Gene Cheung

National Institute of Informatics

15th July, 2016

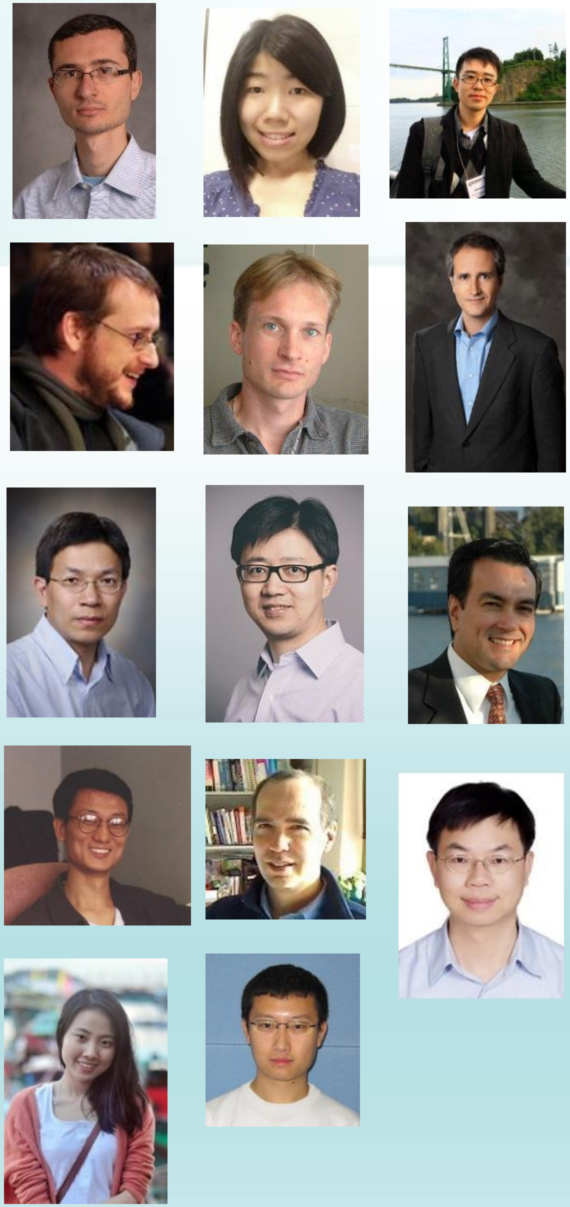
NII
大学共同利用機関法人 情報・システム研究機構
国立情報学研究所
National Institute of Informatics

Interactive Media Streaming Applications Using Merge Frames

Acknowledgement

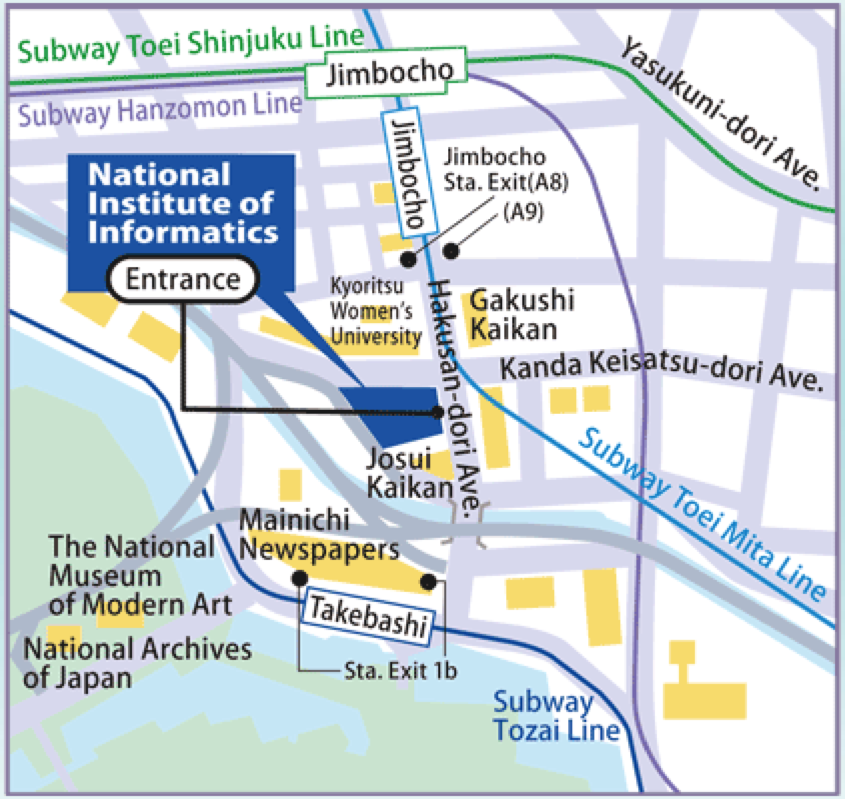
Collaborators:

- 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)
- L. Fang (USTC, China)
- C.-W. Lin (National Tsing Hua University, Taiwan)



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.

Introduction to APSIPA and APSIPA DL

APSIPA Mission: To promote broad spectrum of research and education activities in signal and information processing in Asia Pacific

APSIPA Conferences: ASPIPA Annual Summit and Conference

APSIPA Publications: Transactions on Signal and Information Processing in partnership with Cambridge Journals since 2012; APSIPA Newsletters

APSIPA Social Network: To link members together and to disseminate valuable information more effectively

APSIPA Distinguished Lectures: An APSIPA educational initiative to reach out to the community

Outline

- What is interactive media navigation?
    - e.g. Multiview / free-viewpoint video
- Merge frame for interactive media navigation
    - Previous works
    - Merge frame / block overview
    - Fixed target merging
    - Optimized target merging
- Interactive Light Field Streaming (ILFS)

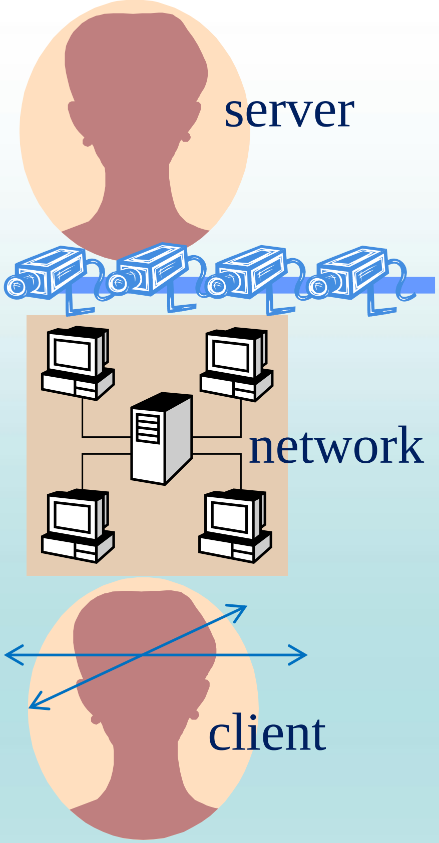
Wei Dai, Gene Cheung, Ngai-Man Cheung, Antonio Ortega, Oscar Au, "**Merge Frame Design for Video Stream Switching using Piecewise Constant Functions**," *IEEE Transactions on Image Processing*, vol. 25, no.8, August 2016

B. Motz, G. Cheung, A. Ortega, “**Redundant Frame Structure using M-frame for Interactive Light Field Streaming**,” (accepted to) *IEEE International Conference on Image Processing*, Phoenix, USA, September, 2016

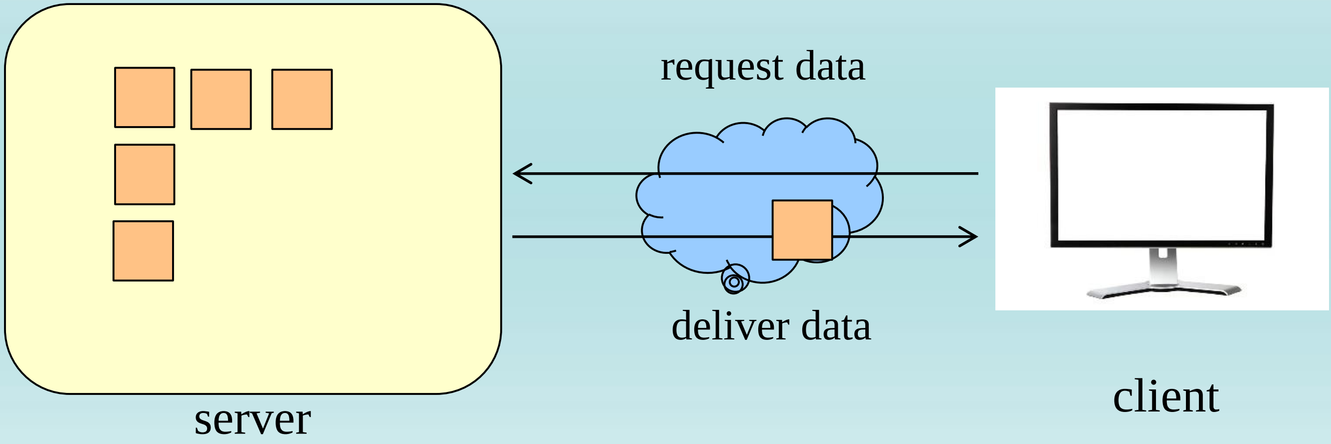
Outline

- What is interactive media navigation?
    - e.g. Multiview / free-viewpoint video
- Merge frame for interactive media navigation
    - Previous works
    - Merge frame / block overview
    - Fixed target merging
    - Optimized target merging
- Interactive Light Field Streaming (ILFS)

What is interactive media navigation / streaming?



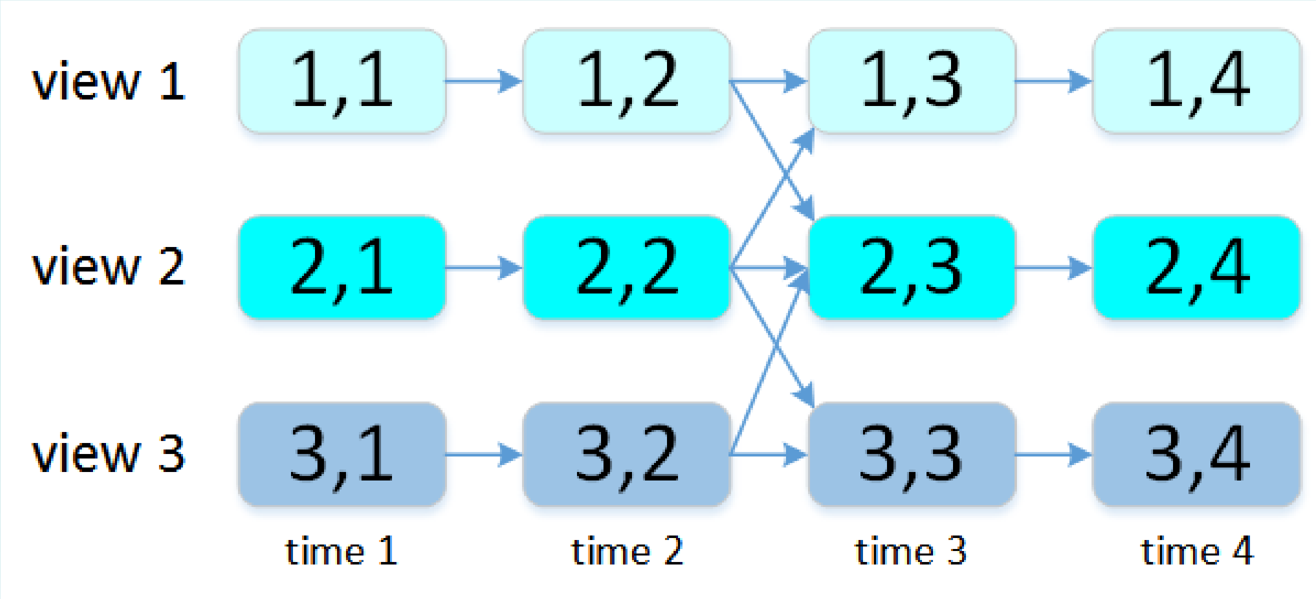
- **Server:** a very large correlated media data set.
  - *e.g.*, multiview video, light field data, etc.
- **Client:** can observe only small data subset at a time.
- **Network**: cannot deliver whole dataset before start of navigation.
- **Interactive navigation**: client requests data, server sends data. Repeat.



G. Cheung, A. Ortega, N.-M. Cheung, B. Girod, "**On Media Data Structures for Interactive Streaming in Immersive Applications**," in *SPIE Visual Communications and Image Processing*, Huang Shan, China, July, 2010.

Interactive Multiview Video Streaming (IMVS)

- **Server:** multiple views of same video captured synchronously in time.
- **Client:** can observe only 1 view at a time.
- **Interactive navigation**:
  - Client plays back video in time uninterrupted.
  - Client requests view, server sends view. Repeat.



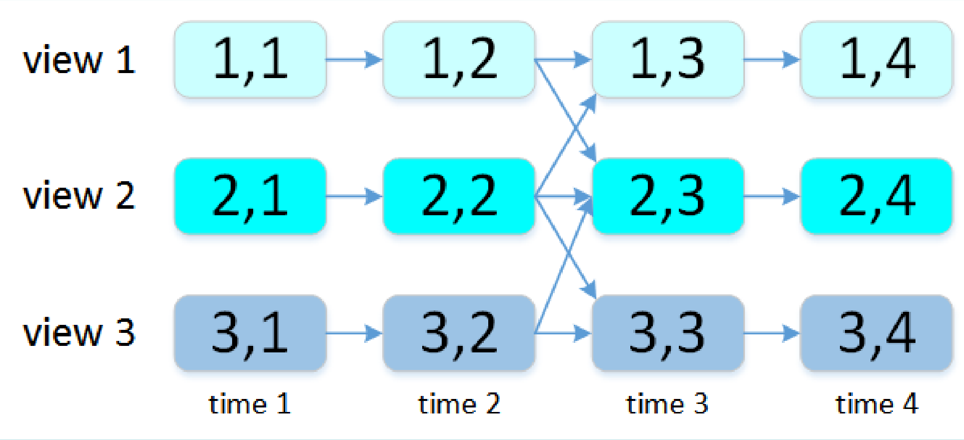
G. Cheung, A. Ortega, N.-M. Cheung, "**Interactive Streaming of Stored Multiview Video using Redundant Frame Structures**," *IEEE Transactions on Image Processing*, vol.20, no.3, pp.744-761, March 2011.

Outline

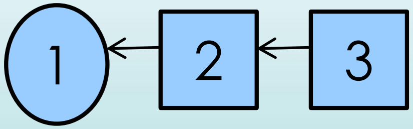
- What is interactive media navigation?
    - e.g. Multiview / free-viewpoint video
- Merge frame for interactive media navigation
    - Previous works
    - Merge frame / block overview
    - Fixed target merging
    - Optimized target merging
- Interactive Light Field Streaming (ILFS)

Merge Frame for Media Navigation: conflicting coding requirements

- Inherent tension between coding efficiency & flexible decoding.



- ***Differential coding*** assumes ***single*** order of frame decoding.



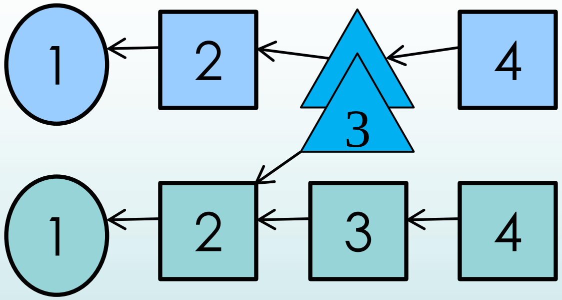
- ***Flexible decoding*** assumes ***several*** orders (paths) of frame decoding.

- **Other examples**:

Research Question: How to enable flexible decoding *without* great sacrifice of coding performance?

Merge Frame for Media Navigation: previous works 1

- **SP frames** (H.264 extended profile):
  - ***Primary SP-frame***: motion prediction + extra quantization. (small).
  - ***Secondary SP-frame***: motion prediction + lossless encoding. (large).
- **Pros**: small primary SP-frame.
- **Cons**:
  - very large secondary SP-frames.
  - As many secondary SP-frames as decoding paths.

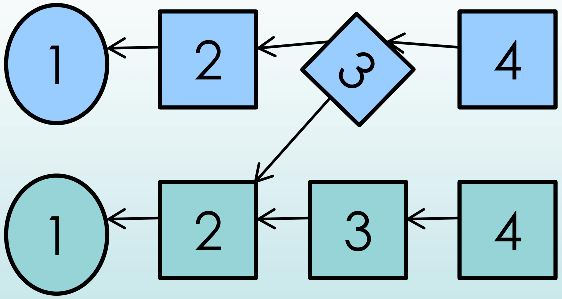


M. Karczewicz and R. Kurceren, "**The SP- and SI-frames design for H.264/AVC**," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no.7, July 2003, pp. 637–644.

X. Sun, F. Wu, S. Li, G. Shen, and W. Gao, "**Drift-free switching of compressed video bitstreams at predictive frames**," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no.5, May 2006, pp. 565–576.

Merge Frame for Media Navigation: previous works 2

- **DSC frames**:
  - ***Key Idea***: treat merging as *noise removal*.
  - Divide **side information** (SI) frames into block, perform DCT, quantization.
  - Examine *bit-planes* of quantized coefficients.
    - If bit-planes different from target, ***channel coding*** to "denoise" SI bit-planes to target bit-planes.



- **Pros**: one merge frame for many decoding paths.
- **Cons**:
  - Bit-plane / channel coding are complex.
  - Channel coding works well only for *average statistics*.

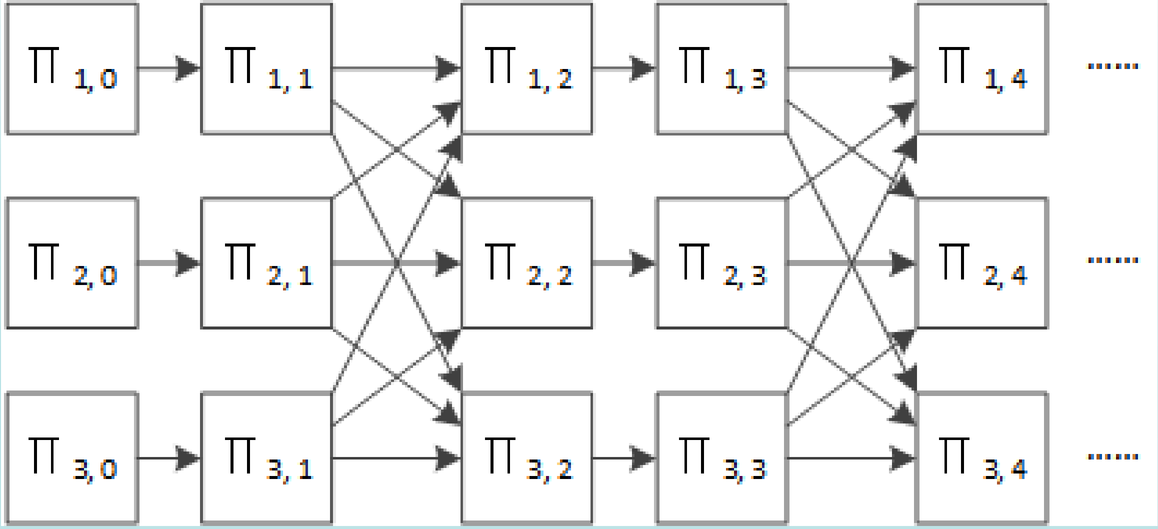
P. Ramanathan, M. Kalman, and B. Girod, "**Rate-distortion optimized interactive light field streaming**," in *IEEE Transactions on Multimedia*, vol. 9, no.4, June 2007, pp. 813–825.

N.-M. Cheung, A. Ortega, and G. Cheung, "**Distributed source coding techniques for interactive multiview video streaming**," in *27th Picture Coding Symposium*, Chicago, IL, May 2009.

Merge Frame for Media Navigation: definition

- **Interactive Video Stream Switching (IVSS)**
  - Multiple ***related*** pre-encoded video streams.
  - Designated ***switching points*** to switch from one to another.

- Picture Interactive Graph
  - Dynamic View Switching: switch to neighboring view of next time instant.
  - No loops in PIG.
  - *Optimized target merging*.



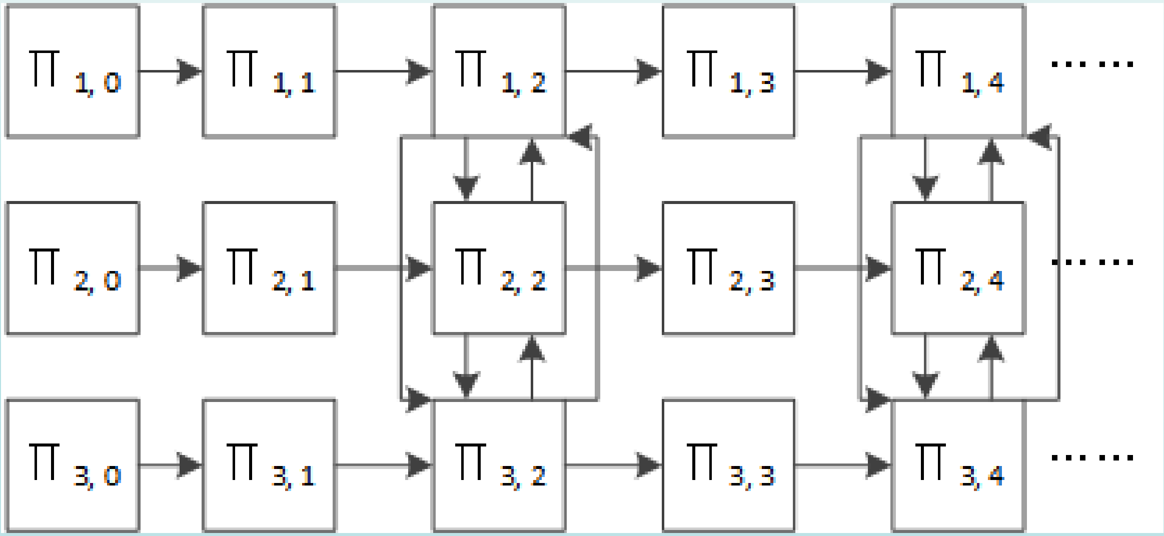
W. Dai, G. Cheung, N.-M. Cheung, A. Ortega, O. Au, "**Rate-distortion Optimized Merge Frame using Piecewise Constant Functions**," *IEEE International Conference on Image Processing*, Melbourne, Australia, September, 2013.

Wei Dai, Gene Cheung, Ngai-Man Cheung, Antonio Ortega, Oscar Au, "**Merge Frame Design for Video Stream Switching using Piecewise Constant Functions**," *IEEE Transactions on Image Processing*, vol. 25, no.8, August 2016

Merge Frame for Media Navigation: definition

- **Interactive Video Stream Switching (IVSS)**
  - Multiple ***related*** pre-encoded video streams.
  - Designated ***switching points*** to switch from one to another.

- Picture Interactive Graph
  - Static View Switching: switch to neighboring view of same time instant.
  - **Loops** in PIG.
  - *Fixed target merging*.

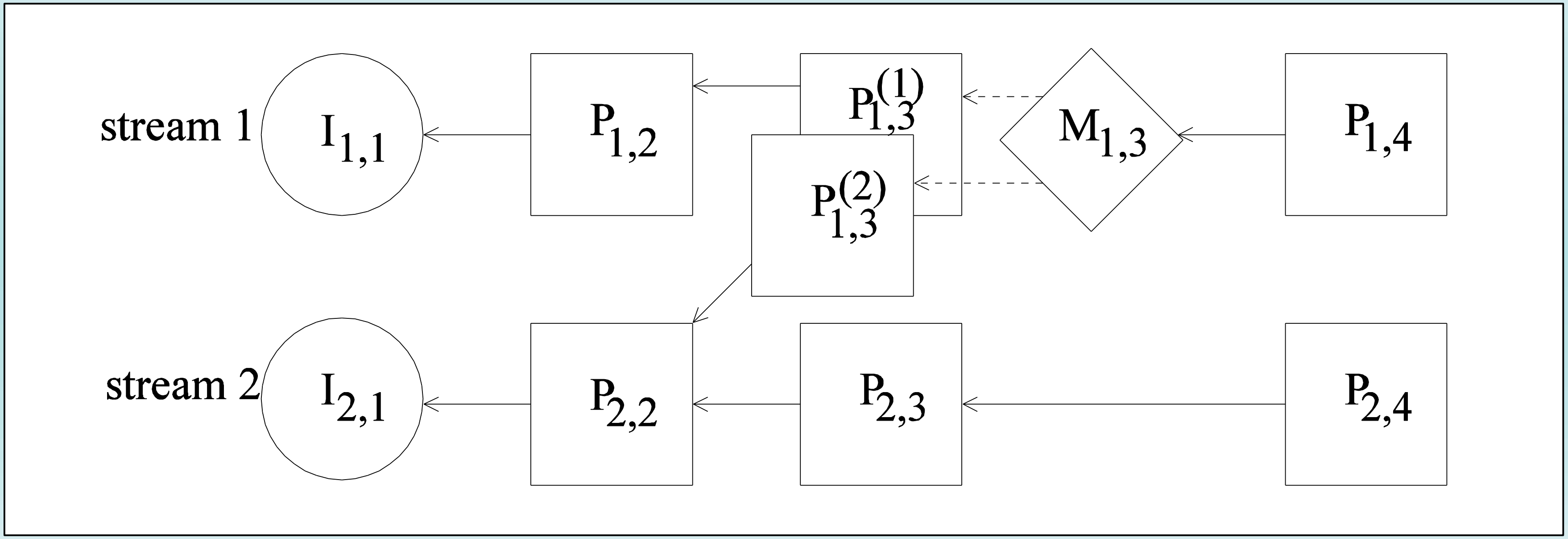


J.-G. Lou, H. Cai, and J. Li, “**A real-time interactive multi-view video system**,” in *ACM International Conference on Multimedia*, Singapore, November 2005.

N.-M. Cheung and A. Ortega, “**Compression algorithms for flexible video decoding**,” in *IS&T/SPIE Visual Communications and Image Processing (VCIP’08)*, San Jose, CA, January 2008.

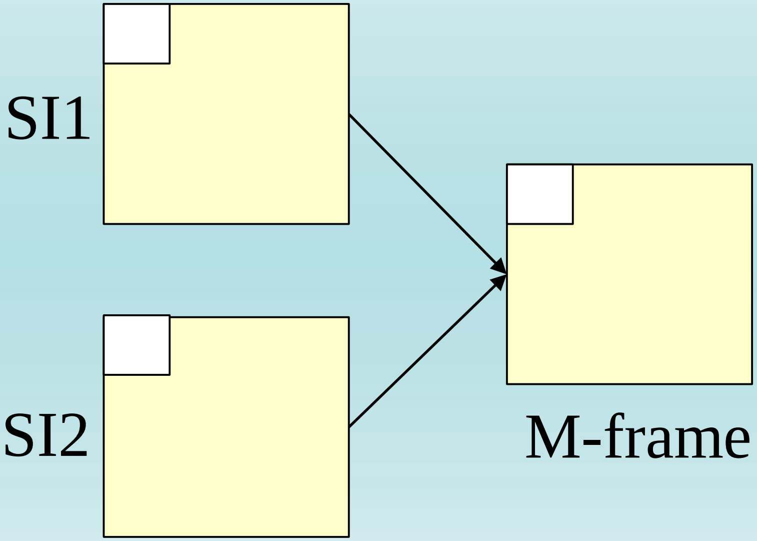
Merge Frame for Media Navigation: framework

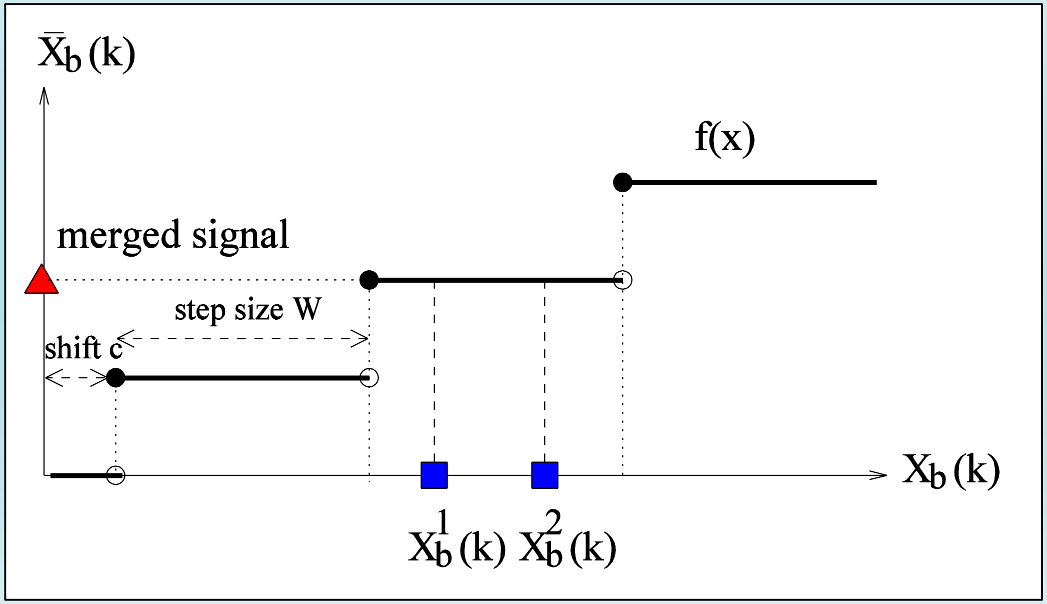
- **Switching Mechanism**
  - **Side Information (SI) frame**: P-frame predicted from diff. streams.
  - **Merge frame:** merge diff. among SI frames into same frame.
  - **Interactive Transmission:** transmit one SI frame + merge frame according to chosen decoding path.



Merge Frame for Media Navigation: merge frame (M-frame) overview

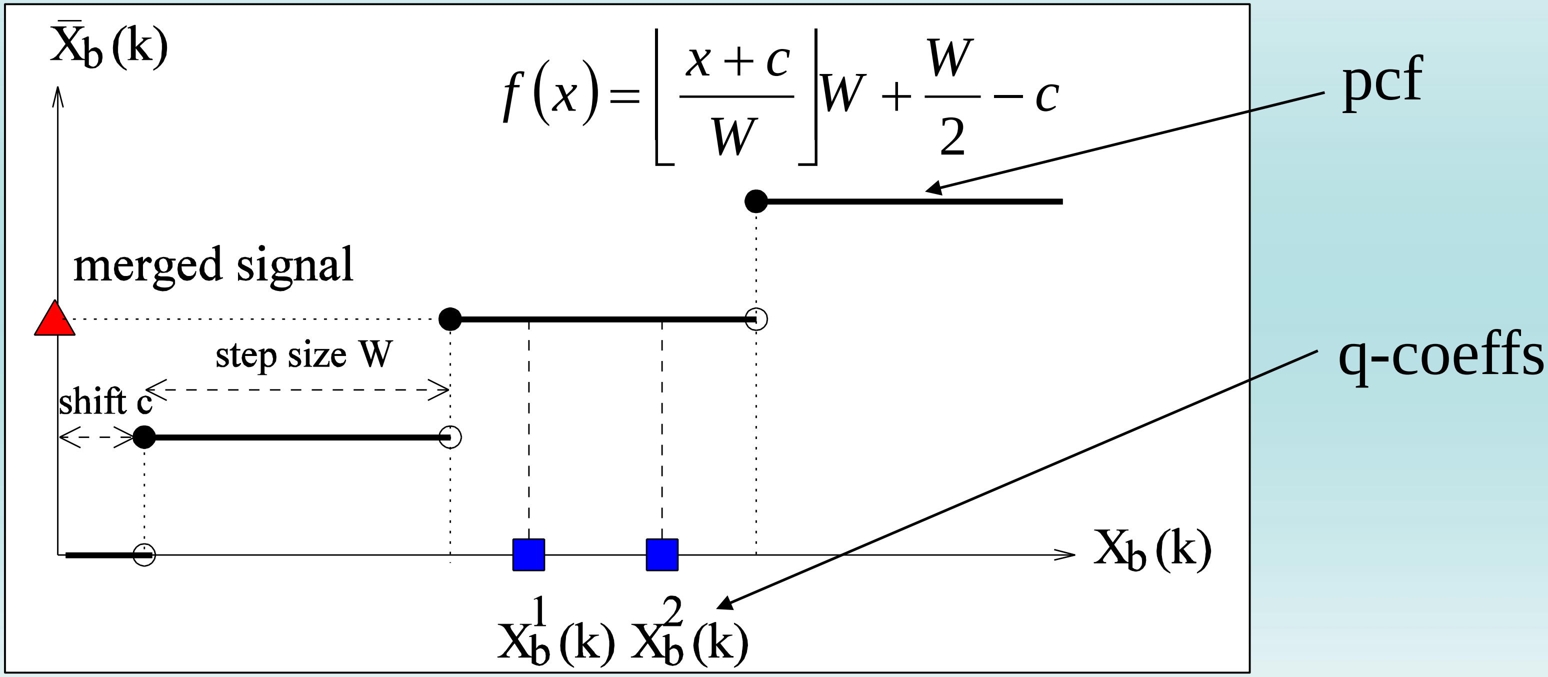
1. Each decoded SI frame is divided into 8x8 blocks, DCT transform and coefficient quantized (**q-coeff**).
2. Given block *b*, if q-coeffs of SI frames very different, use *I*-block.
3. If q-coeffs of SI frames the same, use *skip* block.
4. If q-coeffs of SI frames slightly different, use ***merge block***.





Merge Frame for Media Navigation: merge block overview

- Use ***piecewise constant function*** (pcf) for merging of SI's q-coeffs:
  - Q-coeff's must land on the same "step" for **identical merging**.
- pcf defined by step size *W* and shift c:
  - Choose *W* per frequency of all merge blocks (cheap).
  - Choose c per block per frequency (expensive).



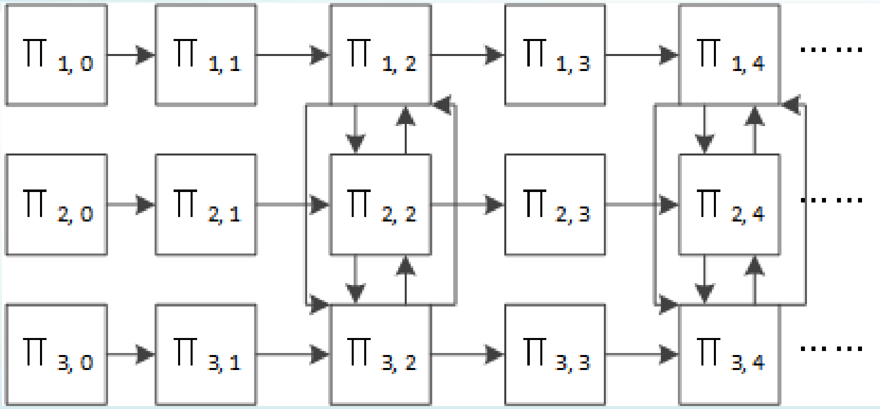
$$f(x) = \left\lfloor \frac{x+c}{W} \right\rfloor W + \frac{W}{2} - c$$

Merge Frame for Media Navigation: 2 merging problems

Fixed Target Merging:

- Find M-frame M to reconstruct any SI frame S^n, *n=1,…,N*, <u>*identically*</u> to a **fixed target** T.
- Difficult to optimize M-frame parameters.

Static view switching



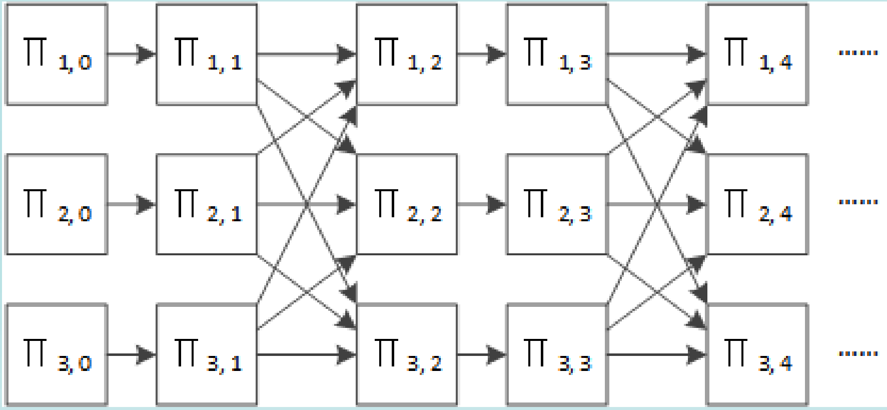
Optimized Target Merging:

- Find M-frame M to reconstruct any SI frame S^n, *n=1,…,N*, <u>*identically*</u> to a **floating target** $\overline{\mathrm{T}}(\mathrm{M})$, such that:

$$\mathrm{M}^* = \arg\min_{\mathrm{M}} D\left(\mathrm{T}, \overline{\mathrm{T}}(\mathrm{M})\right) + \lambda\, R(\mathrm{M})$$

- Optimize M-frame parameters in RD manner.

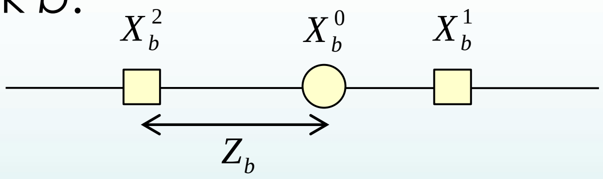
Dynamic view switching



Merge Frame for Media Navigation: step W, shift c (fixed target merging)

- **Choosing step size *W* for given freq *k*:**
  - Compute max diff. from **target q-coeff** in each block *b*:

$$Z_b = \max_{n \in \{1,\ldots,N\}} \left| X_b^0 - X_b^n \right|$$



  - Choose step size W to be roughly 2 * max diff:

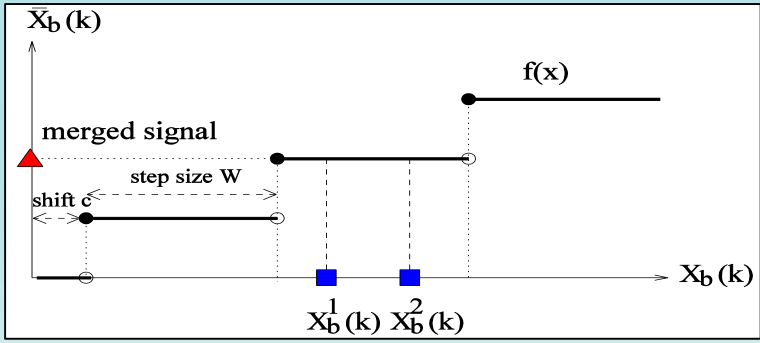
$$W_b^\# = 2Z_b + 2$$

pcf:

$$f(x) = \left\lfloor \frac{x+c}{W} \right\rfloor W + \frac{W}{2} - c$$

- **Choosing shift *c* for each block *b*:**
  - Choose shift: $c_b = W_b^\# / 2 - X_{b,2}^0$, where $X_{b,2}^0 = X_b^0 \bmod W_b^\#$
  - **Lemma V.1**: given this choice of step and shift,

$$f\left(X_b^n\right) = X_b^0, \quad \forall n \in \{0,\ldots,N\}$$



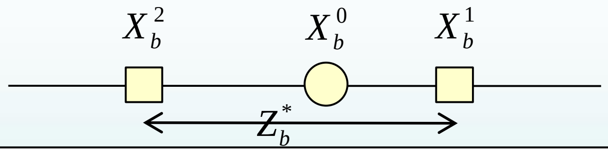
- **Merge block group *B*$_m$**, use a bigger step:

$$Z_{B_M} = \max_{b \in B_M} Z_b \qquad W_{B_m}^\# = 2Z_{B_m} + 2 \qquad X_{b,2}^0 = X_b^0 \bmod W_{B_m}^\#$$

Merge Frame for Media Navigation: step *W*, shift c (optimized target merging)

- **Choosing step size *W* for given freq *k*:**
  - Compute max diff. bet'n 2 q-coeffs in block *b*, then block-wise max diff.:

$$Z_b^* = \max_{i,j\in\{0,\dots,N\}} X_b^i - X_b^j \qquad Z_{B_M}^* = \max_{b\in B_M} Z_b^*$$



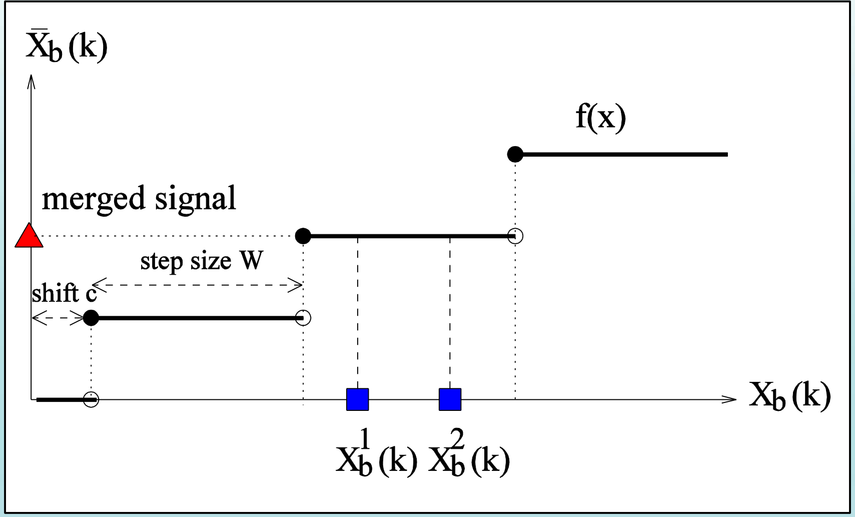
  - Choose step size W to be roughly max diff:

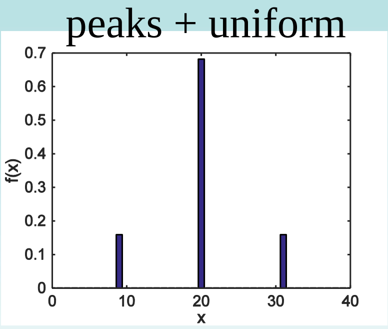
$$W_{B_M} = Z_{B_M}^* + 1$$

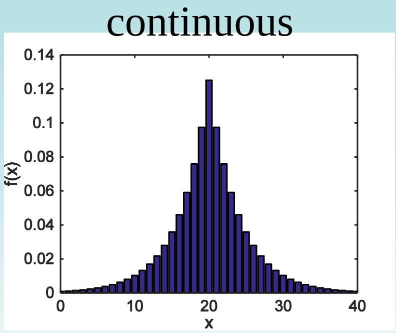
- **Choosing shift *c* for each block *b*:**
  - Given step *W*, range F_b of shifts c can lead to *identical merging*.
  - Choose c in F_b to min RD cost:

$$\min_{0\le c_b\le W_{B_M} \,|\, c_b\in F_b} d_b + \lambda\left(-\log P(c_b)\right)$$

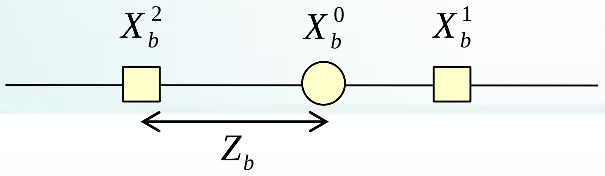
- **Initialize *P(c_b)*:**
  - Initialize a "peaks + uniform" distribution.
  - Rate-constrained LM till convergence.







Comparison with Coset Coding



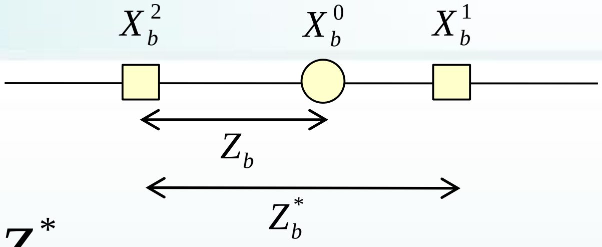
- **Coset Coding**:
  - SI values X_b^n are noisy observations of target X_b^0
  - Compute first largest difference w.r.t. to target:

$$Z_b = \max_n \left| X_b^n - X_b^0 \right|$$

  - **Encoder**: select coset size $W > 2Z_b$, transmit coset index $i_b = X_b^0 \bmod W$
  - **Decoder**: compute $\hat{X}_b = \arg\min_{X \in Z} \left| X_b^n - X \right| \quad s.t. \quad i_b = X \bmod W$

- **Fixed Target Merging**:
  - Step *W* is roughly $2Z_b$: $W_b^\# = 2Z_b + 2$
  - Shift *c* given *W* is remainder of target: $c_b = W_b^\# / 2 - X_{b,2}^0$, where $X_{b,2}^0 = X_b^0 \bmod W_b^\#$
  - <u>Expect the same coding rate as coset coding!</u>

Comparison with Coset Coding



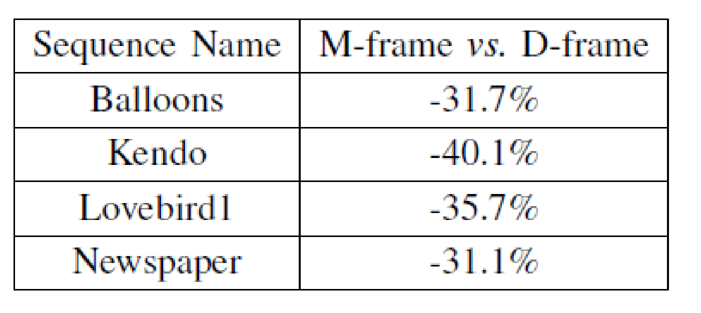
- **Optimized Target Merging**:
  - Step *W* is roughly Z_b: $W_b = Z_b^* + 1$, where $Z_b \le Z_b^*$
  - Compared to coset size $W > 2Z_b$, nearly half the step size!
  - Feasible range of shifts to select from via RD optimization:

$$\min_{0 \le c_b \le W_{B_M} \,|\, c_b \in F_b} d_b + \lambda\left(-\log P(c_b)\right)$$

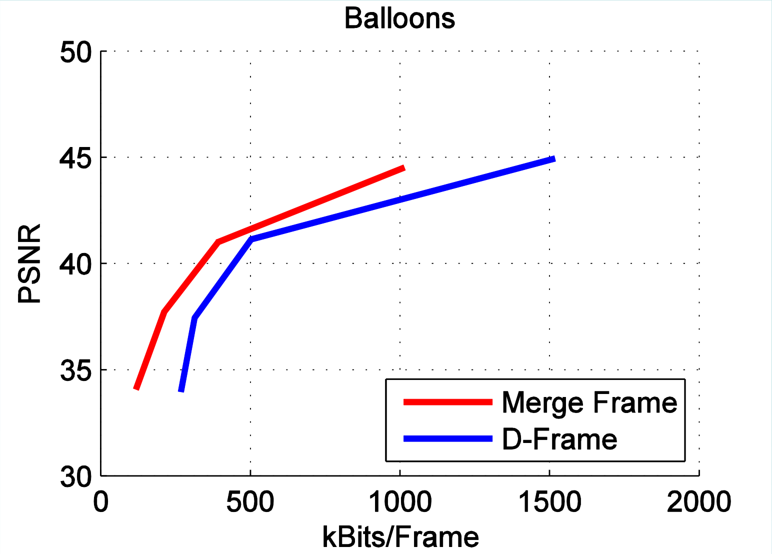
  - Expect significant coding gain, especially at low rates.

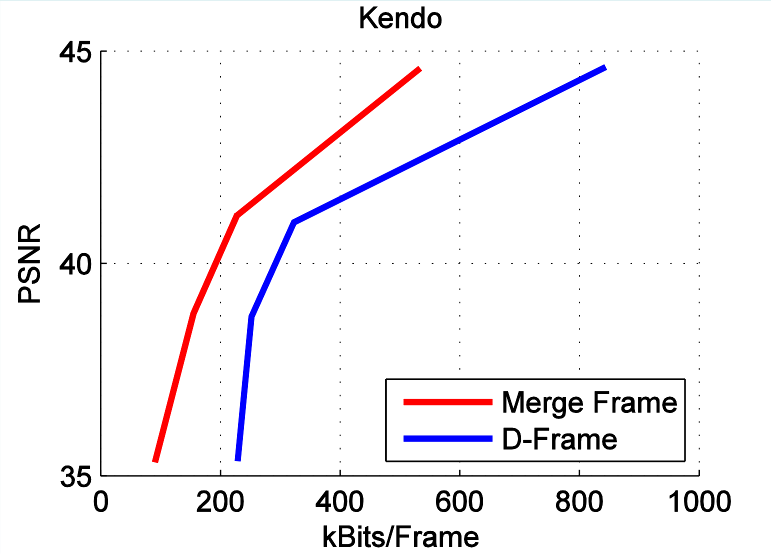
Merge Frame for Media Navigation: experiments

- **Exp Setup**: Static view switching
  - **Fixed target merging**: 3 views with the same QP.
  - H.264 for I- and P-frames.
  - Compared w/ DSC frames.



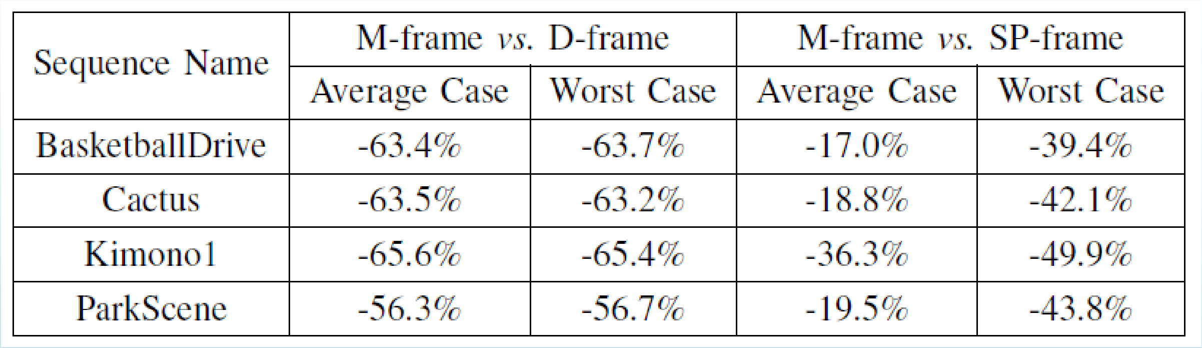
Sequence Name	M-frame *vs.* D-frame
Balloons	-31.7%
Kendo	-40.1%
Lovebird1	-35.7%
Newspaper	-31.1%



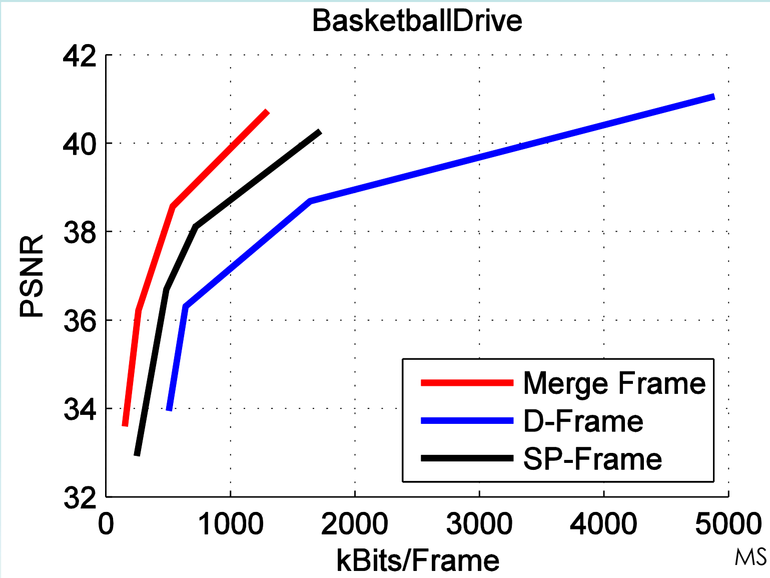


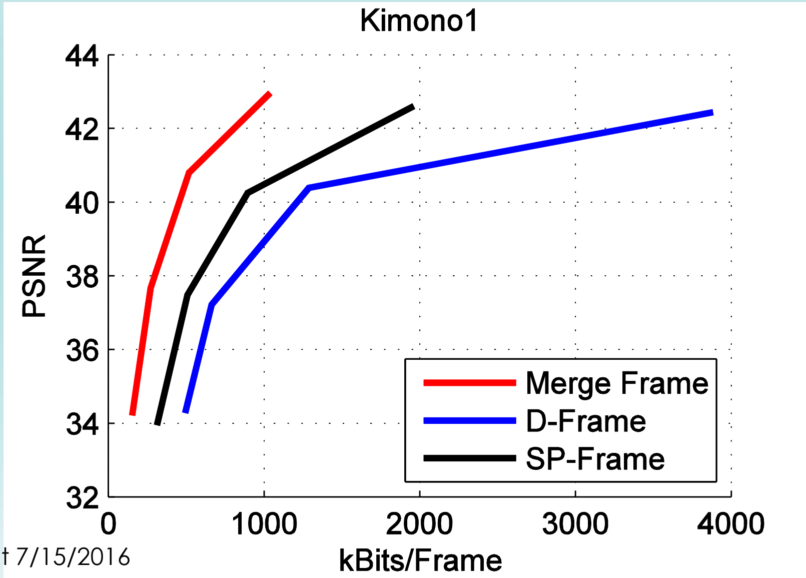
Merge Frame for Media Navigation: experiments 2

- **Exp Setup**: Bit-rate adaptation
  - **Optimized target merging**: 3 streams of same sequence at diff. rates (TFRC).
  - H.264 for I- and P-frames.
  - vs. DSC frames, SP-frames.
  - Worst case plots.



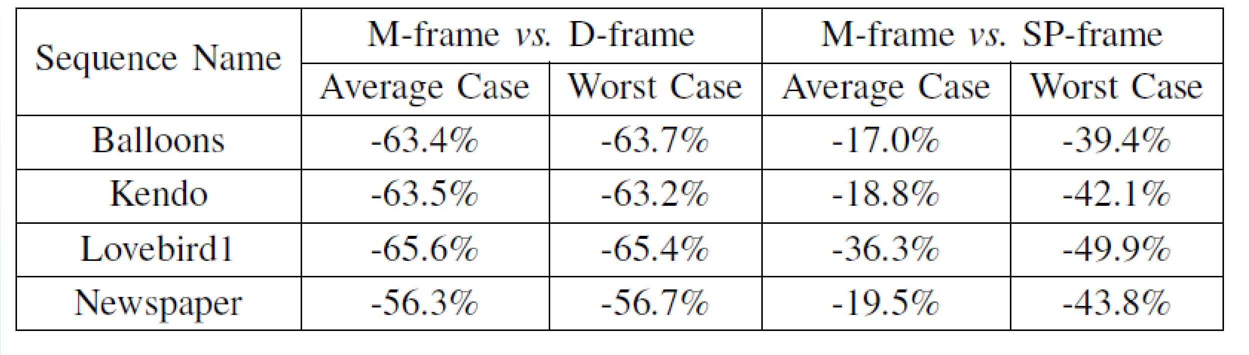
Sequence Name	M-frame *vs.* D-frame		M-frame *vs.* SP-frame	
	Average Case	Worst Case	Average Case	Worst Case
BasketballDrive	-63.4%	-63.7%	-17.0%	-39.4%
Cactus	-63.5%	-63.2%	-18.8%	-42.1%
Kimono1	-65.6%	-65.4%	-36.3%	-49.9%
ParkScene	-56.3%	-56.7%	-19.5%	-43.8%



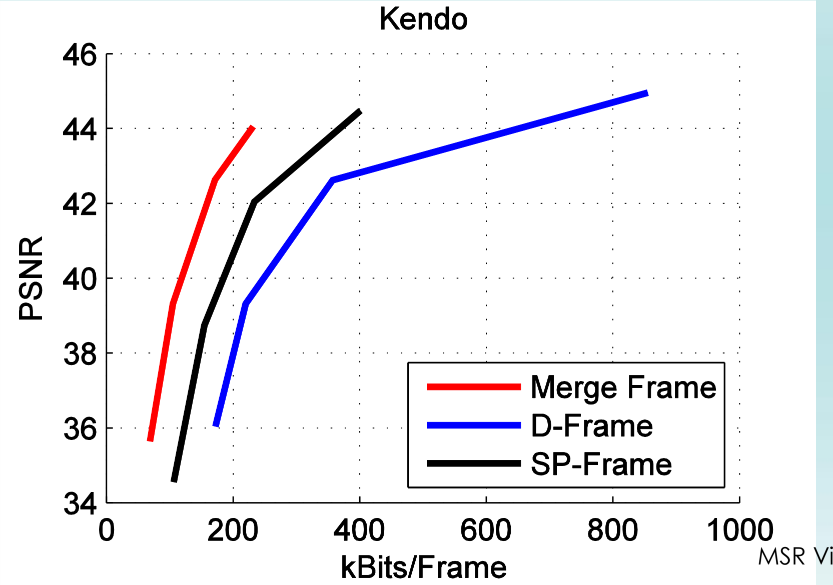


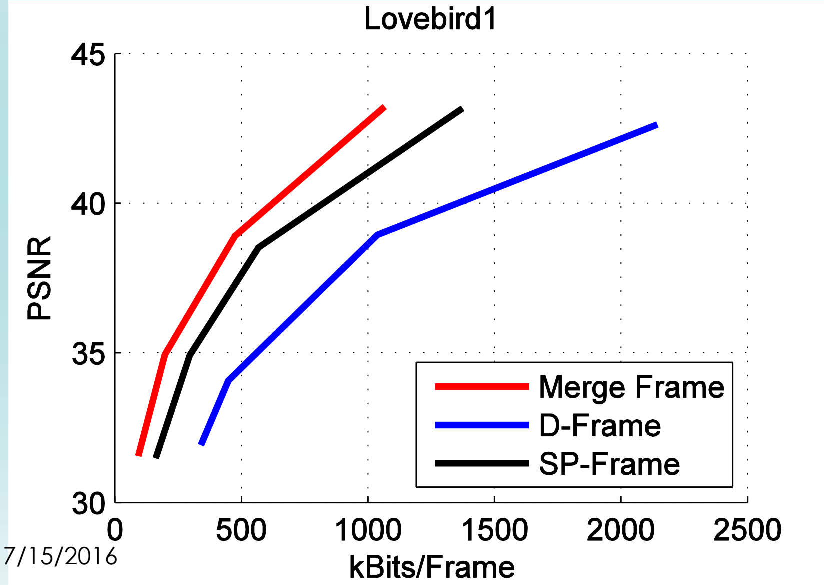
Merge Frame for Media Navigation: experiments 3

- **Exp Setup**: Dynamic view switching
  - **Optimized target merging**: 3 views with the same QP.
  - H.264 for I- and P-frames.
  - vs. DSC frames, SP-frames.
  - Worst case plots.



Sequence Name	M-frame *vs.* D-frame		M-frame *vs.* SP-frame	
	Average Case	Worst Case	Average Case	Worst Case
Balloons	-63.4%	-63.7%	-17.0%	-39.4%
Kendo	-63.5%	-63.2%	-18.8%	-42.1%
Lovebird1	-65.6%	-65.4%	-36.3%	-49.9%
Newspaper	-56.3%	-56.7%	-19.5%	-43.8%





Outline

- What is interactive media navigation?
    - e.g. Multiview / free-viewpoint video
- Merge frame for interactive media navigation
    - Previous works
    - Merge frame / block overview
    - Fixed target merging
    - Optimized target merging
- Interactive Light Field Streaming (ILFS)

Interactive Light Field Streaming (ILFS)

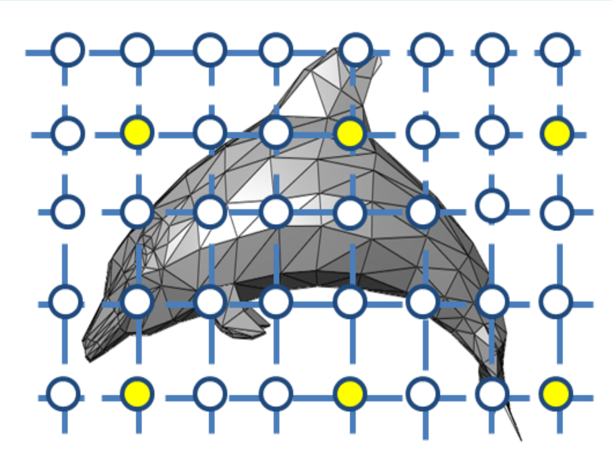
LYTRO

Light Field:

- Capture light intensity and direction per pixel.
  - Micro-lenses placed in front of a traditional image sensor.
- Generate 2D array of viewpoint images for users to navigate.

Goal:

- Design coding structures to facilitate view-switches, while achieving low trans. Rate.

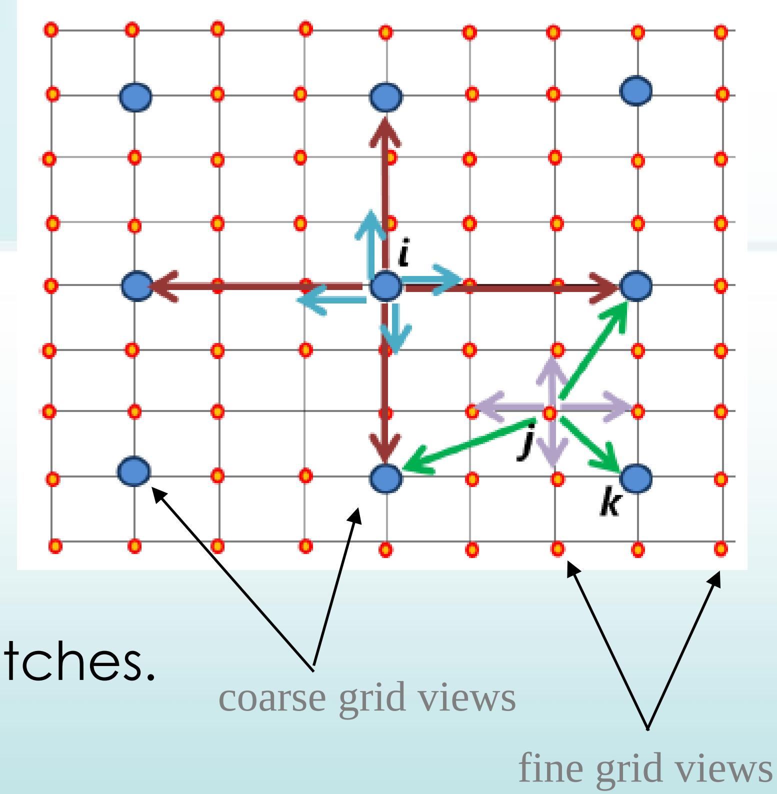


Idea:

- Build ***redundant*** structures using I-frames, P-frames, M-frames as building blocks, given storage size.

B. Motz, G. Cheung, A. Ortega, "**Redundant Frame Structure using M-frame for Interactive Light Field Streaming**," (accepted to) *IEEE International Conference on Image Processing*, Phoenix, USA, September, 2016.

User Interaction Model



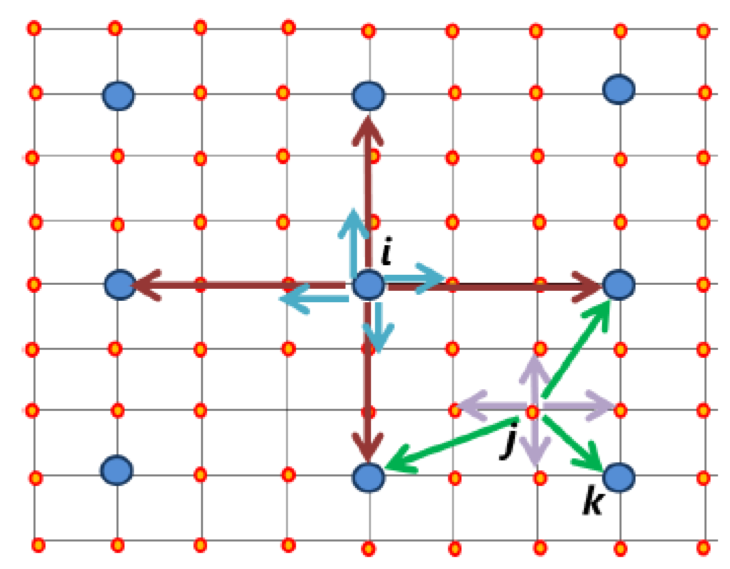
1. **View Navigation Model**:
   - Define permissible view-switches.
2. **User Behavior Model**:
   - Define probabilities of permissible view-switches.

View Navigation Model (#):

- **WALK**: navigate locally on fine grid.
  - move to horizontal/vertical adjacent fine views {n, e, s, w}.
- **JUMP**: navigate neighborhoods on coarse grid.
  - move to earest horizontal/vertical coarse views {N, E, S, W}.

#W. Cai, G. Cheung, S.-J. Lee, and T. Kwon, "**Optimal frame structure design using landmarks for interactive light field streaming**," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, Kyoto, Japan, March 2012.

User Behavior Model



Memoryless Model: $p_{i,j}$

- Prob of next view *j* depends on curr. view *i*.

1-hop Memory Model: $p_{k,i,j}$

- Prob of next view *j* depends on curr. view *i* & past view *k*.
- Tend to select same direction repeatedly.

Define $p_{k,i,j}$ when i is fine grid view:

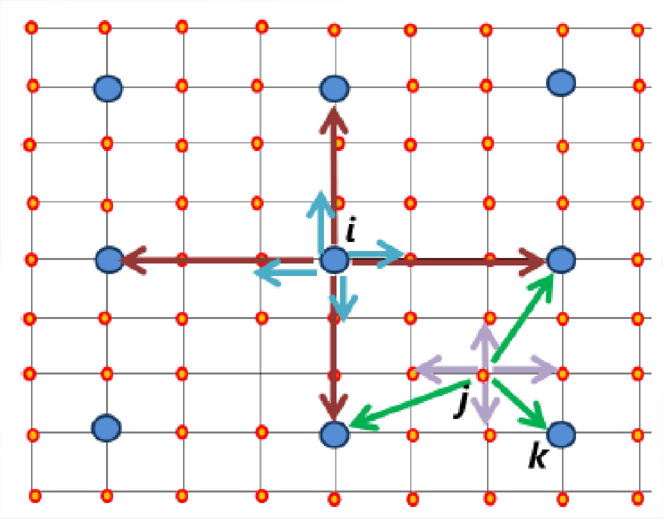
$$p_{k,i,j} = \begin{cases} q_1/3 & if \;\; j \in G^c \\ q_0(1-q_1) & if \;\; \phi(k,i) = \phi(i,j) \\ (1-q_0)(1-q_1)/3 & o.w. \end{cases}$$

switches to coarse grid views

switch to same-direction fine grid view

switches to different-direction fine grid views

Redundant Frame Structure

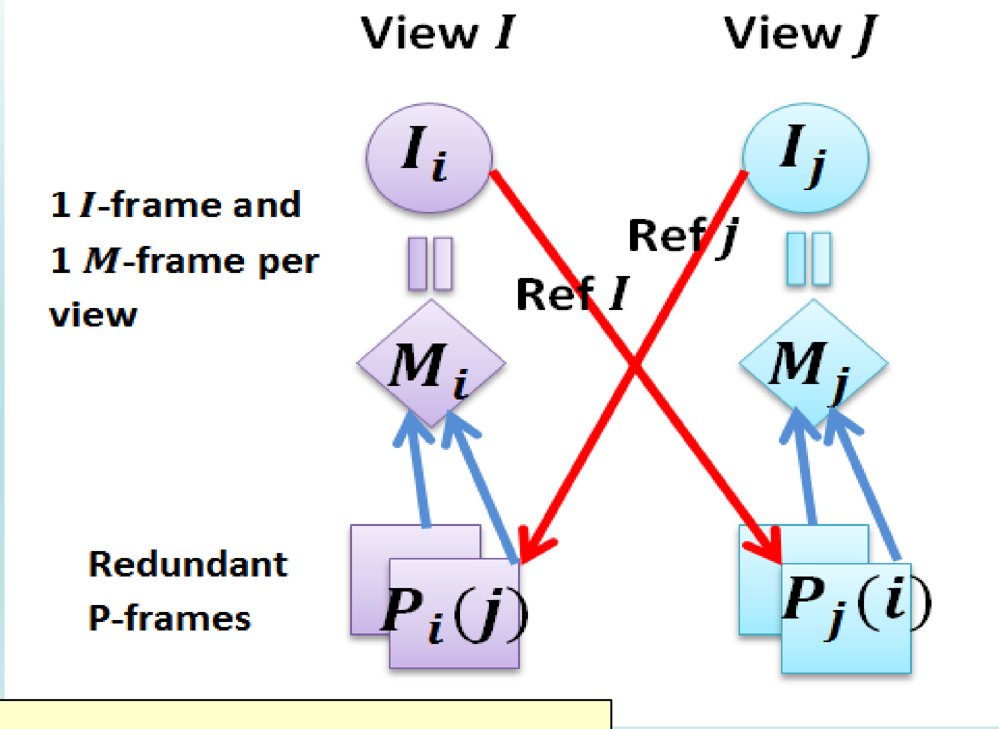


Default Structure:

- 1 I-frames, 1 M-frames per view.
- View navigation possible using I-frames.

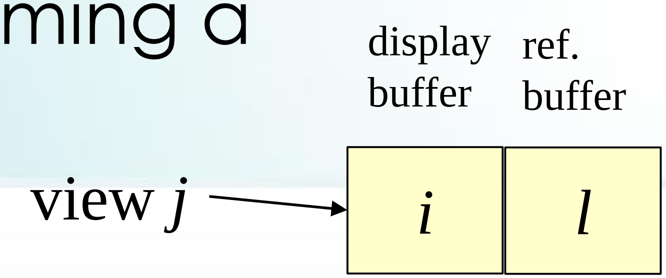
Redundancy in P-frames:

- Add P-frame $P_i(j)$ to facilitate switch from view *j* to view *i*.
- Diff. P-frames $P_i(j)$ reconstruct to same I-frame I_i using M-frame M_i .
- P-frame can enable ***2-hop trans***.



Question: which P-frames to add given storage constraint?

Expected transmission cost assuming a flexible 1-frame Buffer



Flexible 1-Frame Buffer:

- In addition to 1 I-frame ***display buffer***, there is 1-frame ***ref. buffer***.
- Simplified buffer model to keep optimization tractable.

- Assume lifetime of *T* view-switches.

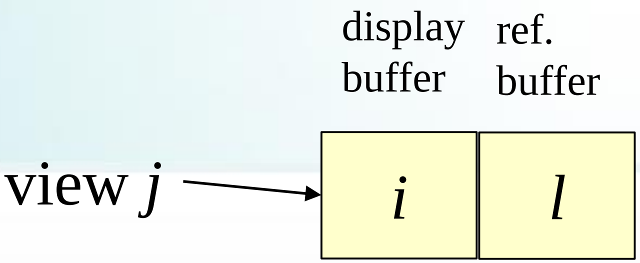
Expected Transmission cost for user at view *i* at instant *t*, given prev. view *k* and buffered view *l*:

$$c_{i|k}^{(t)}(l) = \sum_{j} p_{k,i,j} \min\left[h_i^{(t)}(l,j), \dot{h}_i^{(t)}(l,j), \ddot{h}_i^{(t)}(l,j)\right]$$

0-hop trans. 1-hop trans. 2-hop trans.

view-switching prob.

0-hop transmission cost (I-frame)



display buffer

ref. buffer

view *j*

i

l

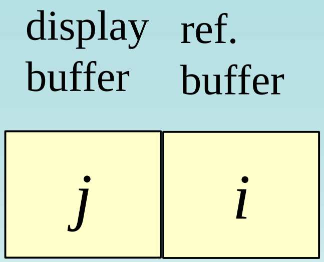
- Send I-frame I_j given curr. view *i* and ref. view *l*.
- A choice of keeping view *i* or *l* in ref. buffer.

$$h_i^{(t)}(l, j) = r_j^I + 1(t < T)\min_{\gamma\in\{l,i\}} c_{j|i}^{(t+1)}(\gamma)$$

I-frame trans. cost

Recurse only if there are view-switches

recursive cost with choice of ref frame

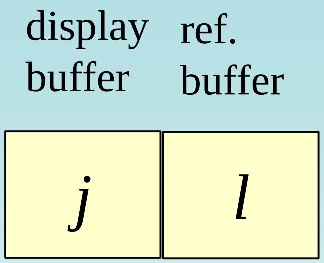


display buffer

ref. buffer

j

i



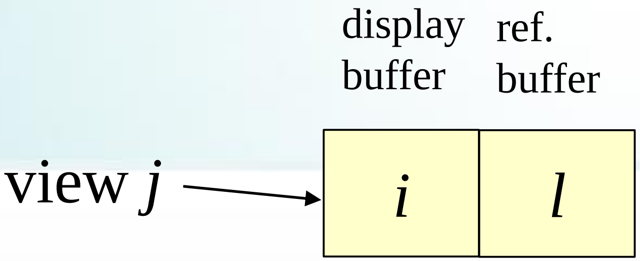
display buffer

ref. buffer

j

l

1-hop transmission cost (one P-frame)


view *j*

display buffer | ref. buffer

i | *l*

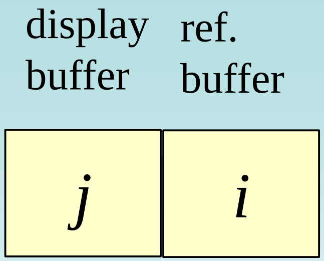
- Send P-frame $P_j(i)$ or $P_j(l)$ plus M-frame M_j given curr. view *i* and ref. view *l*.
- Occupancy of ref. buffer depends on P-frame used.

$$\dot{h}_i^{(t)}(l,j) = \min_{\gamma\in\{l,i\}} \left[r_j^P(\gamma) + 1(t<T)\, c_{j|i}^{(t+1)}(\gamma) \right]$$

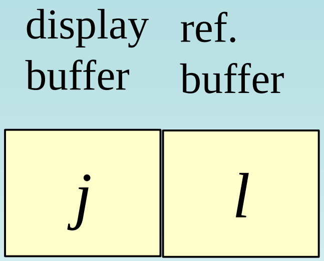
P-frame trans. cost

Recurse only if there are view-switches

recursive cost with different ref frames

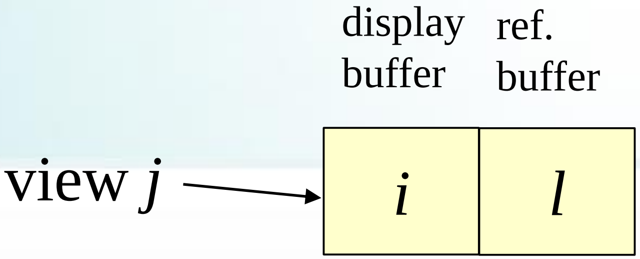

display buffer | ref. buffer

j | *i*


display buffer | ref. buffer

j | *l*

2-hop transmission cost (two P-frames)



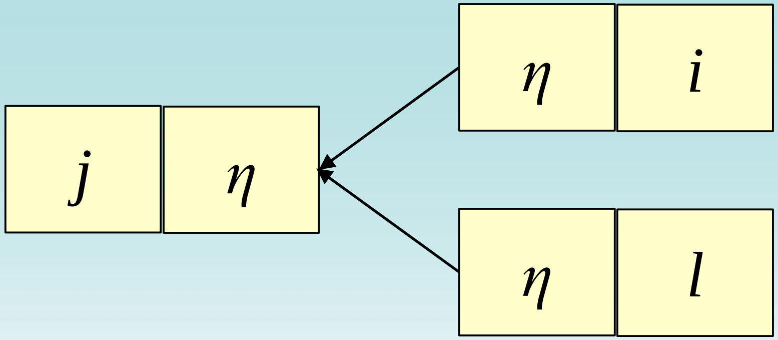
- Transition to **intermediate view** η, then transition from η to destination j.
- Occupancy of ref. buffer depends on P-frame used.

$$\ddot{h}_t^{(t)}(l, j) = \min_{\eta} \left[r_j^P(\eta) + 1(t < T) c_{j|i}^{(t+1)}(\eta) + \min_{\gamma \in \{l,i\}} r_\eta^P(\gamma) \right]$$

P-frame trans. cost from intermediate view η

recursive cost

transition cost to intermediate view η



Structure Optimization

- **Question**: how to add P-frames given storage constraint?
- **Greedy Alg**: add P-frame that maximally lower Lagrangian cost, one at a time:

$$\min_{\theta} c_s^{(0)}(\theta) + \lambda\, b(\theta)$$

expected trans. cost

storage cost

$$b(\theta) = \sum_{P_j(i) \in \theta} \left| P_j(i) \right|$$

P-frames in structure θ

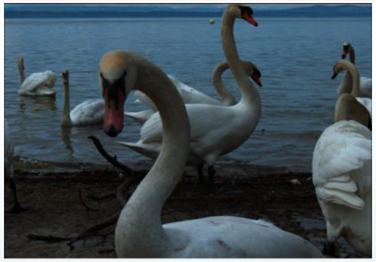
Experimental Setup

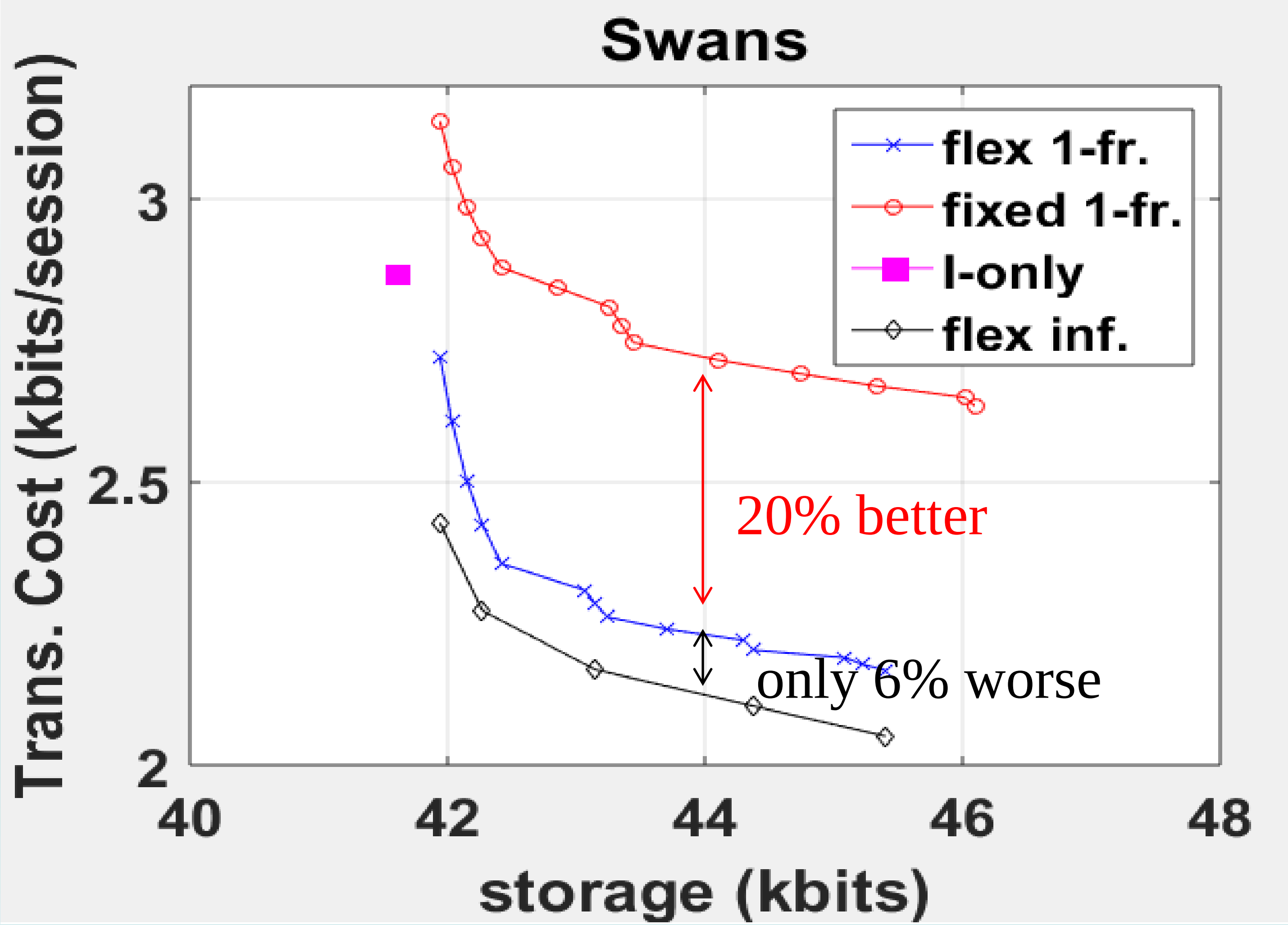
- 2 Light field images of size 432x624
- We select a 6x6 fine grid and 2x2 coarse grid
- The user can switch T=12 times
- HEVC HM-15.0 for I-, P-frames. QP is set s.t. PSNR=36dB

	Flowers	**Swans**
I-frames cost	x 5 Vertical P-frames x10 horizontal P-frames	x10 P-frames
M-frames cost	x3 Vertical P-frames x6 horizontal P-frames	x5 P-frames

- $q_0 = 0.4$ and $q_1 = 0.6$
- **COMPARISON SCHEME:**
  - **I-only:** structure with only I frames
  - **Fixed 1 frame buffer:** ref. view is previous displayed view.
  - **Flexible infinite buffer:** Client keeps all traversed frames for ref. Simulate 100 clients for average.

Simulation Results (Swans Dataset)

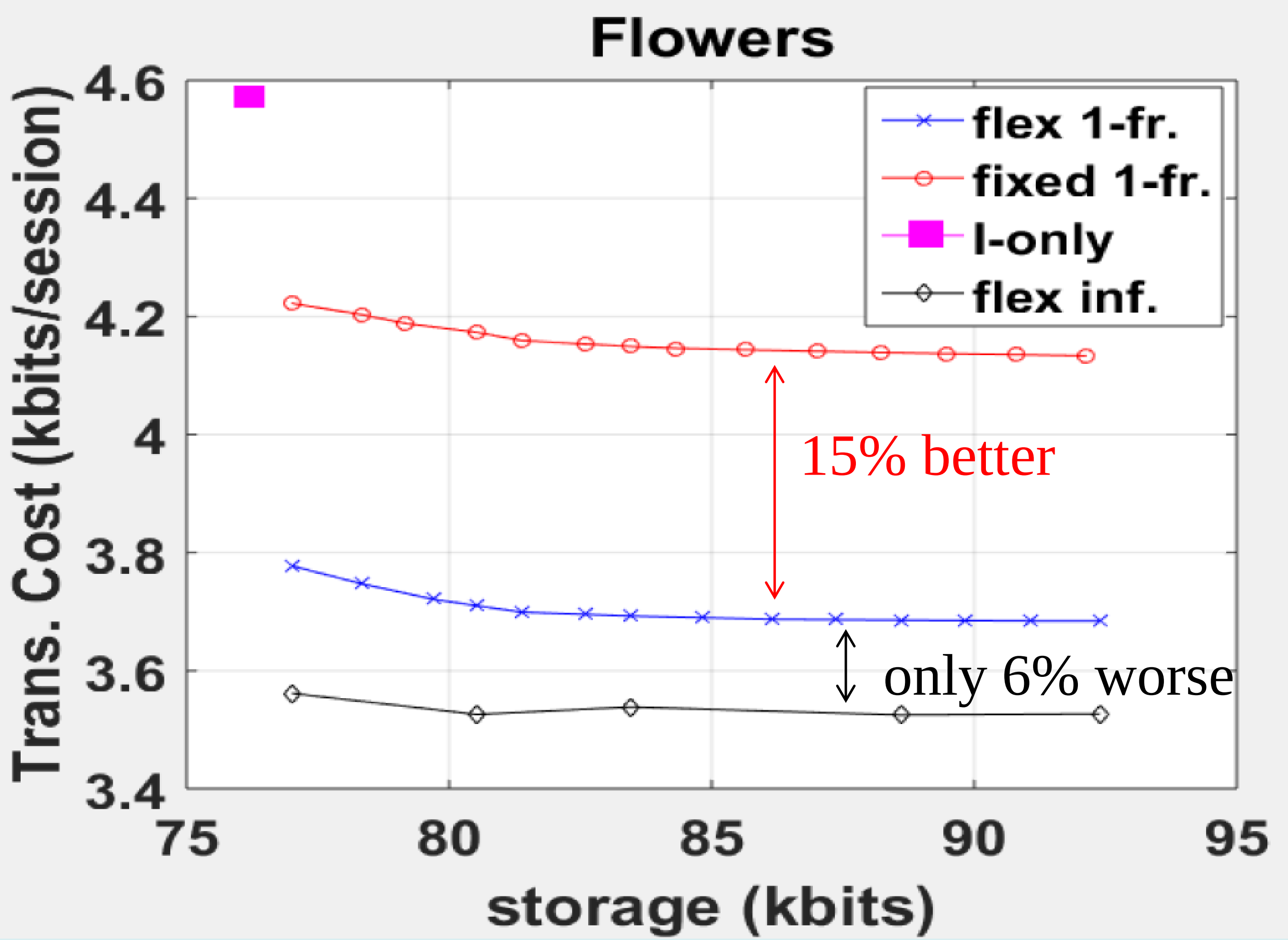




20% better

only 6% worse

Simulation Results (Flowers Dataset)



15% better

only 6% worse

Summary

- Interactive media navigation
  - Difficult to achieve to good compression efficiency & flexible decoding.
- Merge frame to facilitate interactive navigation
  - Fixed target merging
  - Optimized target merging
- Interactive light field streaming
  - Redundancy to enable faster switches

Q&A

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