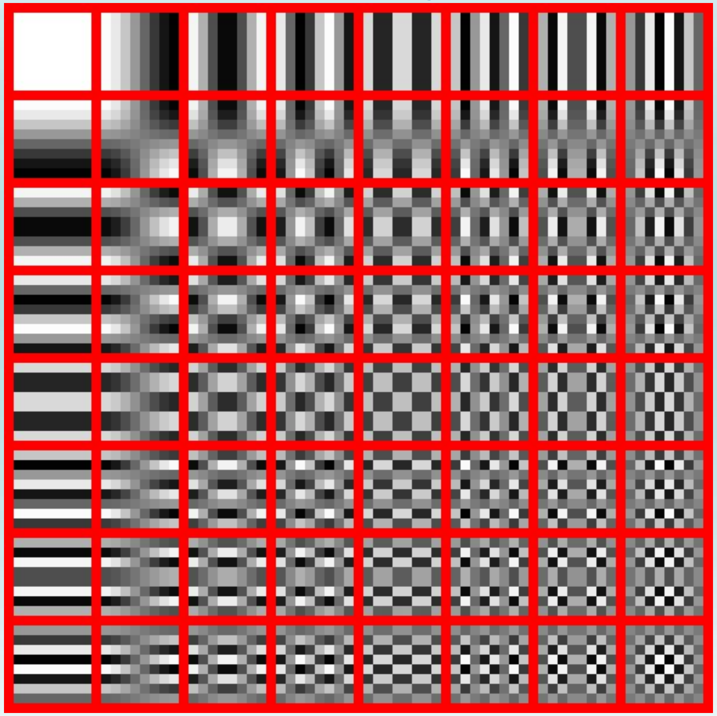
- Wealth of SP tools (transforms, wavelets, dictionaries, etc) for tasks such as:
  - compression, denoising, restoration, classification.

Smoothness of Signals

- Many known signals are **smooth**.
- Notion of *frequency*, *bandlimited*.
- 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.



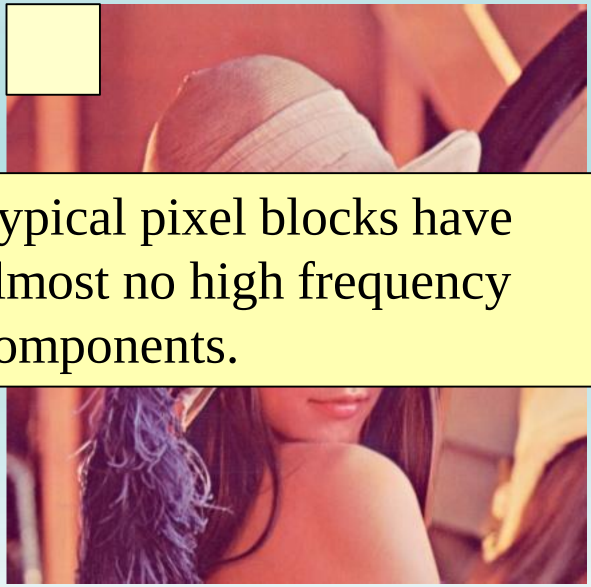
$$\mathrm{a} = \Phi\, \mathrm{x}$$

desired signal

transform

transform coeff.

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



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*

- Many known signals are **sparse**.

(over-complete) dictionary

desired signal

(sparse) code vector

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

covariance matrix

$$\mathrm{E}\left[\mathbf{x}\mathbf{x}^T\right] = \mathbf{C}$$

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

eigen-matrix

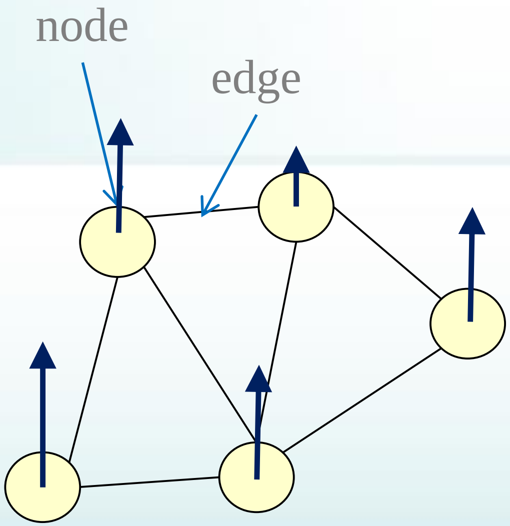
diagonal matrix of eigen-values

$$\mathbf{C}\mathbf{Q} = \mathbf{Q}\Lambda$$

- **DCT** approximates KLT*.

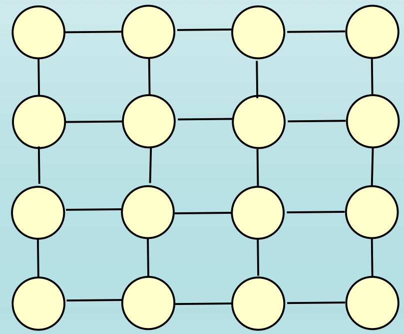
*Assuming simple AR process with same innovation variance.

Graph Signal Processing



- Signals live on graph.
  - Graph is sets 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.



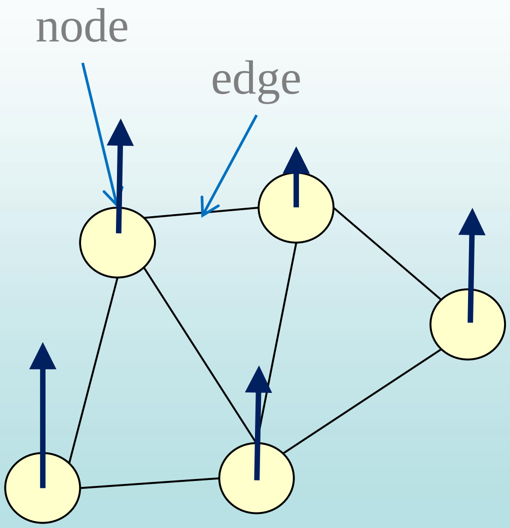
Graph Signal Processing (GSP) addresses the problem of processing signals that live on graphs.

[22] 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

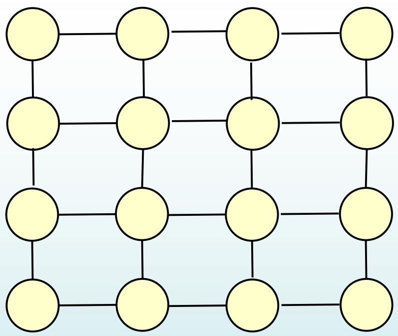
Research questions:

- **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.

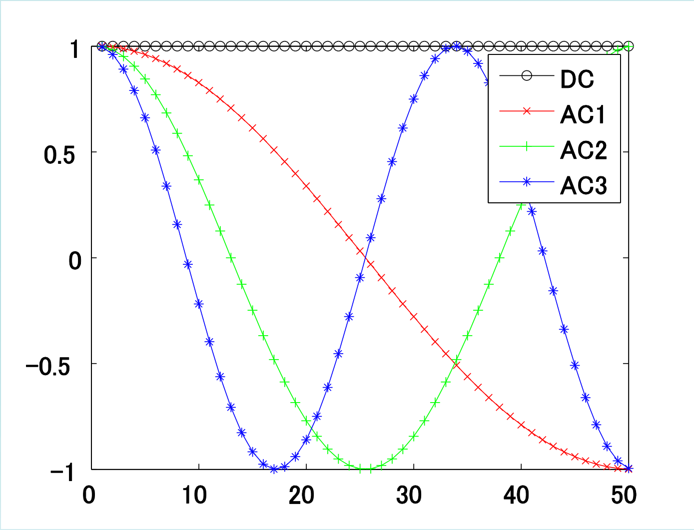


Graph Fourier Transform (GFT) for Graph-signals

Graph Fourier Transform:



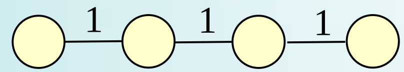
- *Signal-adaptive* transform:
  1. If two connected pixels are "similar", then edge weight is large → adjacency matrix A.
  2. Compute ***graph Laplacian*** L = D-A.
  3. Perform eigen-decomposition on L for GFT.



$$x = \sum_i a_i \varphi_i$$

- **Intuition**: Embed geometric structure of signal as edge weights in graph.

Facts of Graph Laplacian & GFT



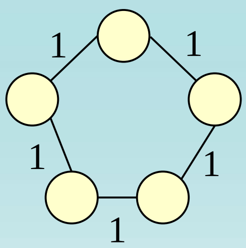
$$L = \begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 1 \end{bmatrix}$$

L for 4-node un-weighted line graph

- L is a high-pass filter.
- $\mathrm{x}^T \mathrm{L} \mathrm{x}$ is one measure of variation in signal.

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

- L is *positive semi-definite*; eigenvalues λ's ≥ 0 → eigenvalues are ***graph frequencies***.
- L = D – A; λ = 0 must be eigenvalue w/ vector $[1 \ \ldots \ 1]^T$.
- Use eigenvectors for spectral decomposition of signal.
  - GFT defaults to *DCT* for un-weighted connected line.
  - GFT defaults to *DFT* for un-weighted connected circle.



Usage Example: first non-zero eigenvalue → *spectral clustering* (Shi & Malik'00).

Summary

- 3D Imaging: texture-plus-depth image capture.
  - Immersive Visual Communication: video conf., mixed reality
  - 3D Data sensing & analysis: health monitoring
- Key problems: 3D video coding & streaming
- Other problems: denoising, interpolation, bit-depth enhancement, de-quantization
- Applications: joint gaze correction / face beautification, sleep monitoring, heart rate detection
- Fundamentals of GSP:
  - Graph Fourier Transform