Gene Cheung
National Institute of Informatics
2nd October, 2013

3D visual communication: media representation, transport and rendering

.

Acknowledgement

Collaborators:

- Yu Mao (NII, Japan)
- Wei Hu, Wenxiu Sun, Wei Dai, Prof. Oscar Au (HKUST, HK)
- Prof. Antonio Ortega (USC, USA)
- Dr. Dinei Florencio, Cha Zhang, Phil Chou (MSR)
- Xiaoyu Xiu, Hadi Hadizadeh, Prof. Jie Liang, Prof. Ivan Bajic (SFU, Canada)
- Prof. Ngai-Man Cheung (SUTD, Singapore)
- Prof. Bruno Machiavello, Camilo Dorea, Mintsu Hung (UofBrasilia, Brazil)
- Dr. Wai-tian Tan (HPL, now Cisco)

Slides Contributors:

- Dr. Philipp Merkle (HHI, Germany)
- Prof. Minh Do (UIUC, USA)
- Prof. Patrick Le Callet (UofNantes, France)
- Dr. Thomas Maugey, Prof. Pascal Frossard (EPFL, Switzerland)
- Mr. Hiroshi Sankoh (KDDI Labs, Japan)

Presentation Outline

- **Background & Motivation** (3D, not your mother's 2D)
- **3D Video representation / coding:**
  - Depth map coding
    - HEVC tools for depth maps
    - Graph-based Transform (GBT) for depth maps
  - Depth map denoising
  - Denoising + compression?
  - Why code depth images?
- **3D Video streaming:**
  - Video compression with flexible decoding for interactive streaming
  - Loss-resilient texture-plus-depth video streaming (skip)
- **3D view synthesis:**
  - Robust view synthesis for free viewpoint video
  - Synthesized image interpolation for *z*-dimension camera movement

Presentation Outline

- **Background & Motivation** (3D, not your mother's 2D)
- **3D Video representation / coding:**
  - Depth map coding
    - HEVC tools for depth maps
    - Graph-based Transform (GBT) for depth maps
  - Depth map denoising
  - Denoising + compression?
  - Why code depth images?
- **3D Video streaming:**
  - Video compression with flexible decoding for interactive streaming
  - Loss-resilient texture-plus-depth video streaming (skip)
- **3D view synthesis:**
  - Robust view synthesis for free viewpoint video
  - Synthesized image interpolation for *z*-dimension camera movement

Biography (how I got started in 3D)

- 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 to 2009.
  - *Topic 1*: 2D video coding & streaming optimization (2000~2007).
  - *Topic 2*: Interactive multiview video, w/ Prof. Ortega (2007~).
- Faculty in NII from 11/2009 to now.
  - *Topic 1*: Immersive visual communication:
    - Free viewpoint video coding, streaming, view synthesis.
  - *Topic 2*: Visual saliency & gaze analysis.

Biography (how I got started in 3D)

2D video Communication (12 yrs)

- 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 to 2009.
  - *Topic 1*: 2D video coding & streaming optimization (2000~2007).
  - *Topic 2*: Interactive multiview video, w/ Prof. Ortega (2007~).
- Faculty in NII from 11/2009 to now.
  - *Topic 1*: Immersive visual communication:
    - Free viewpoint video coding, streaming, view synthesis.
  - *Topic 2*: Visual saliency & gaze analysis.

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

Biography (how I got started in 3D)

2D video Communication (12 yrs)

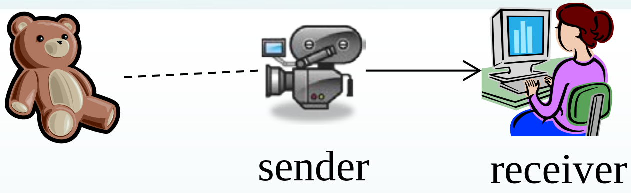
- MS from UC Berkeley in EECS in 1998.
  - *Thesis*: Joint source / channel coding for wireless video.
- PhD from UC Berkeley in EECS in 2000.
  - *Thesis*: Computation / memory / distortion tradeoff in signal compression.
- Senior researcher in HP Labs Japan from 2000 to 2009.
  - *Topic 1*: 2D video coding & streaming optimization (2000~2007).
  - *Topic 2*: Interactive multiview video, w/ Prof. Ortega (2007~).
- Faculty in NII from 11/2009 to now.

3D video Communication (7 yrs)

  - *Topic 1*: Immersive visual communication:
    - Free viewpoint video coding, streaming, view synthesis.
  - *Topic 2*: Visual saliency & gaze analysis.

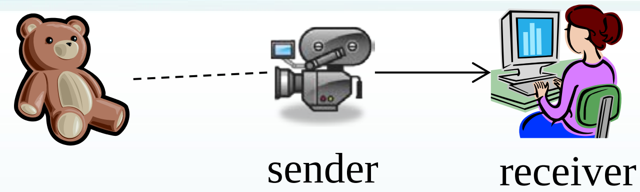
Video Communication: 2D to 2.5D to 3D

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

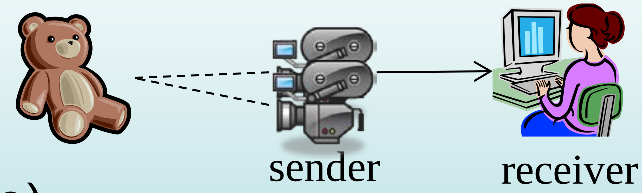


Video Communication: 2D to 2.5D to 3D

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

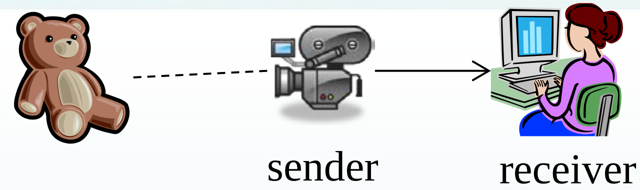


- **2.5D Video (stereoscopic)**
  - 2 capturing cameras @ sender.
  - 1 stereoscopic display @ receiver (non-interactive).

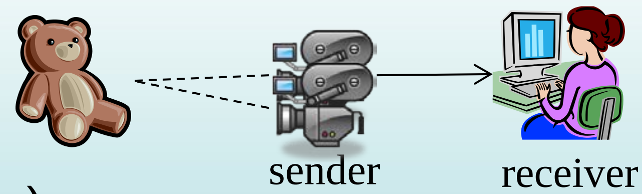


Video Communication: 2D to 2.5D to 3D

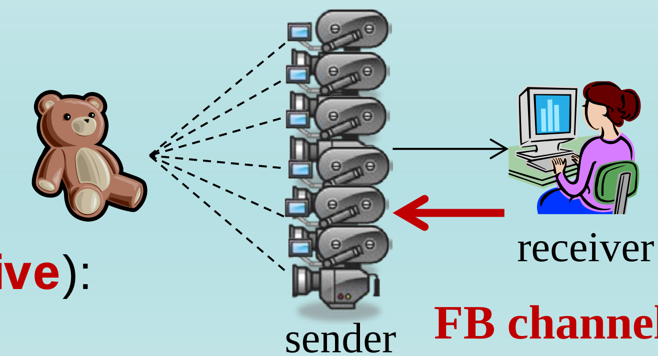
- **2D Video**
  - 1 capturing camera @ sender.
  - 1 2D display @ receiver (non-interactive).



- **2.5D Video (stereoscopic)**
  - 2 capturing cameras @ sender.
  - 1 stereoscopic display @ receiver (non-interactive).

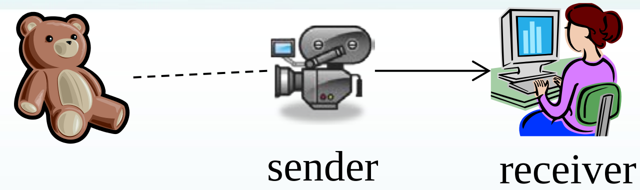


- **3D Video (multiview, free viewpoint)**
  - Multiple capturing cameras @ sender.
  - 1 2D / stereoscopic display @ receiver (**interactive**):
    - Receiver observes subset of high dimension media available @ sender!

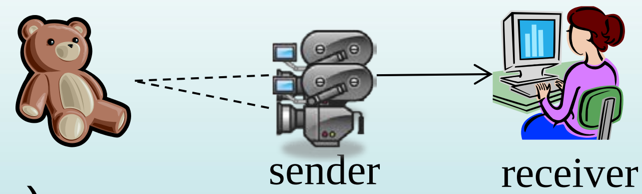


Video Communication: 2D to 2.5D to 3D

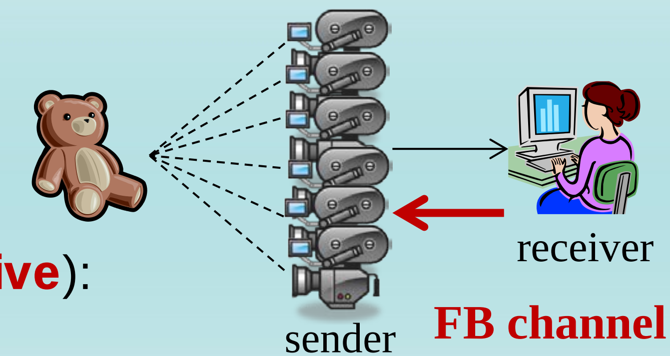
- **2D Video**
  - 1 capturing camera @ sender.
  - 1 2D display @ receiver (non-interactive).



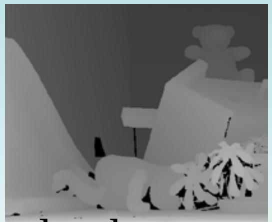
- **2.5D Video (stereoscopic)**
  - 2 capturing cameras @ sender.
  - 1 stereoscopic display @ receiver (non-interactive).



- **3D Video (multiview, free viewpoint)**
  - Multiple capturing cameras @ sender.
  - 1 2D / stereoscopic display @ receiver (**interactive**):
    - Receiver observes subset of high dimension media available @ sender!

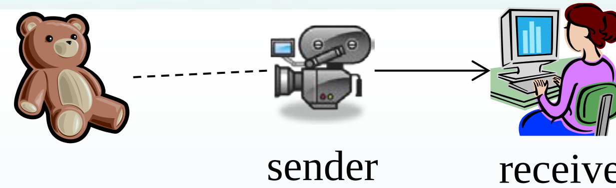


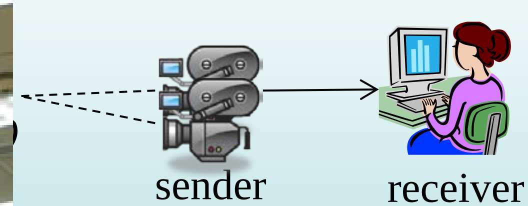

texture map

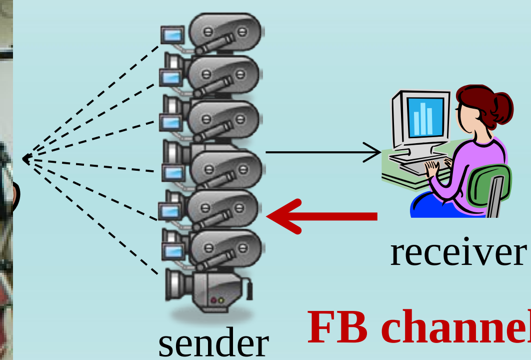

depth map

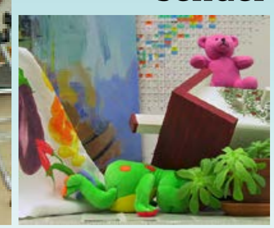
Video Communication: 2D to 2.5D to 3D

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

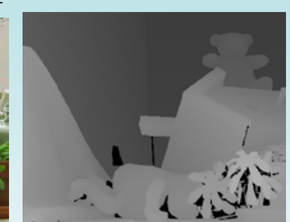








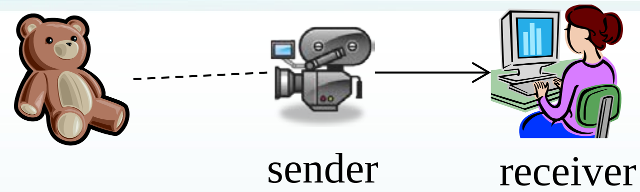
texture map



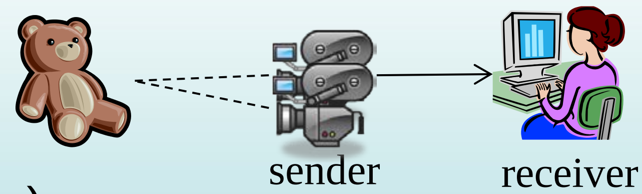
depth map

Video Communication: 2D to 2.5D to 3D

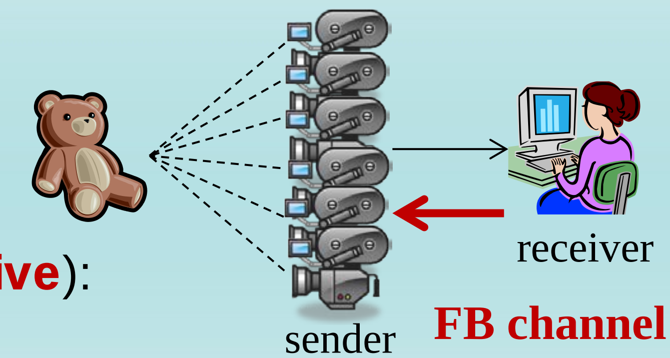
- **2D Video**
  - 1 capturing camera @ sender.
  - 1 2D display @ receiver (non-interactive).

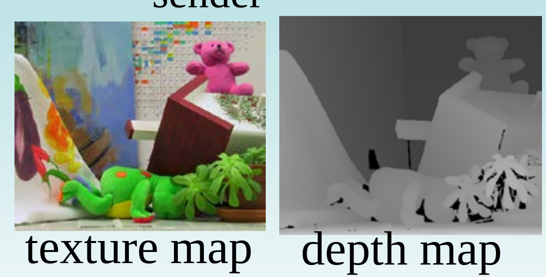


- **2.5D Video (stereoscopic)**
  - 2 capturing cameras @ sender.
  - 1 stereoscopic display @ receiver (non-interactive).



- **3D Video (multiview, free viewpoint)**
  - Multiple capturing cameras @ sender.
  - 1 2D / stereoscopic display @ receiver (**interactive**):
    - Receiver observes subset of high dimension media available @ sender!





Multiview Video Streaming

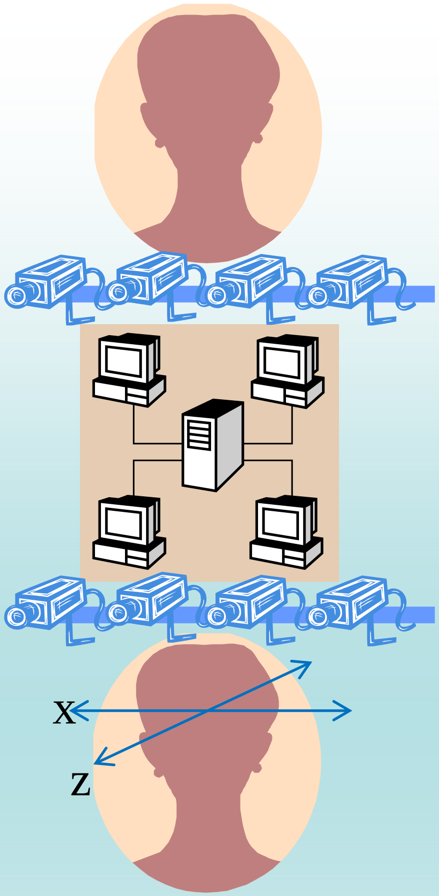
- Interactive view-switches among captured camera viewpoints.

Free Viewpoint Video Streaming

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

*Courtesy of KDDI Laboratories, Japan

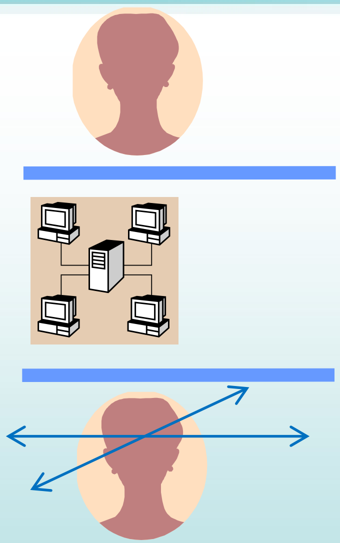
Immersive Communication



- **Goal**: ultra-realistic networked visual communication.
- **Application**: HQ teleconferencing, tele-medicine, distance learning.
- **Features**:

1. Gaze-corrected view.
2. **Motion Parallax**: fast, smooth interactive view-switching triggered by tracked observer's head position.
3. Low-delay, loss-resilient network transmission.

Potential Impact

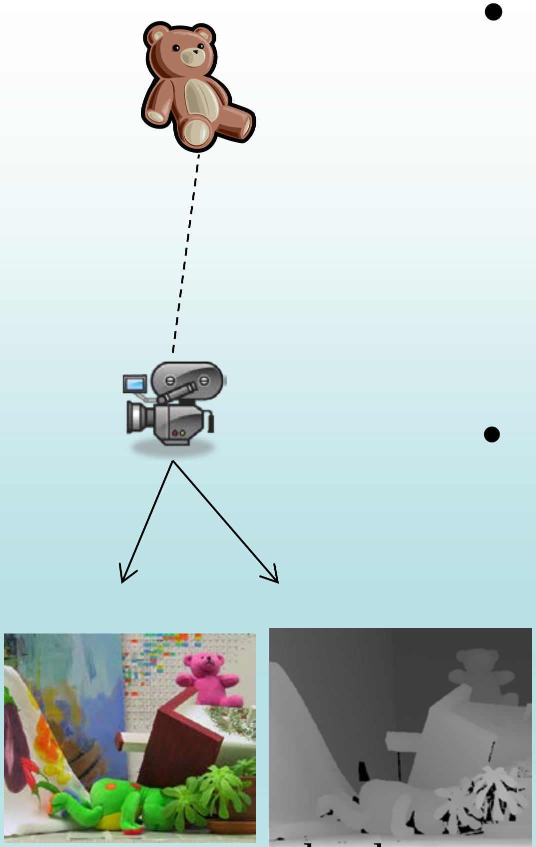


- Immersive Communication ≠ Skype calls!
  - Non-verbal means (postures, gestures) are important.
  - Eye-contact is important.
  - Depth perception via ***motion parallax***.
- Substitute for face-to-face meetings.
  - Reduce travel cost, improve productivity.
  - Reduce carbon footprints.
  - Example apps: HQ teleconferencing, tele-medicine.
- Enhance Virtual Reality is 1 of 14 grand challenges chosen by ***National Academy of Engineering*** for 21^{st} century.
  - Treatment of social anxieties, phobias, children ***autism***.
  - Training & teaching: virtual surgeries, etc.

Presentation Outline

- **Background & Motivation** (3D, not your mother's 2D)
- **3D Video representation / coding:**
  - Depth map coding
    - HEVC tools for depth maps
    - Graph-based Transform (GBT) for depth maps
  - Depth map denoising
  - Denoising + compression?
  - Why code depth images?
- **3D Video streaming:**
  - Video compression with flexible decoding for interactive streaming
  - Loss-resilient texture-plus-depth video streaming (skip)
- **3D view synthesis:**
  - Robust view synthesis for free viewpoint video
  - Synthesized image interpolation for *z*-dimension camera movement

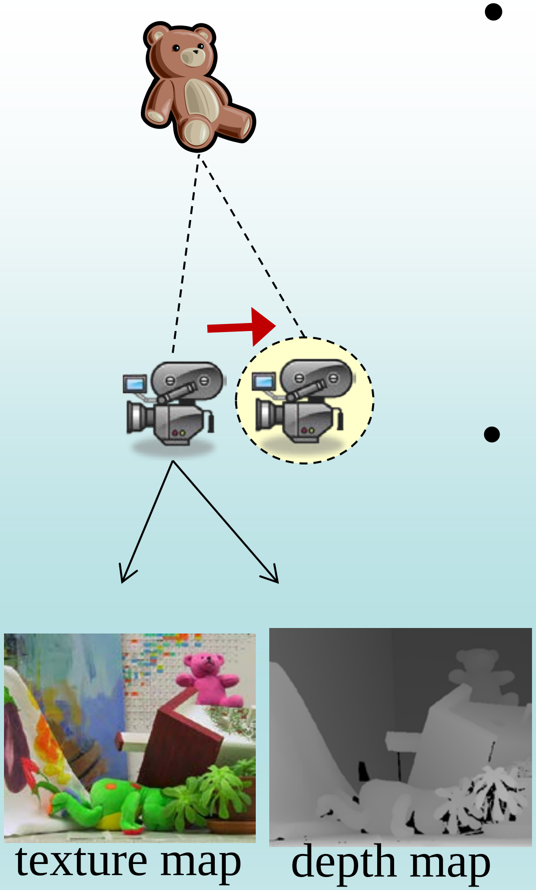
3D Video Representation



texture map depth map

- Texture + depth maps from 1 or more camera viewpoints.
  - **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).
  - Computation-efficient.
  - Unlike model-based approach, complexity not scene-dependent.

3D Video Representation

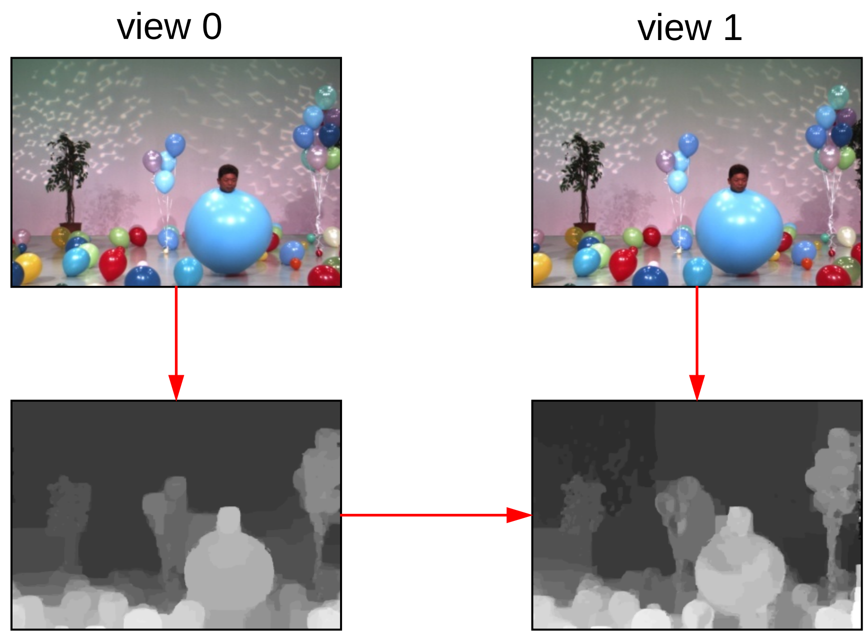


- Texture + depth maps from 1 or more camera viewpoints.
  - **Texture map**: color image like RGB.
  - **Depth map**: per-pixel distance bt'n captured objects in 3D scene & capturing camera.
- Synthesis of intermediate views via *depth-image-based rendering* (DIBR).
  - Computation-efficient.
  - Unlike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

Coding of depth maps

Coding of depth or disparity maps

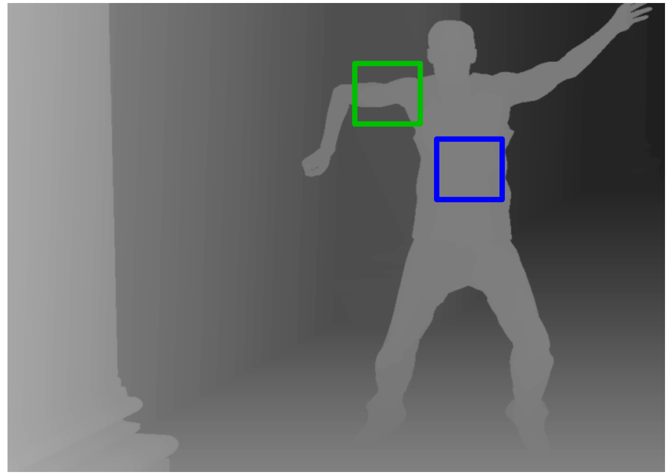
- Inter-view and additionally inter-component correlations are exploited by prediction-based coding
- Tools:
  - Disparity-compensated prediction for dependent view
  - Depth modeling modes
  - Motion parameter inheritance
  - Synthesized view distortion optimization



K. Müller, "3D High-Efficiency Video Coding for Multi-View Video and Depth Data," IEEE Transactions on Image Processing, Sep. 2013.
Courtesy of Fraunhofer HHI, Berlin, Germany.

Depth Modeling Modes

Fraunhofer
Heinrich Hertz Institute



Depth map properties:

- Sharp edges representing object borders
- Large areas of slowly varying values representing object areas
- Edges in depth maps are correlated with edges in video pictures

New intra prediction modes

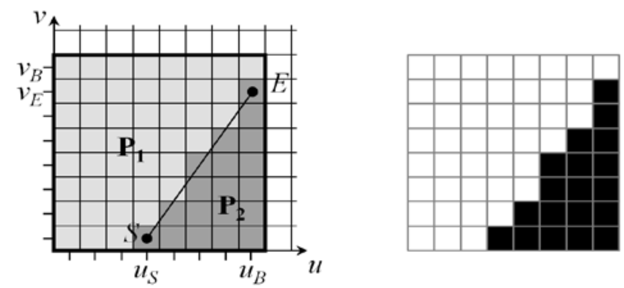
- Representation of depth edges
- Partition block into two regions with constant sample values
- Prediction based on co-located texture block
- Optional transform coding of residual

P. Merkle et al., "Coding of depth signals for 3D video using wedgelet block segmentation with residual adaptation," ICME 2013.
Courtesy of Fraunhofer HHI, Berlin, Germany.

Depth Modeling Modes – Intra Wedgelet

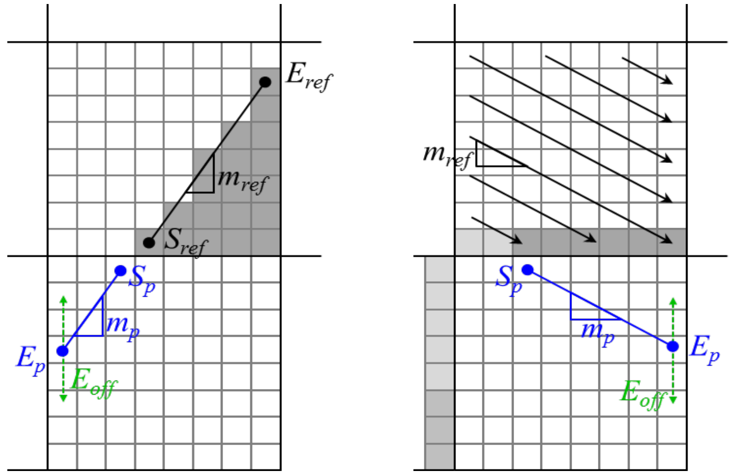
Explicit Wedgelet signaling

- Wedgelet partition of current block is estimated at the encoder by minimum distortion search using original depth signal
- Pre-defined lists of Wedgelet patterns for fast search and efficient signaling



Intra-predicted Wedgelet partitioning

- Separation line for current block is predicted from neighboring blocks
- Prediction from Wedgelet block by continuing separation line in current block
- Prediction from conventional intra block by combining direction and maximum slope point
- Transmission of line end refinement



P. Merkle et al., “Coding of depth signals for 3D video using wedgelet block segmentation with residual adaptation,“ ICME 2013.
Courtesy of Fraunhofer HHI, Berlin, Germany.

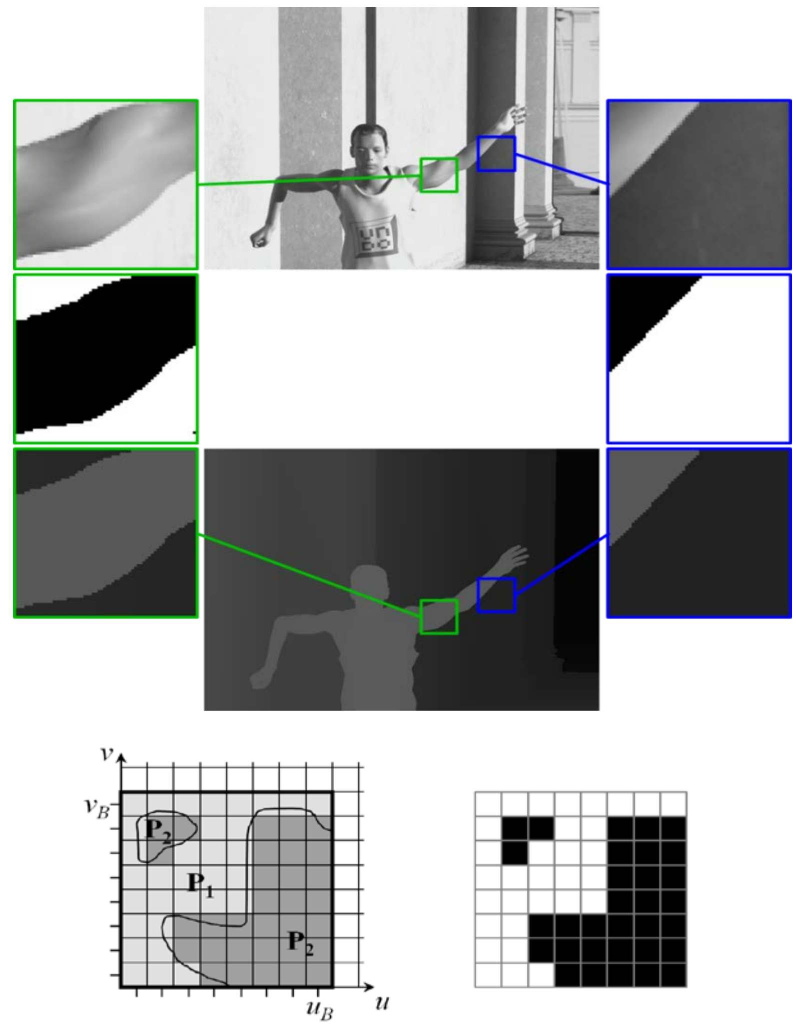
Depth Modeling Modes – Inter-component

Inter-component prediction of Wedgelet

- Wedgelet partition of current block is predicted from co-located block of reconstructed video picture by minimum distortion search
- Disable mode when co-located texture block has insignificant texture information (using mean absolute difference)

Inter-component prediction of Contour

- Contour partition of current block is predicted from co-located block of reconstructed video picture by thresholding segmentation
- Disable mode when co-located texture block has insignificant texture information

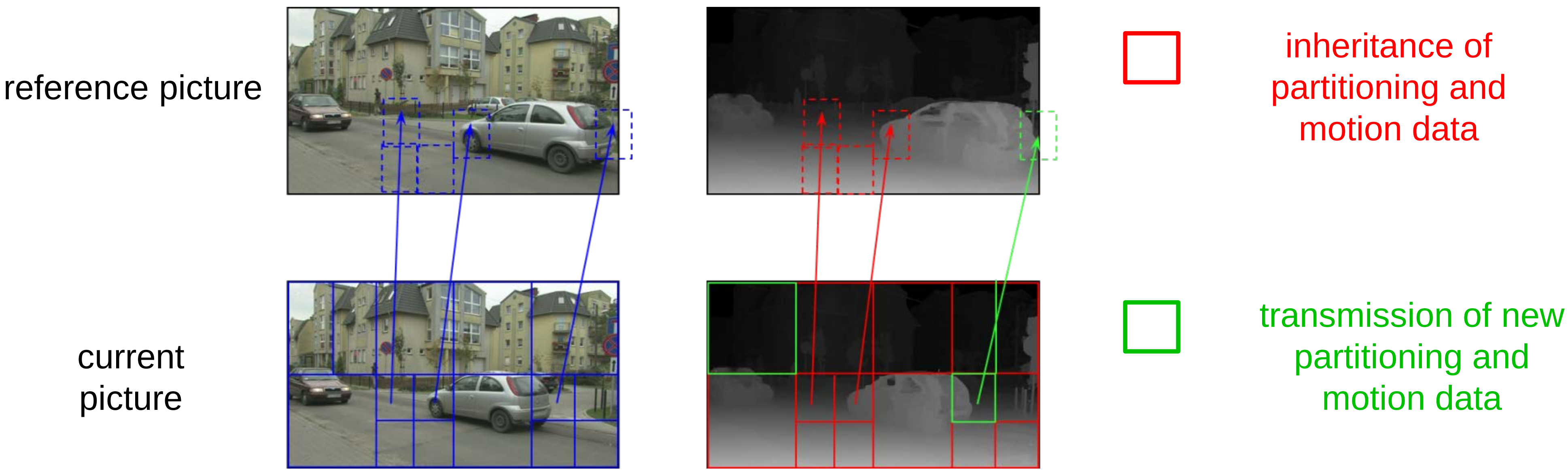


P. Merkle et al., "3D video: Depth coding based on inter-component prediction of block partitions," PCS 2012.
Courtesy of Fraunhofer HHI, Berlin, Germany.

Motion Parameter Inheritance

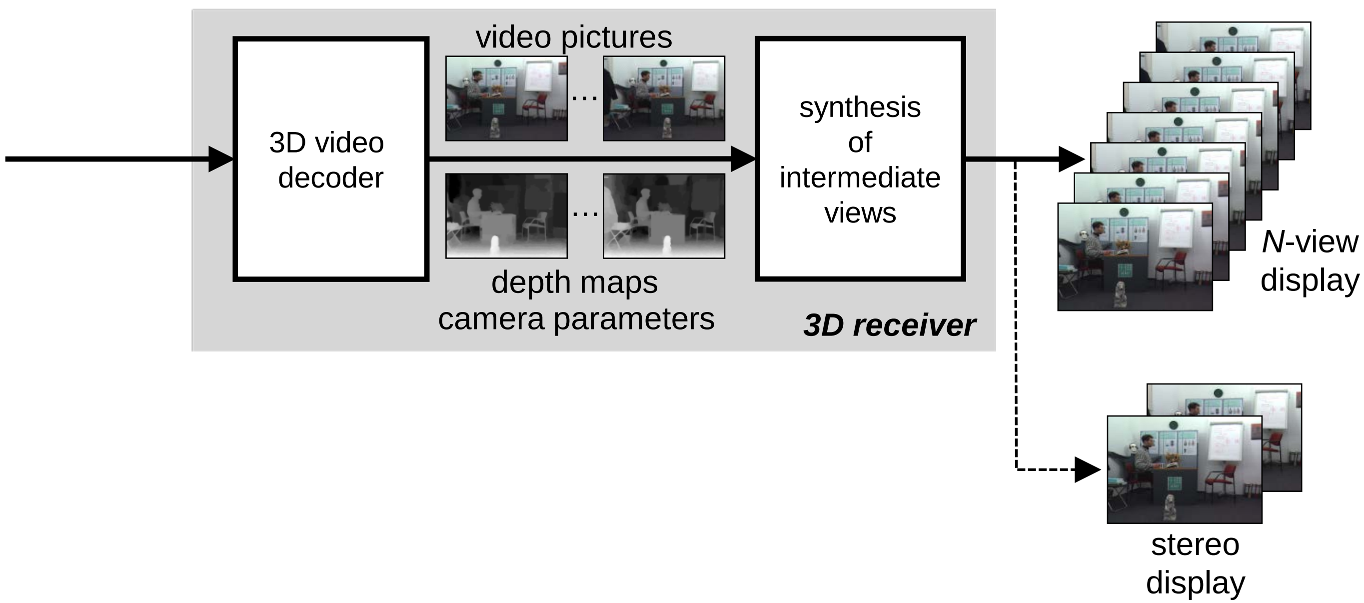
Inheritance of partitioning and motion data from co-located video block

- Block-adaptive signalling
- Use merge syntax: Insert as first entry in candidate list
- Only supported if complete co-located video block is inter-coded



M. Winken et al., “Motion vector inheritance for high efficiency 3D video plus depth coding,“ PCS 2012.
Courtesy of Fraunhofer HHI, Berlin, Germany.

Synthesized view distortion optimization

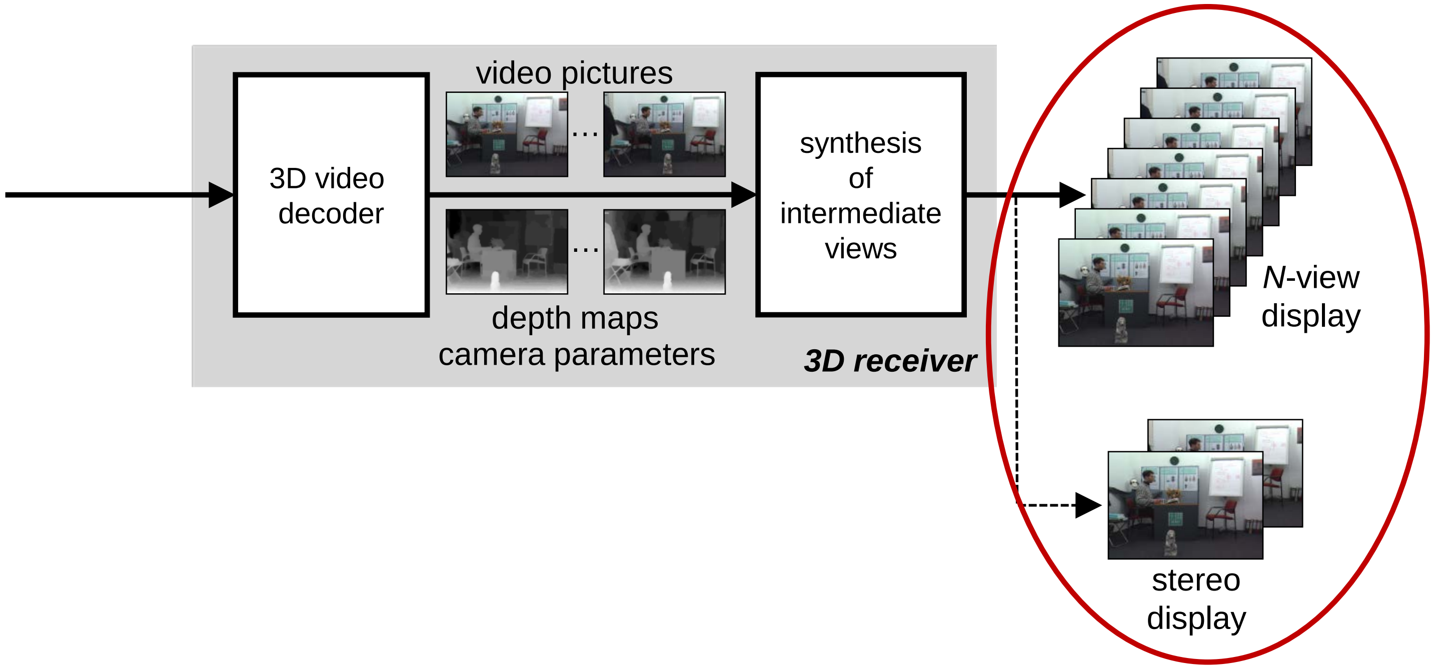


G. Tech et al., "3D video coding using the synthesized view distortion change," PCS 2012.
Courtesy of Fraunhofer HHI, Berlin, Germany.

Encoder control

Synthesized view distortion optimization

- Coding artifacts in depth data are only indirectly perceivable in synthesized video data
- Decoded depth map itself is not visible

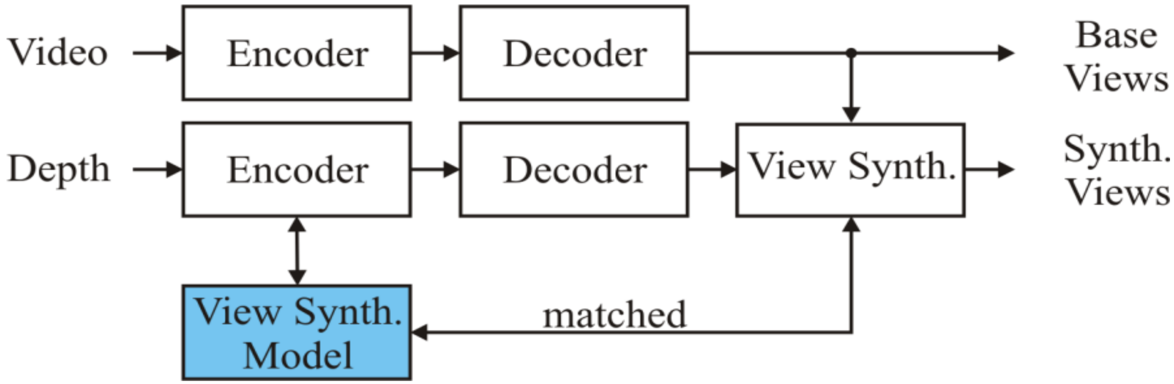


G. Tech et al., “3D video coding using the synthesized view distortion change,“ PCS 2012.
Courtesy of Fraunhofer HHI, Berlin, Germany.

Synthesized view distortion optimization

- Coding artifacts in depth data are only indirectly perceivable in synthesized video data
- Decoded depth map itself is not visible

→ Consider errors in synthesized views in encoder

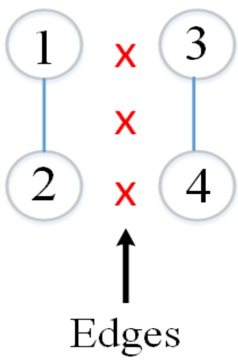


G. Tech et al., “3D video coding using the synthesized view distortion change,“ PCS 2012.
Courtesy of Fraunhofer HHI, Berlin, Germany.

Presentation Outline

- **Background & Motivation** (3D, not your mother's 2D)
- **3D Video representation / coding:**
  - Depth map coding
    - HEVC tools for depth maps
    - Graph-based Transform (GBT) for depth maps
  - Depth map denoising
  - Denoising + compression?
  - Why code depth images?
- **3D Video streaming:**
  - Video compression with flexible decoding for interactive streaming
  - Loss-resilient texture-plus-depth video streaming (skip)
- **3D view synthesis:**
  - Robust view synthesis for free viewpoint video
  - Synthesized image interpolation for *z*-dimension camera movement

Graph-Based Transform (GBT)



$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad D = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$L = \begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & -1 & 1 \end{bmatrix} \quad U = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

$\mathbf{A} = [a_{ij}]$ graph connectivity

$$\mathbf{D} = [d_{ij}], \; d_{ii} = \sum_j a_{ij}$$

$$\mathcal{L} = \mathbf{D} - \mathbf{A}$$

- An adaptive transform that avoids filtering across edges
- Equal to KLT under some specific statistic model when a_{ij} represents pixel correlation

G. Shen, W.-S. Kim, S.K. Narang, A. Ortega, J. Lee, and H. Wey, "Edge-adaptive transforms for efficient depth map coding," *IEEE Picture Coding Symposium*, Nagoya, Japan, December 2010.

D. Shuman, S.K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The Emerging Field of Signal Processing on Graphs," *IEEE Signal Processing Magazine*, pp.83-98, May 2013.

Depth Map Coding using Graph-Based Transform

- Depth map: Piecewise Smoothness (PWS)
- GBT gives compact compression for depth maps
  - sparse transform domain representation
    (avoid filtering across edges)
  - simple transform description
    (the statistics of depth maps is simple: pixel correlation is either 0 or 1)
- Example

a 4x4 block

GBT $\alpha_{\mathbf{1}} = \begin{bmatrix} 237 & 0 & 0 & 0 \\ 163 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ DCT $\alpha_{\mathbf{2}} = \begin{bmatrix} 285 & -29 & -5 & -4 \\ 16 & 1 & -16 & -4 \\ -5 & 3 & 5 & -7 \\ -1 & -4 & 1 & 9 \end{bmatrix}$

- Complexity issue: real-time eigen-decomposition, only operate on small blocks

Multi-resolution Graph-based Transform

- Objective: Encode *large* blocks with GBT in *low complexity*

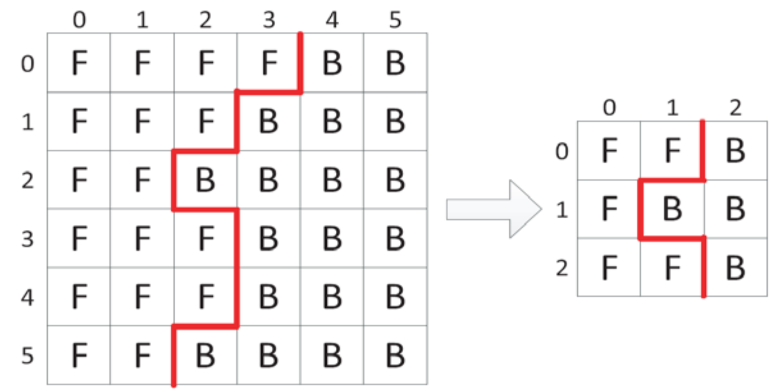
- Key Idea

  - Encode sharp edges in original high resolution:

    preserve sharpness

  - Encode smooth surfaces in low-pass-filtered and down-sampled low resolution:

    save bits & reduce complexity



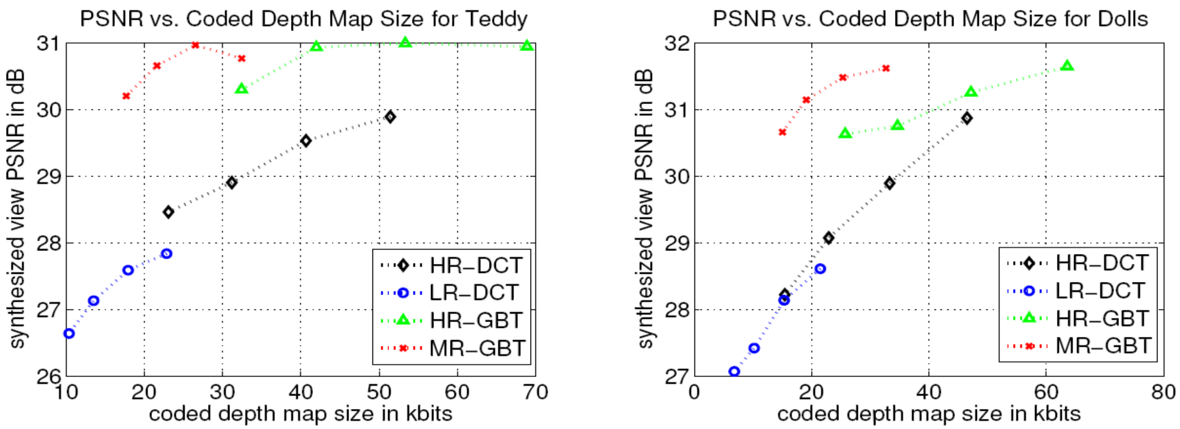
  - At the decoder, the LR surfaces are up-sampled and interpolated while respecting the losslessly encoded HR edges.

W. Hu, G. Cheung, X. Li and O. Au, "Depth Map Compression using Multi-resolution Graph-based Transform for Depth-image-based Rendering," *IEEE International Conference on Image Processing*, Orlando, FL, September 2012.

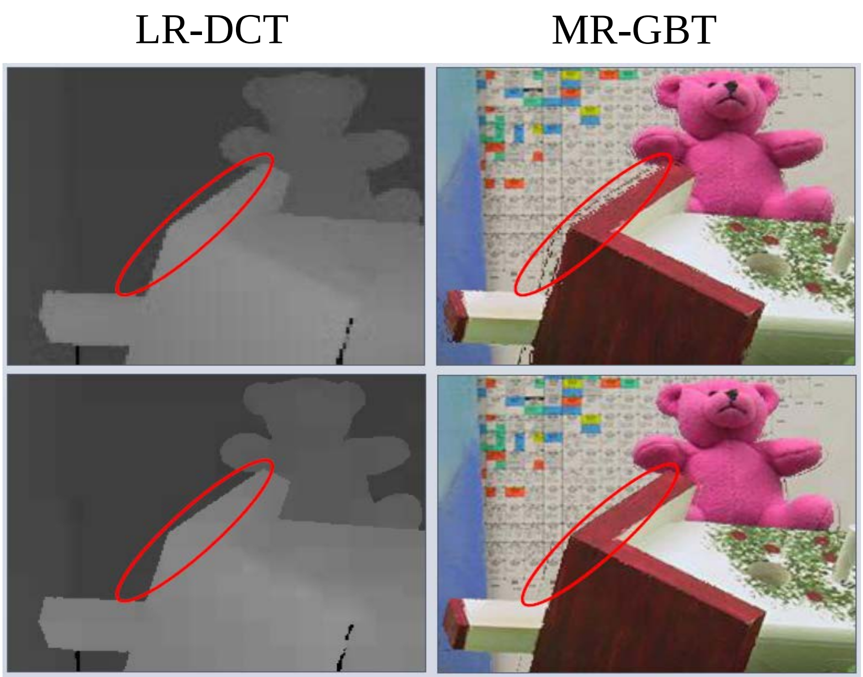
Experimentation

Experimental Setup

- H.264/AVC Reference Software JM17.1
- Test images: Middlebury multiview image sets
- QP: 24, 28, 32, 36
- Distortion metric: PSNR of synthesized views



(a) (b)



reduce bitrate by 68% compared to HR-DCT
and 55% compared to HR-GBT

Presentation Outline

- **Background & Motivation** (3D, not your mother's 2D)
- **3D Video representation / coding:**
  - Depth map coding
    - HEVC tools for depth maps
    - Graph-based Transform (GBT) for depth maps
  - Depth map denoising
  - Denoising + compression?
  - Why code depth images?
- **3D Video streaming:**
  - Video compression with flexible decoding for interactive streaming
  - Loss-resilient texture-plus-depth video streaming (skip)
- **3D view synthesis:**
  - Robust view synthesis for free viewpoint video
  - Synthesized image interpolation for *z*-dimension camera movement

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
  - $g(p)$: 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

spatial Gaussian

err Gaussian

color Gaussian

pixel p

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

dth bin

neighbors of pixel p

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

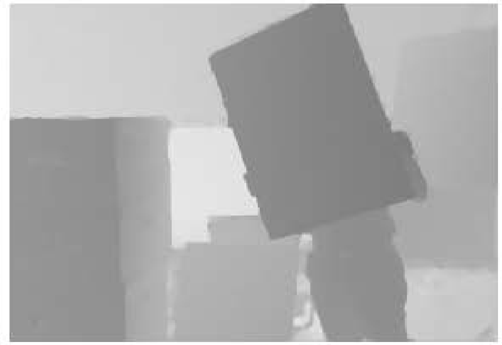
D. Min, J. Lu, and M. N. Do, "Depth video enhancement based on weighted mode filtering," IEEE Trans. on Image Processing, 2012.

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

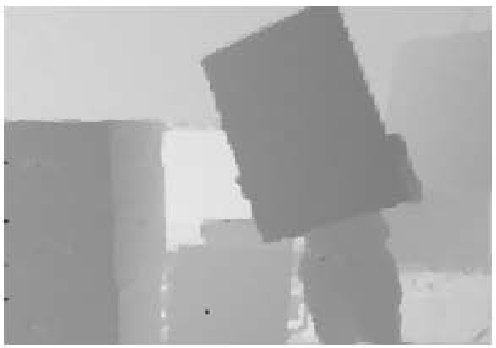
Result Comparison

(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).

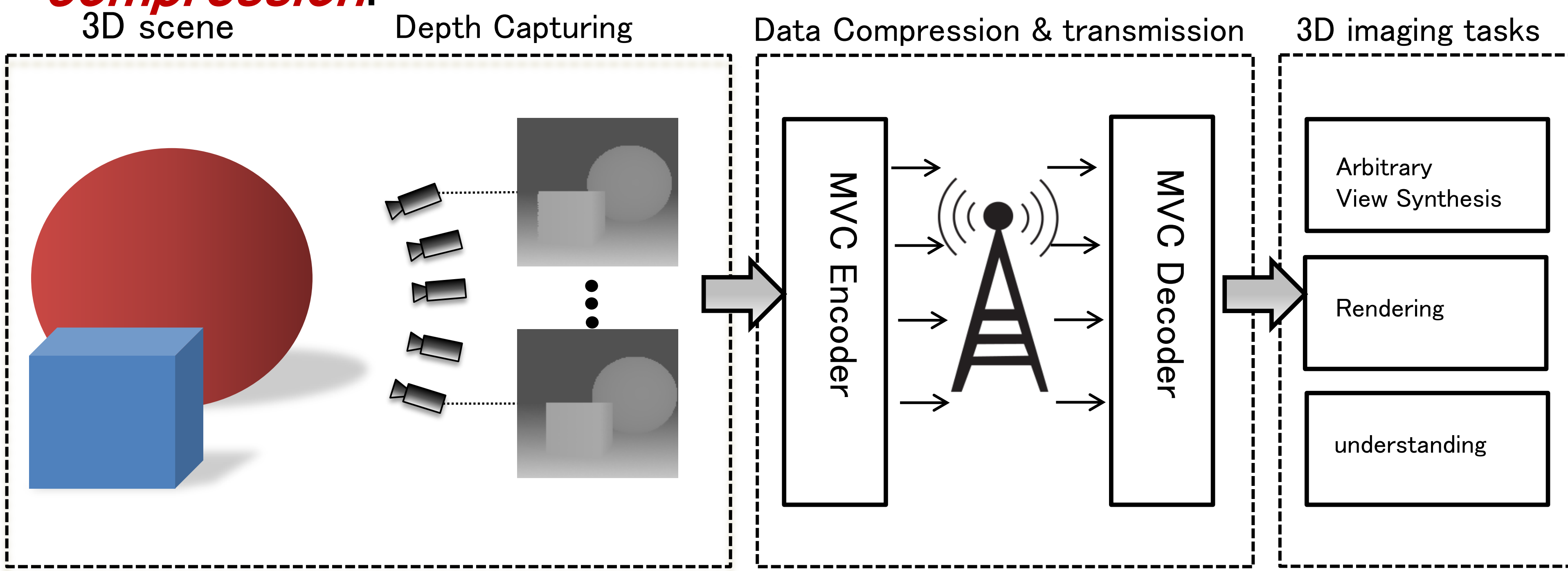
*Courtesy of Prof. M. Do, UIUC, USA

Presentation Outline

- **Background & Motivation** (3D, not your mother's 2D)
- **3D Video representation / coding:**
  - Depth map coding
    - HEVC tools for depth maps
    - Graph-based Transform (GBT) for depth maps
  - Depth map denoising
  - Denoising + compression?
  - Why code depth images?
- **3D Video streaming:**
  - Video compression with flexible decoding for interactive streaming
  - Loss-resilient texture-plus-depth video streaming (skip)
- **3D view synthesis:**
  - Robust view synthesis for free viewpoint video
  - Synthesized image interpolation for *z*-dimension camera movement

Depth Processing in 3D Video Communication

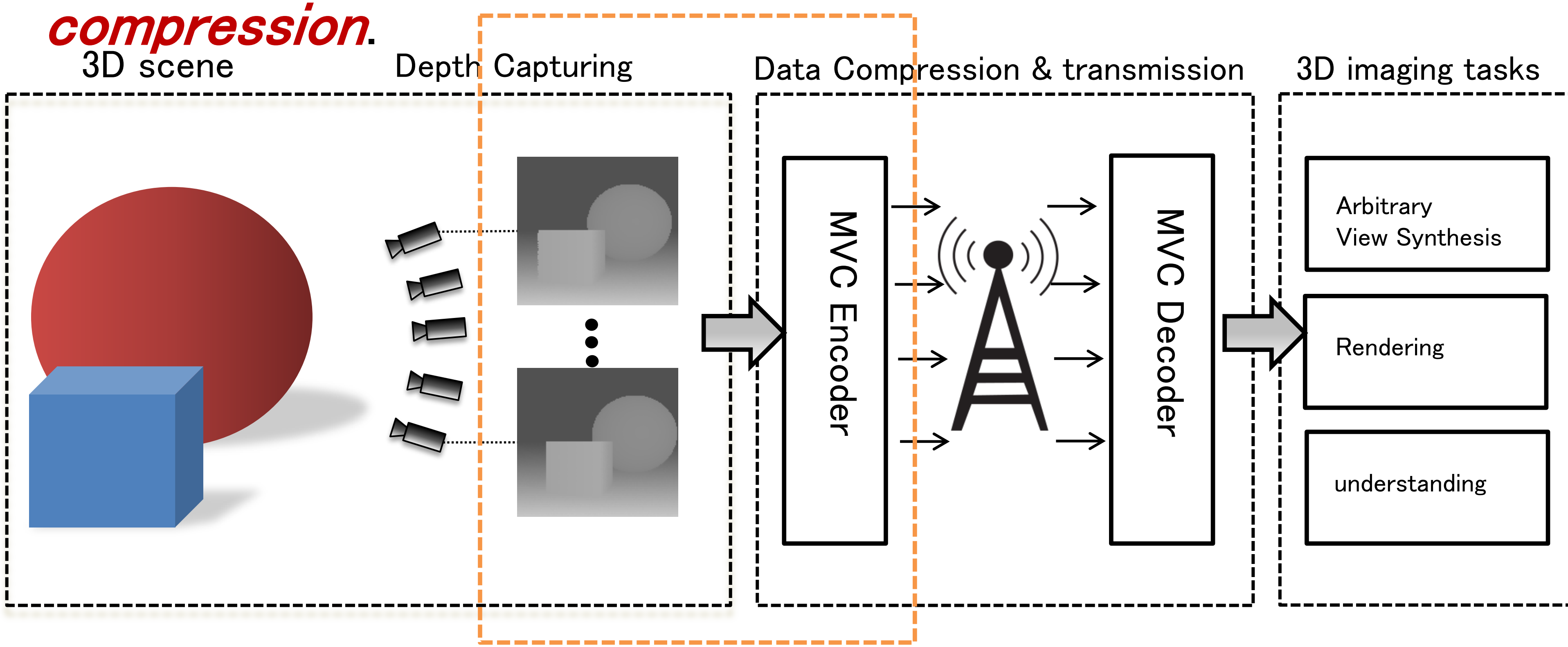
- Pipeline of 3D Video Communication System
- At encoder, ***depth processing*** means ***denoising*** & ***compression***.



*W. Sun et al, "Rate-distortion Optimized 3D Reconstruction from Noise-corrupted Multiview Depth Videos," ICME, 2013.

Depth Processing in 3D Video Communication

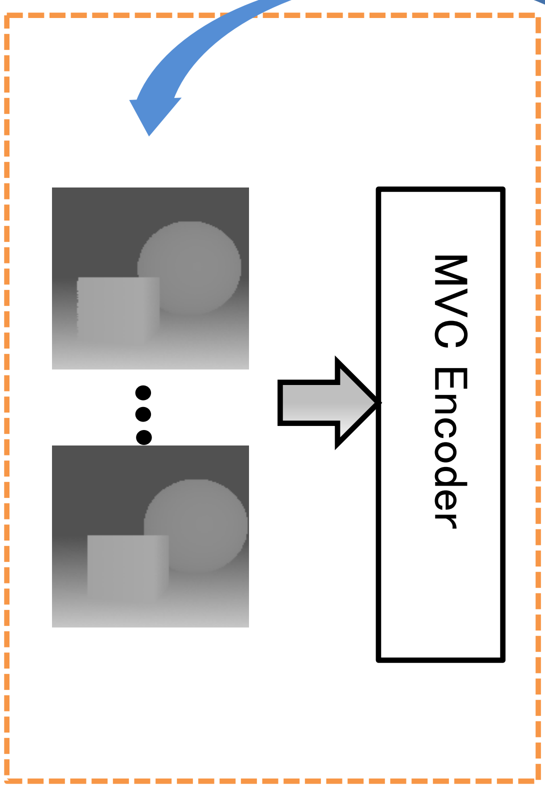
- Pipeline of 3D Video Communication System
- At encoder, ***depth processing*** means ***denoising*** & ***compression***.



*W. Sun et al, "Rate-distortion Optimized 3D Reconstruction from Noise-corrupted Multiview Depth Videos," ICME, 2013.

Two Practical Problems

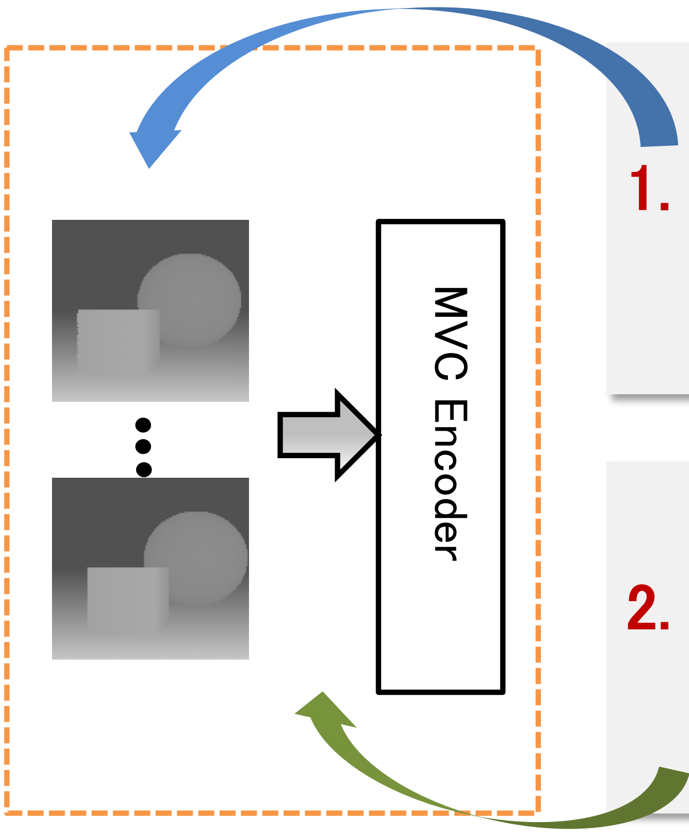
- Two related but different processing problems concerning depth maps (after acquisition):



1. **Denoising**
   - Acquired depth maps inherently noisy.

Two Practical Problems

- Two related but different processing problems concerning depth maps (after acquisition):

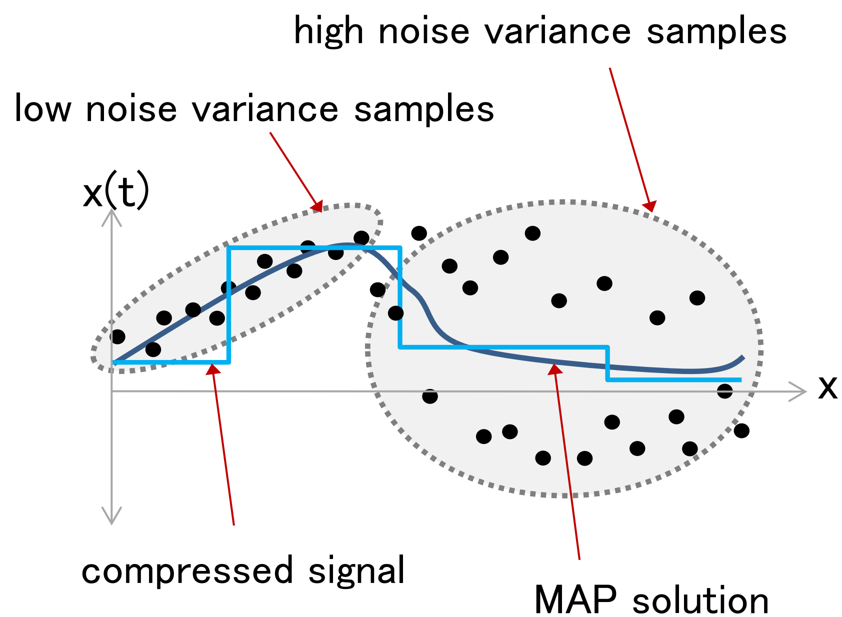


1. Denoising
- Acquired depth maps inherently noisy.

2. Compression
- Reduce transmission bandwidth.

Separate vs. Joint Approach

- Separate 2-step approach:
  1. Denoise depth maps optimally (e.g. MAP formulation) regardless of rep. size;
  2. compress computed MAP surface in deterministic way via conventional codec.



- Joint approach by performing denoising / compression as one:
  - Problem inherently probabilistic.
  - Can compress large noise variance samples aggressively.

Rate-constrained Estimation

- Given observed depth maps $\mathbf{y} = [y_1, y_2, \cdots]$, find optimal 3D surface $\mathbf{s}$.

Bayes Rule

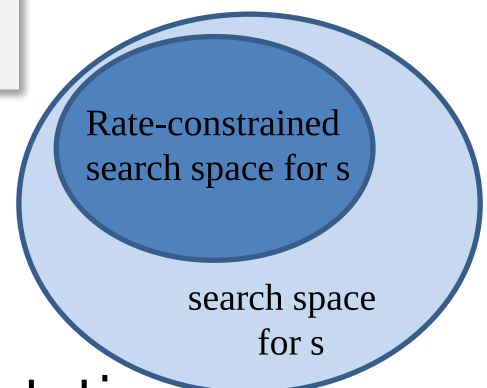
$$\Pr(s \mid y) = \frac{\Pr(y \mid s)\Pr(s)}{\Pr(y)}$$

Rate Term

Rate-constrained MAP

$$\max_{s} \Pr(y \mid s)\Pr(s) \quad \text{s.t.} \quad R(s) \le \overline{R}$$

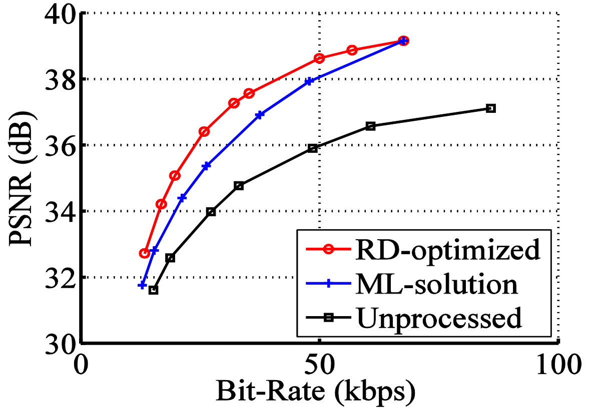
$$\min_{s} -\log \Pr(y \mid s) - \log \Pr(s) + \lambda R(s)$$



- **Distortion term:** select **s** to agree w/ observations.
- **Prior term**: select **s** to agree w/ prior.
- **Rate term**: select **s** that requires few bits for representation.

Experimentation

Up to 2.42dB gain in PSNR



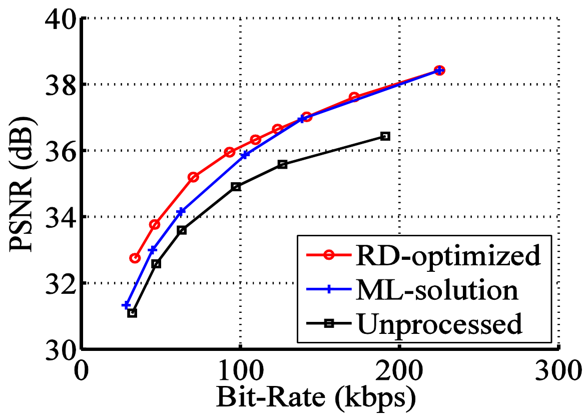


Fig. **PSNR of synthesized virtual views** at decoder versus coding rate for *Lovebird1 (top) and Balloons (bottom).*

Improved virtual view

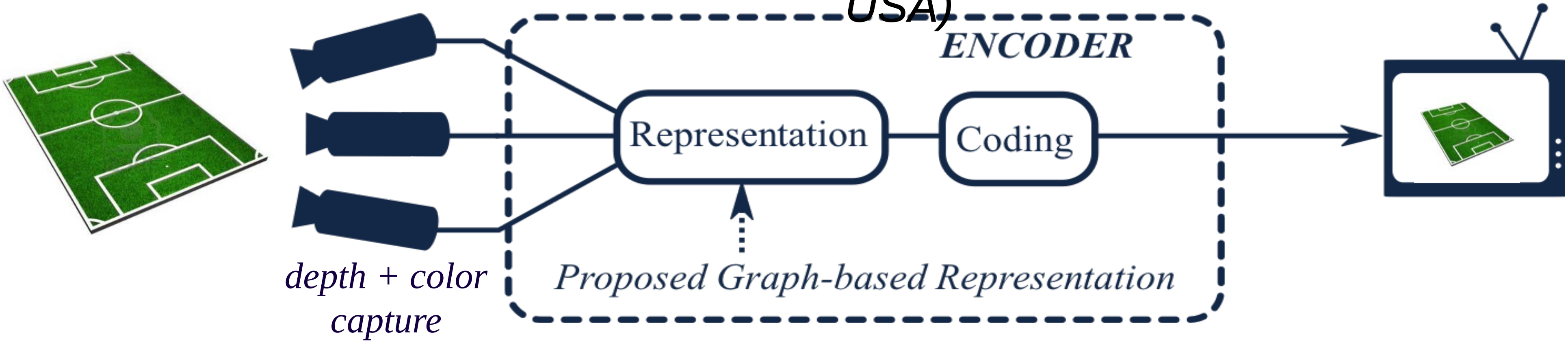
Fig. ***Top Row (Lovebird1):*** synthesized virtual view 5 using texture and depth maps at view 4 and 6. Depth maps are of 48kbps: Unprocessed (left), ML-solution (center), RD-optimized (right). ***Bottom Row (Balloons):*** synthesized virtual view 2 using texture depth maps at view 1 and 3. Depth maps are of 100kbps: Unprocessed (left), ML-solution (center), RD-optimized (right).

Presentation Outline

- **Background & Motivation** (3D, not your mother's 2D)
- **3D Video representation / coding:**
  - Depth map coding
    - HEVC tools for depth maps
    - Graph-based Transform (GBT) for depth maps
  - Depth map denoising
  - Denoising + compression?
  - Why code depth images?
- **3D Video streaming:**
  - Video compression with flexible decoding for interactive streaming
  - Loss-resilient texture-plus-depth video streaming (skip)
- **3D view synthesis:**
  - Robust view synthesis for free viewpoint video
  - Synthesized image interpolation for *z*-dimension camera movement

Graph-based representation

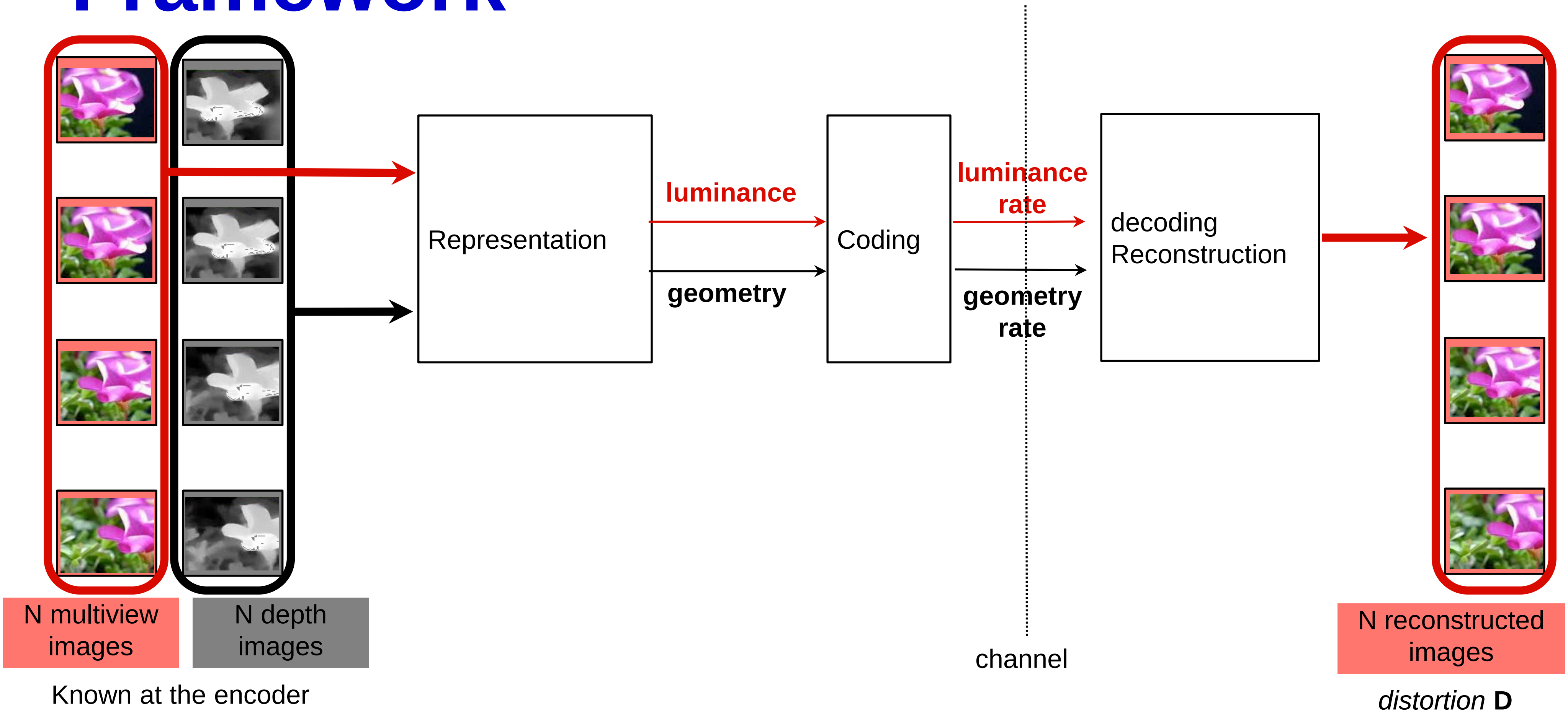
(Collaboration, Antonio Ortega, USC, USA)



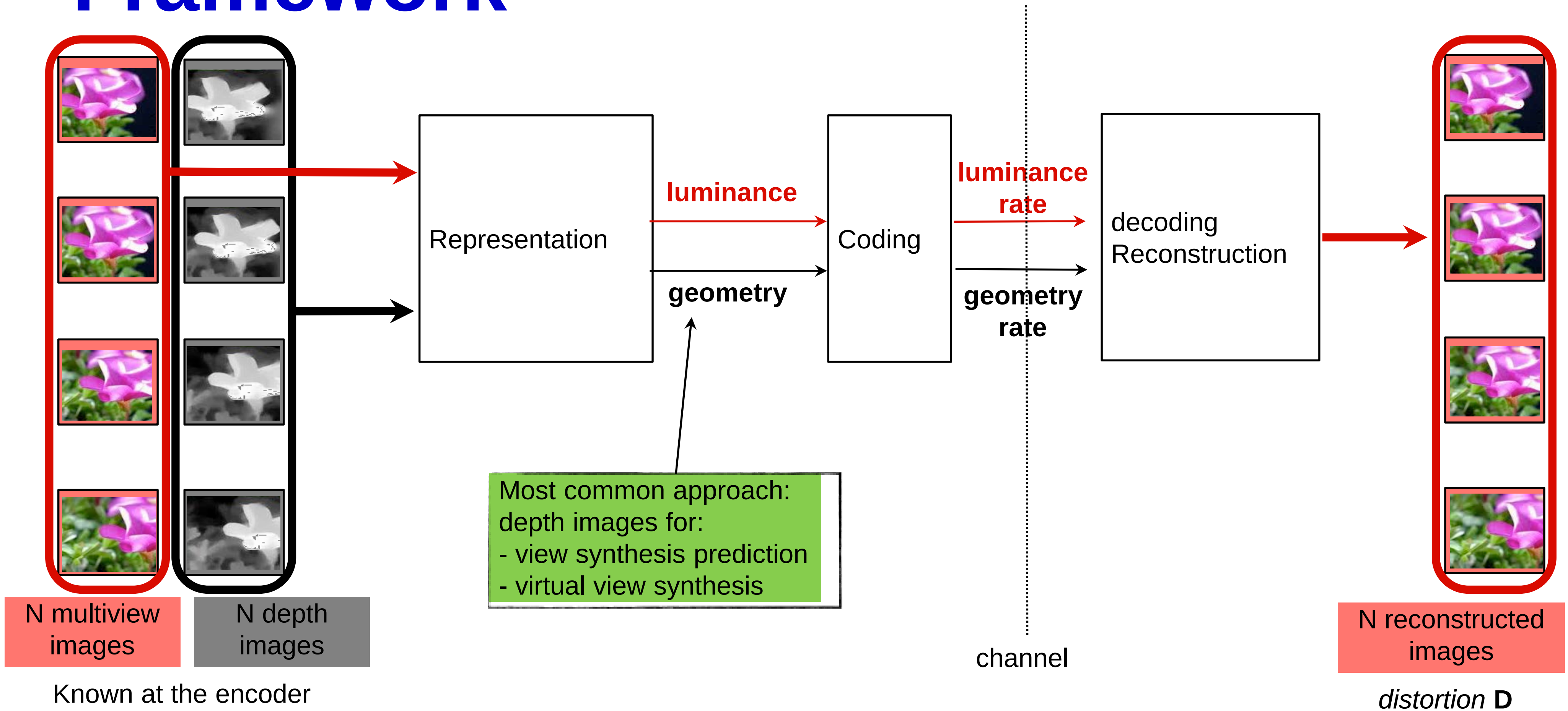
- Find an alternative to depth-based representation:

- Main idea
  - describe the inter-view pixel connections as links in a **graph**

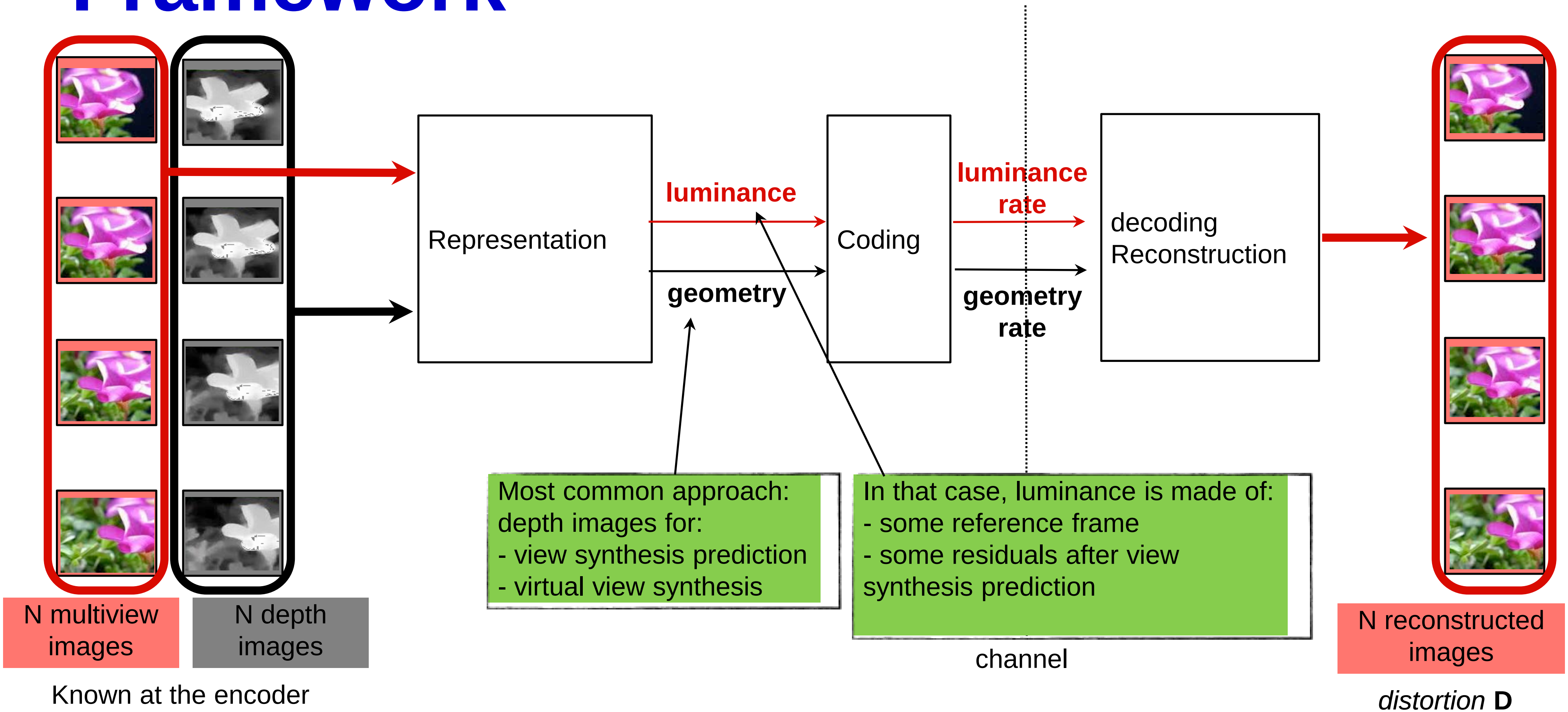
Framework


Representation
luminance
geometry
Coding
luminance
rate
geometry
rate
decoding
Reconstruction
N multiview images
N depth images
Known at the encoder
channel
N reconstructed images
distortion **D**

Framework

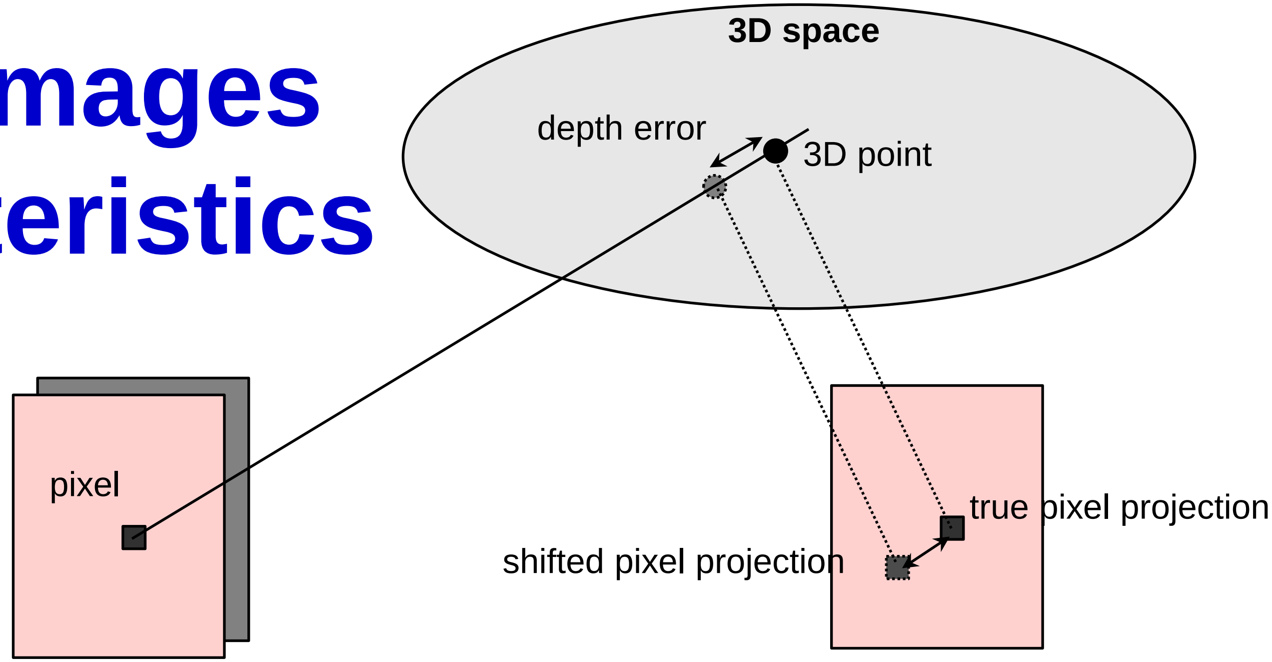

Representation
luminance
geometry
Coding
luminance
rate
geometry
rate
decoding
Reconstruction
Most common approach:
depth images for:
- view synthesis prediction
- virtual view synthesis
N multiview images
N depth images
Known at the encoder
channel
N reconstructed images
distortion **D**

Framework



Representation
luminance
geometry
Coding
luminance
rate
geometry
rate
decoding
Reconstruction
Most common approach:
depth images for:
- view synthesis prediction
- virtual view synthesis
In that case, luminance is made of:
- some reference frame
- some residuals after view
synthesis prediction
channel
N multiview
images
N depth
images
Known at the encoder
N reconstructed
images
distortion **D**

Depth images characteristics

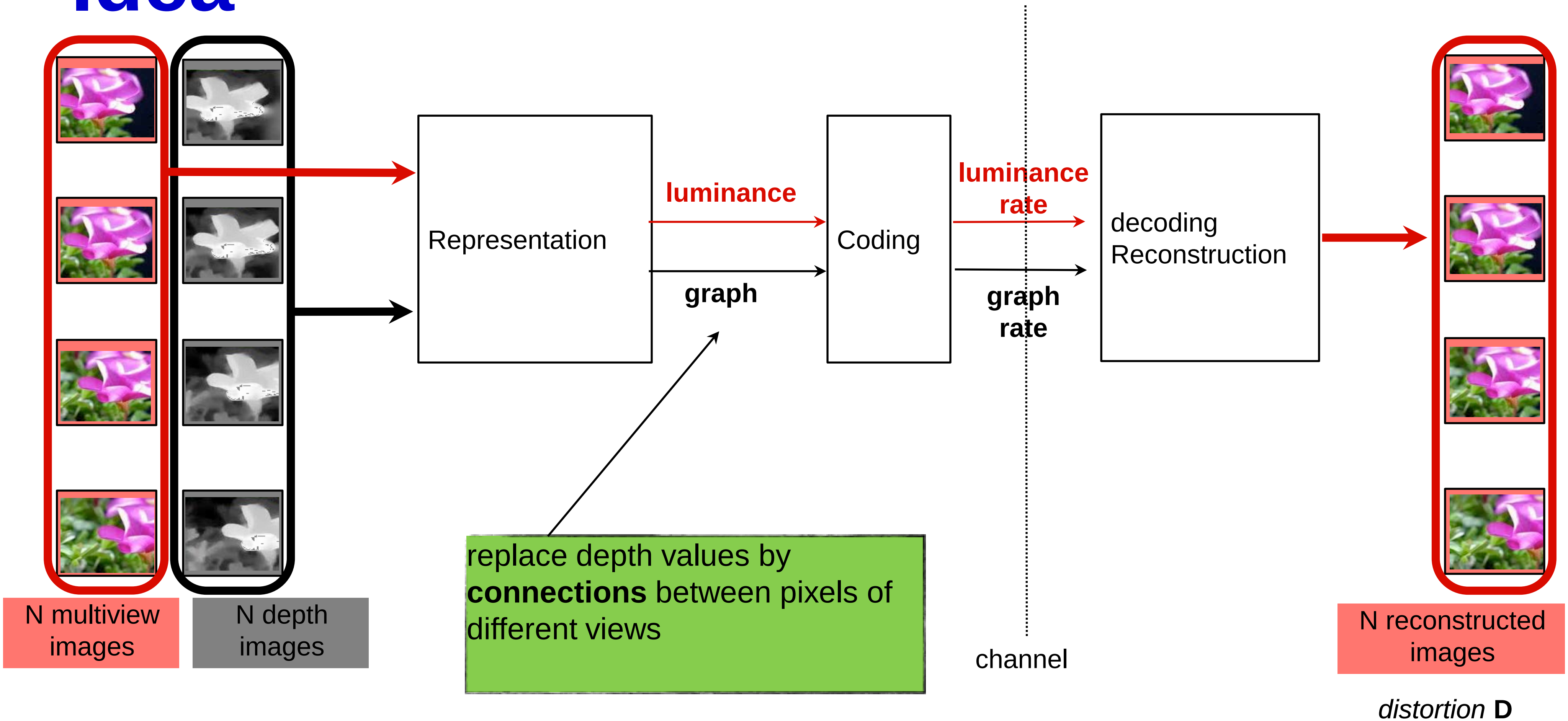


Captured luminance + depth images
viewpoint 1

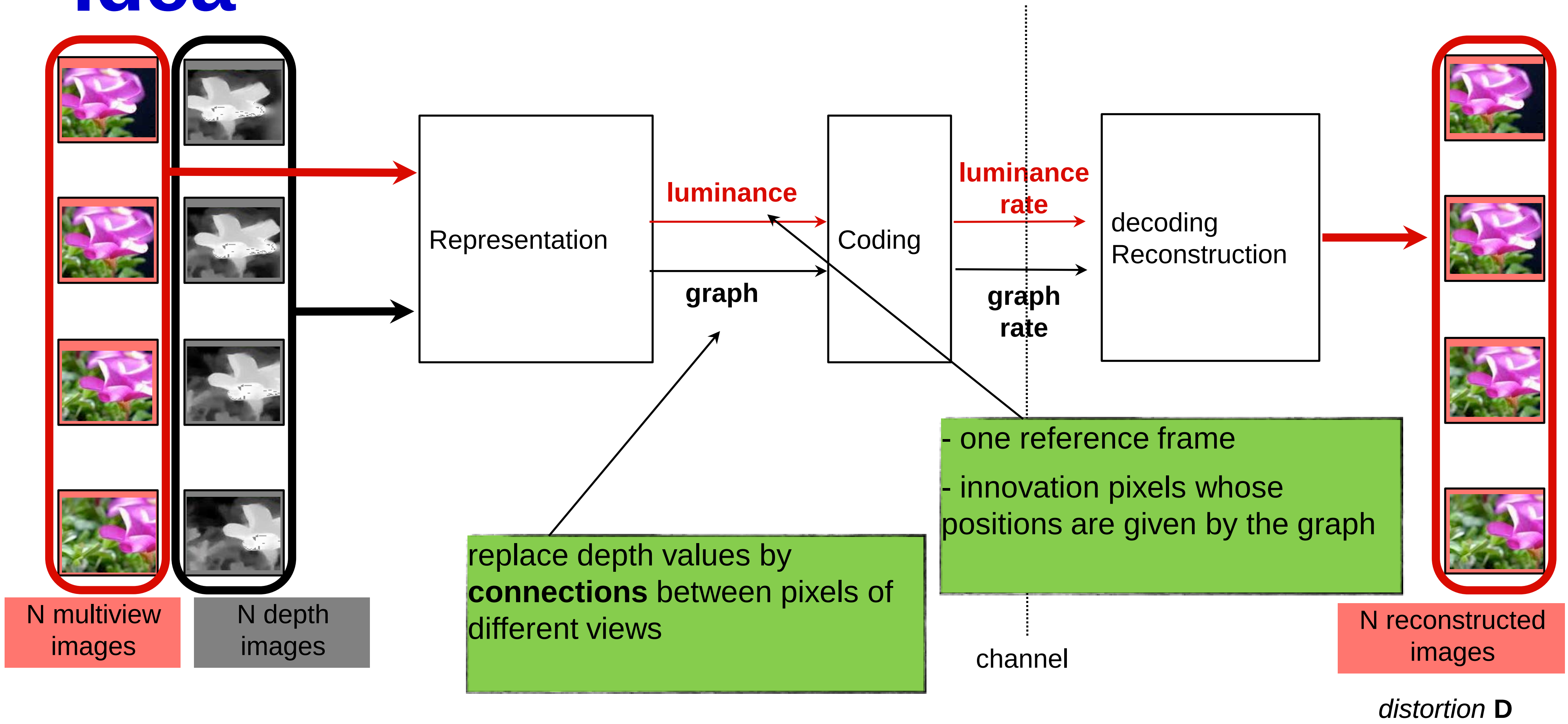
Interpolated image
viewpoint 2

- Depth-based schemes:
  - captured luminance and depth signals at several reference viewpoints
  - depth-based interpolation of intermediate viewpoint at decoder side
- Depth-based representation drawbacks:
  - an error in depth signal (estimation, compression) leads to spatial shift on the synthesized viewpoint
  - the induced error is difficult to model

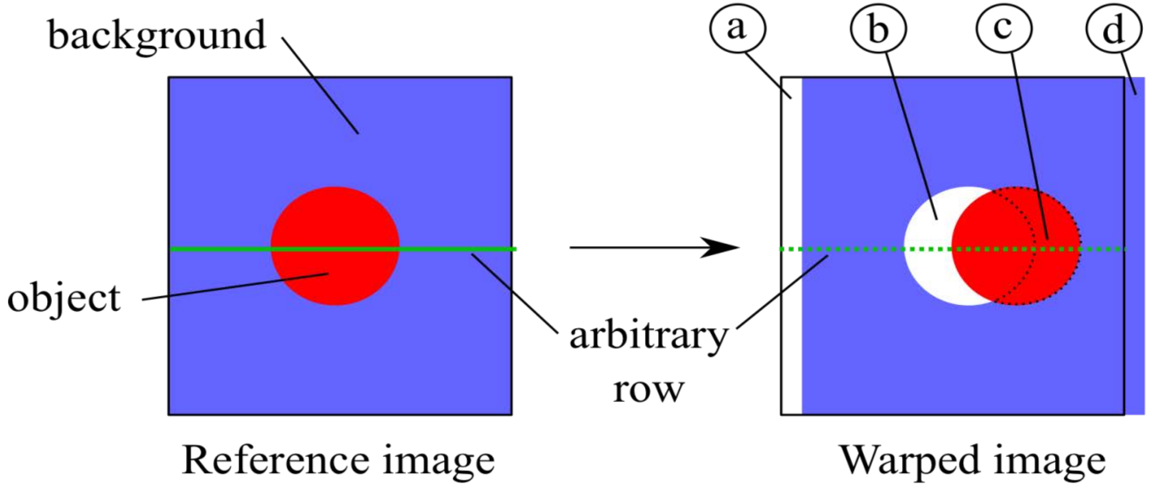
Idea


Representation
luminance
graph
Coding
luminance
rate
graph
rate
decoding
Reconstruction
channel
N multiview images
N depth images
replace depth values by **connections** between pixels of different views
N reconstructed images
distortion **D**

Idea


Representation
luminance
graph
Coding
luminance
rate
graph
rate
decoding
Reconstruction
N multiview
images
N depth
images
replace depth values by
connections between pixels of
different views
- one reference frame
- innovation pixels whose
positions are given by the graph
channel
N reconstructed
images
distortion **D**

Motivation: Pixel Classification

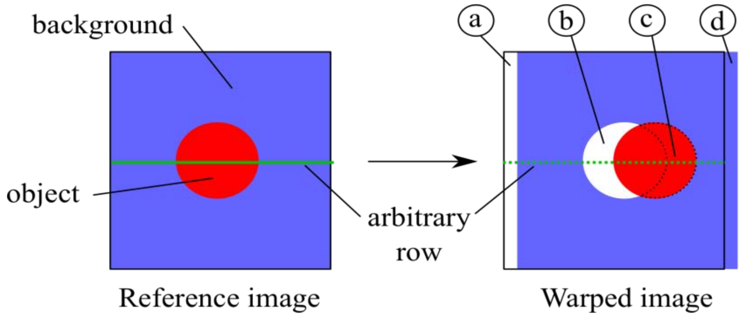


- Pixels categories
  - (a) : appearing pixels
  - (b) : disoccluded pixels
  - (c) : occluded pixels
  - (d) : disappearing pixels

- Warped image description
- links between these pixels and the reference image

- Proposed graph-based representation
- links back to previous frames
- OR explicit new pixels

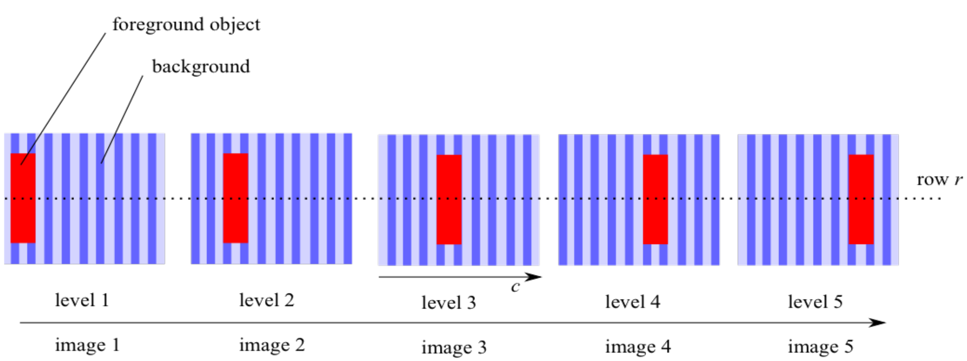
Graph-based representation

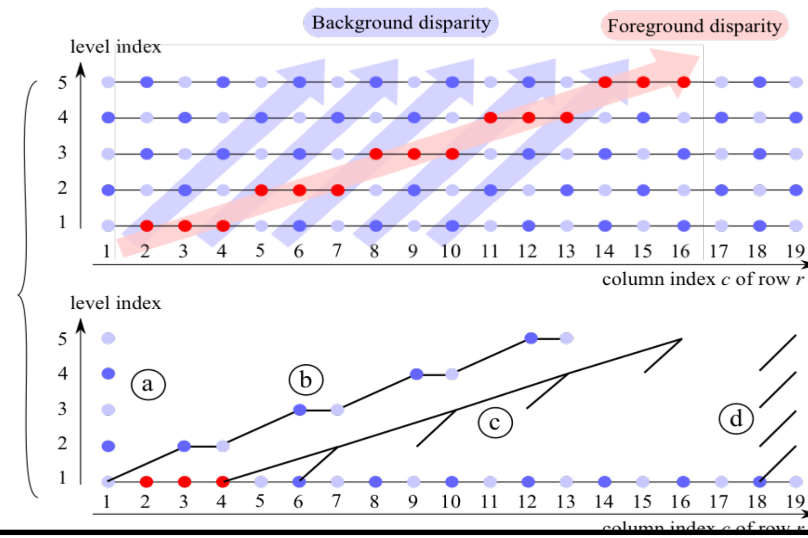


Describe right view with:
- a maximum of references to left view pixels
- Only « new » pixels

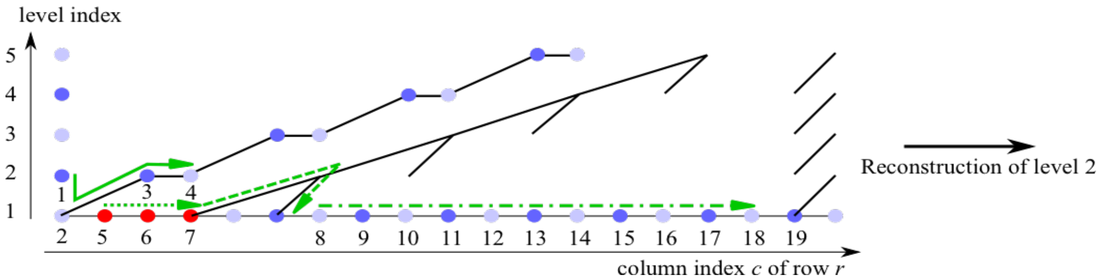
GRAPH RULES

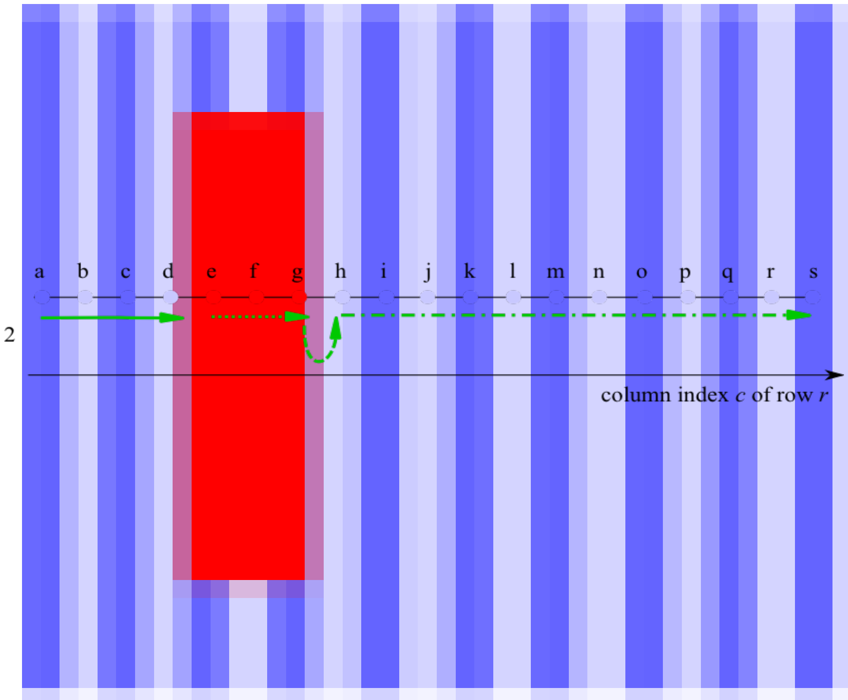
- Only new pixels appear in higher levels
- Connections link these pixels with their neighbor in the previous level
  - The (a) appearing and (b) occluded pixels are described *in the first image/level they appear*
  - The (c) disoccluded and (d) disappearing pixels are represented in the graph *by connections with no luminance values*





View reconstruction





- Reconstruction policy:
  - start at the level that is to be reconstructed and to fill all the appearing pixels
  - follow the connections to upper levels when they occur
  - go down to lower level when it is not possible to continue in the current level

Summary

- Graph links between views:
  - Provide a description of the geometry
  - Give an information of neighborhood between pixels
  - Permits a better control of compresion error

Summary:
3D Video Representation / Coding

- Geometry Representation of 3D scene for Image Synthesis at Receiver.
- Depth Images:
  - Piecewise smooth. Compact representation?
  - Auxiliary info. How to characterize err?
  - Joint denoising / compression?
- Graph-based representation?

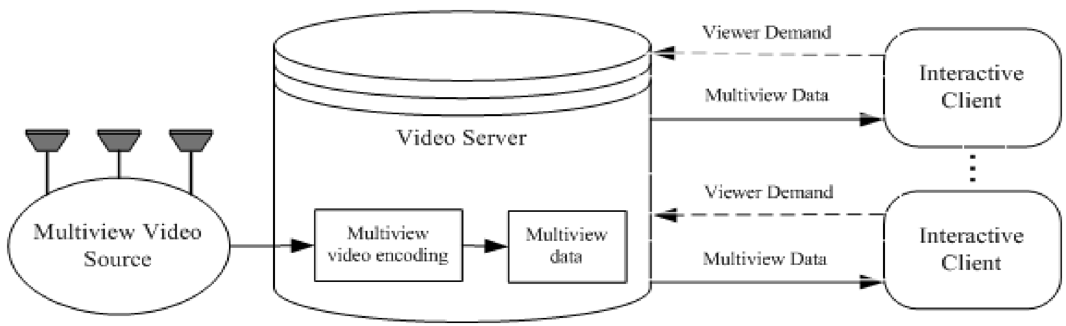
Presentation Outline

- **Background & Motivation** (3D, not your mother's 2D)
- **3D Video representation / coding:**
  - Depth map coding
    - HEVC tools for depth maps
    - Graph-based Transform (GBT) for depth maps
  - Depth map denoising
  - Denoising + compression?
  - Why code depth images?
- **3D Video streaming:**
  - Video compression with flexible decoding for interactive streaming
  - Loss-resilient texture-plus-depth video streaming (skip)
- **3D view synthesis:**
  - Robust view synthesis for free viewpoint video
  - Synthesized image interpolation for *z*-dimension camera movement

Background to Interactive Multiview Video Streaming

- **Multiview Video**
  - Closely spaced cameras capturing pictures periodically and synchronously.
  - The perception of depth via *motion parallax*.

- 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.

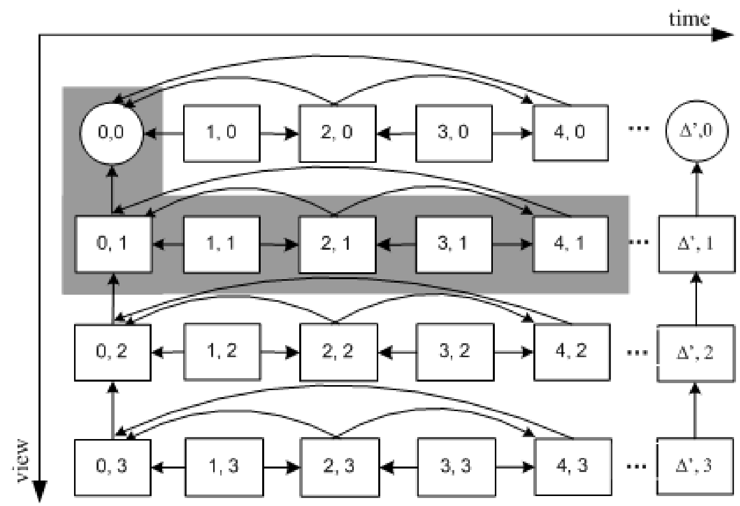


*G. Cheung, A. Ortega, N.-M. Cheung, "Interactive Streaming of stored multiview video Using redundant frame structures," *TIP*, March 2011.

Background to Interactive Multiview Video Streaming

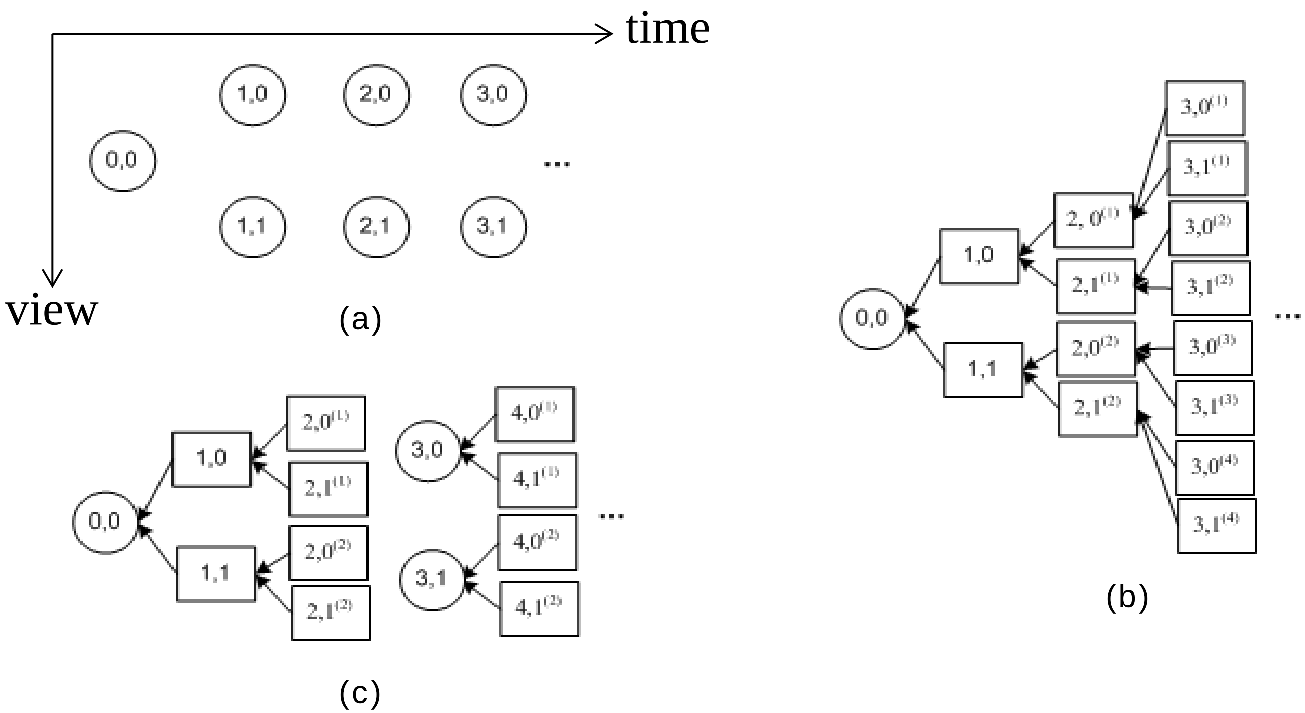
- Multiview Video Coding (MVC)
  - Strong correlation both in temporal and inter-view domains.
  - Efficiently encoding frames of all views in rate-distortion manner.

- Are MVC frame structures suitable for IMVS?
  - Insufficient decoding flexibility for interactive view-switching.
  - Multiple views transmitted but only one single view displayed.



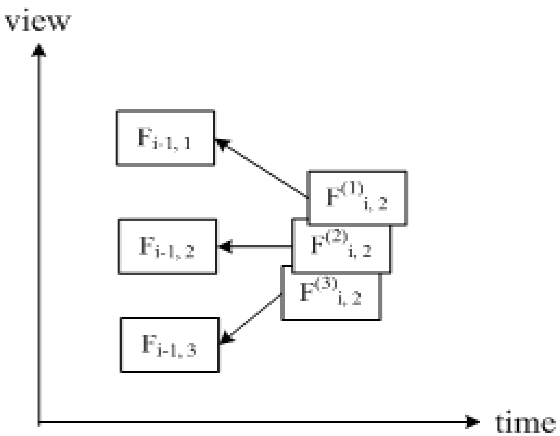
IMVS: 1st attempt w/ I + P-frames

- Frame Structure Optimization [G. Cheung MMSP'08, PV'09]
  - Using I- and P-frames, design ***redundant*** structures trading off transmission rate and storage.
  - Create multiple decoding paths for likely view transitions.

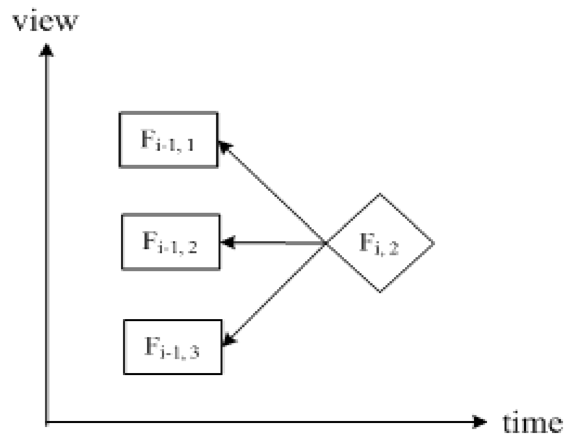


IMVS: 2nd attempt w/ *merge* frame

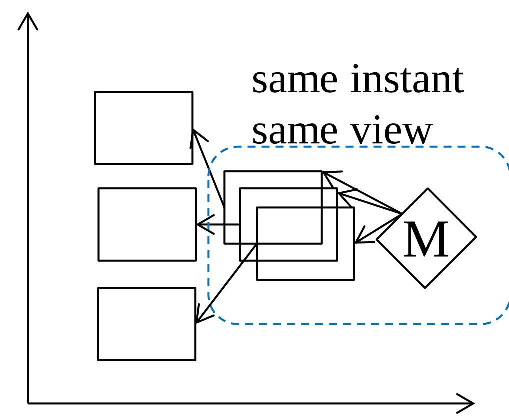
- Merge Frame (M-frame)
  - Identical reconstruction: an identical decoded frame for a set of possible predictors at streaming time.
  - Two novel DSC-based implementations of M-frame [N.-M. Cheung PCS'09, G. Cheung ICIP'09].
  - Application of M-frame in IMVS scenario, with superior performance over I-frame [G. Cheung TIP'11].



(a)



(b)

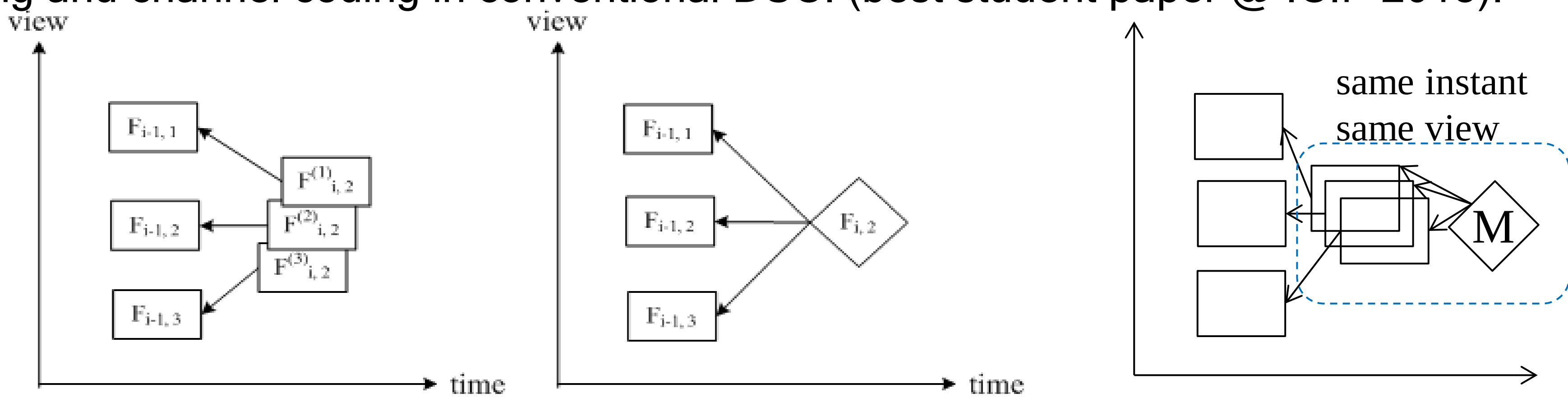


*W. Dai, G. Cheung, N.-M. Cheung, A. Ortega, O. Au, "Rate-Distortion Optimized Merge Frame Using Piecewise Constant Functions," *ICIP'13*.

IMVS: 2nd attempt w/ *merge* frame

- Merge Frame (M-frame)
  - Identical reconstruction: an identical decoded frame for a set of possible predictors at streaming time.
  - Two novel DSC-based implementations of M-frame [N.-M. Cheung PCS'09, G. Cheung ICIP'09].
  - Application of M-frame in IMVS scenario, with superior performance over I-frame [G. Cheung TIP'11].

Recent Advance: developed RD-optimal merge frame without bit-plane coding and channel coding in conventional DSC. (best student paper @ ICIP 2013).

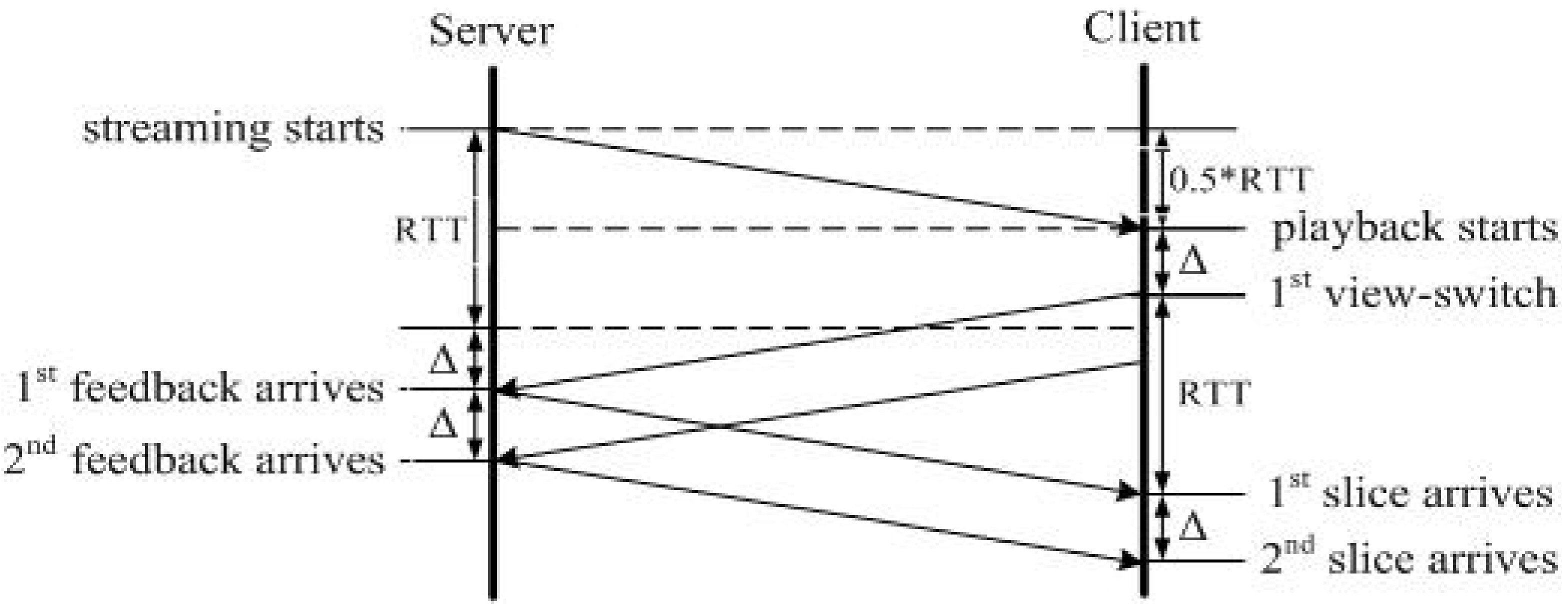


(a) (b)

*W. Dai, G. Cheung, N.-M. Cheung, A. Ortega, O. Au, "Rate-Distortion Optimized Merge Frame Using Piecewise Constant Functions," *ICIP'13*.

IMVS: 3rd attempt w/ network delay

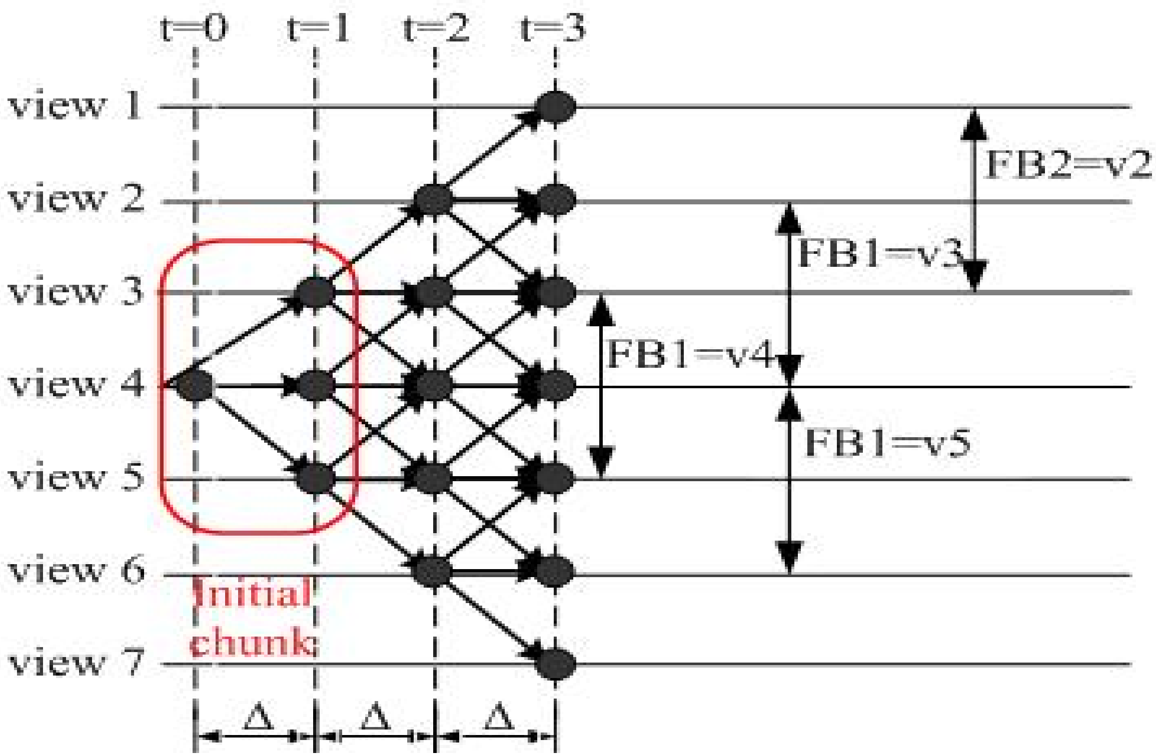
- IMVS with fixed network delay
  - Problem: view-switch request suffers one RTT delay.
  - Key idea: upon each feedback, additional data are sent to cover all view positions client could select when the data arrive at client.



*Xiaoyu Xiu, G. Cheung, A. Ortega, Jie Liang, "Delay-Cognizant Interactive Streaming of Multiview Video using Free Viewpoint Synthesis," *TMM*, March 2012.

IMVS: 3rd attempt w/ network delay

- IMVS with fixed network delay
  - Problem: view-switch request suffers one RTT delay.
  - Key idea: upon each feedback, additional data are sent to cover all view positions client could select when the data arrive at client.



*Xiaoyu Xiu, G. Cheung, A. Ortega, Jie Liang, "Delay-Cognizant Interactive Streaming of Multiview Video using Free Viewpoint Synthesis," *TMM*, March 2012.

Summary:
3D Video Streaming

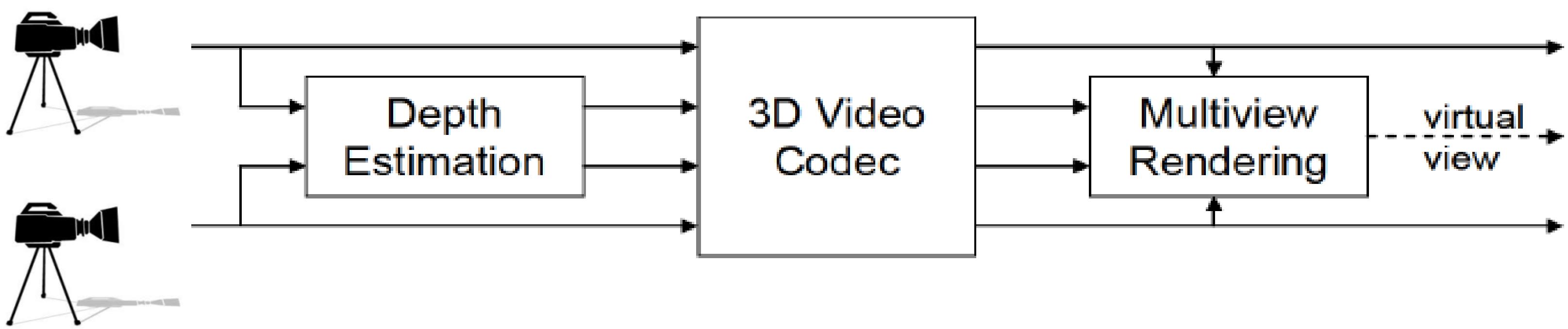
- High dimensional media navigation problem
- Asymmetric info:
  - Sender knows statistical model for navigation.
  - Receiver knows exact navigation path.
- Compression with decoding flexibility

Presentation Outline

- **Background & Motivation** (3D, not your mother's 2D)
- **3D Video representation / coding:**
  - Depth map coding
    - HEVC tools for depth maps
    - Graph-based Transform (GBT) for depth maps
  - Depth map denoising
  - Denoising + compression?
  - Why code depth images?
- **3D Video streaming:**
  - Video compression with flexible decoding for interactive streaming
  - Loss-resilient texture-plus-depth video streaming (skip)
- **3D view synthesis:**
  - Robust view synthesis for free viewpoint video
  - Synthesized image interpolation for *z*-dimension camera movement

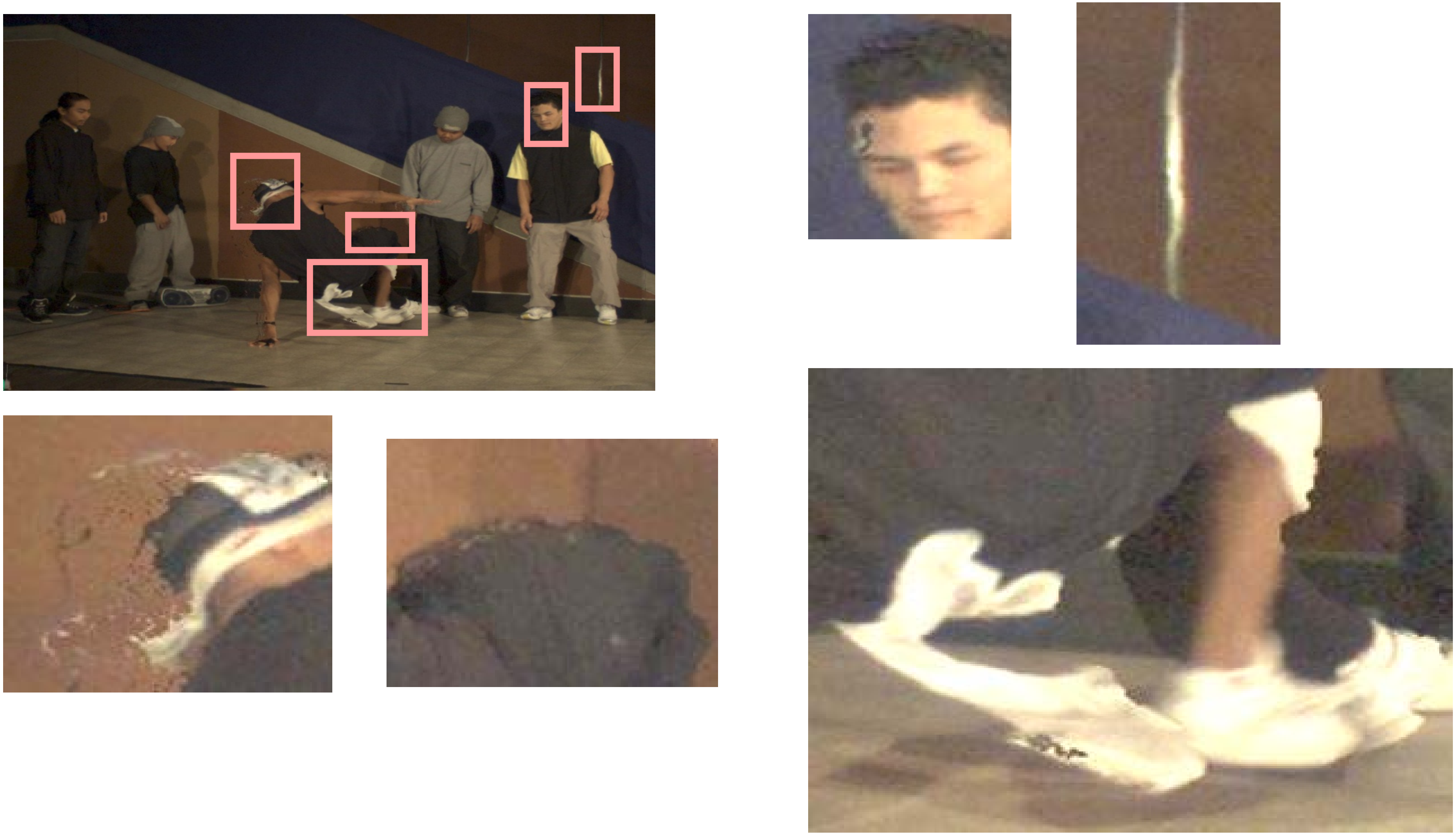
Conversion (FTV context)

- Lossy Conversion
  - Depth Image Based Rendering (DIBR)
  - Depth Estimation from single or multiple viewpoints



*Courtesy of Prof. P. Le Callet, UofNantes, France

DIBR: artefacts



*Courtesy of Prof. P. Le Callet, UofNantes, France

DIBR: current quality metrics are useless

	PSNR	SSIM	MSSIM	VSNR	VIF	VIFP	UQI	IFC	NQM	WSNR	PSNR HVSM	PSNR HVS
CC	40.5	-14.5	-8.9	-21.2	-22.3	-22.9	-25.9	-22.8	42.6	39.6	37.1	37.1

Table 3. Correlation coefficients between subjective and objective scores in percentage.

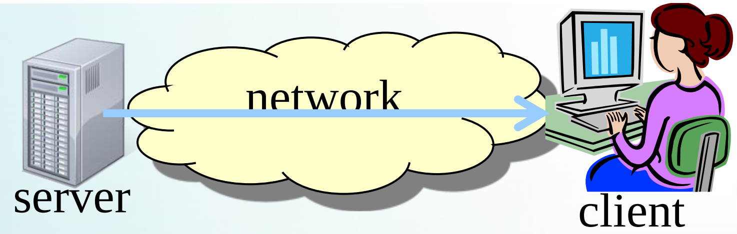
	PSNR	SSIM	MSSIM	VSNR	VIF	VIFP	UQI	IFC	NQM	WSNR	PSNR hsvm	PSNR hsv
PSNR		83.9	79.6	87.3	77.0	70.6	53.6	71.6	95.2	98.2	99.2	99.0
SSIM	83.9		96.7	93.9	93.4	92.4	81.5	92.9	84.9	83.7	83.2	83.5
MSSIM	79.6	96.7		89.7	88.8	90.2	86.3	89.4	85.6	81.1	77.9	78.3
VSNR	87.3	93.9	89.7		87.9	83.3	71.9	84.0	85.3	85.5	86.1	85.8
VIF	77.0	93.4	88.8	87.9		97.5	75.2	98.7	74.4	78.1	79.4	80.2
VIFP	70.6	92.4	90.2	83.3	97.5		85.9	99.2	73.6	75.0	72.2	72.9
UQI	53.6	81.5	86.3	71.9	75.2	85.9		81.9	70.2	61.8	50.9	50.8
IFC	71.6	92.9	89.4	84.0	98.7	99.2	81.9		72.8	74.4	73.5	74.4
NQM	95.2	84.9	85.6	85.3	74.4	73.6	70.2	72.8		97.1	92.3	91.8
WSNR	98.2	83.7	81.1	85.5	78.1	75.0	61.8	74.4	97.1		97.4	97.1
PSNR hsvm	99.2	83.2	77.9	86.1	79.4	72.2	50.9	73.5	92.3	97.4		99.9
PSNR hsv	99.0	83.5	78.3	85.8	80.2	72.9	50.8	74.4	91.8	97.1	99.9	

Table 2. Correlation coefficients between objective metrics in percentage.

Towards a new quality metric for 3D synthesized views assessment – in IEEE ICIP 2011
Emilie Bosc, R.pépion, P. Le Callet, M. Köppel, P. Ndjiki-Nya, M. Pressigout, L. Morin

*Courtesy of Prof. P. Le Callet, UofNantes, France

Saliency-based Error Concealment

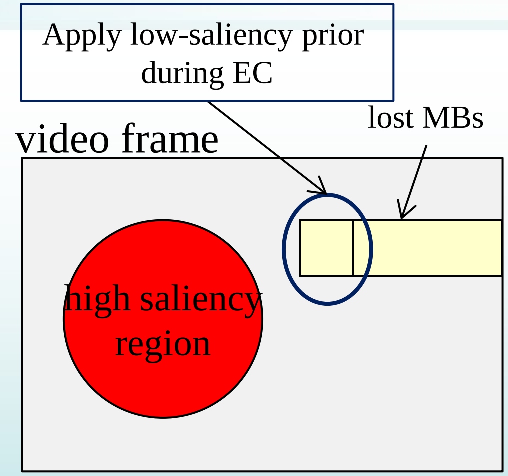


Goal: Packets are dropped in network during video streaming. Reconstruct a missing pixel block ***b*** by minimizing some cost function:

$$\min_{\boldsymbol{b}} fit\_err(\boldsymbol{b})$$

Problem: The problem is *under-determined*.

Solution: Add a *convex* **saliency term** as follows:



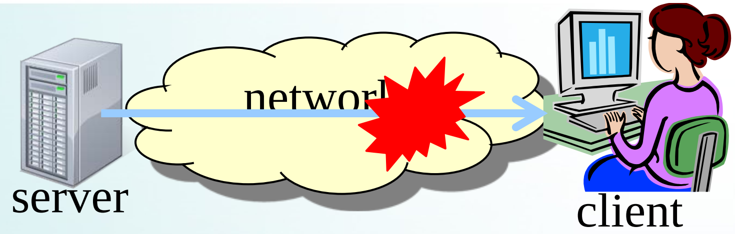
$$\min_{\boldsymbol{b}} \{fit\_err(\boldsymbol{b}) + \lambda\, saliency(\boldsymbol{b})\}$$

Advantages:

1. Potential wrong candidates become **less attention-grabbing**.
2. It serves as **a true prior** in an ROI-based streaming application.

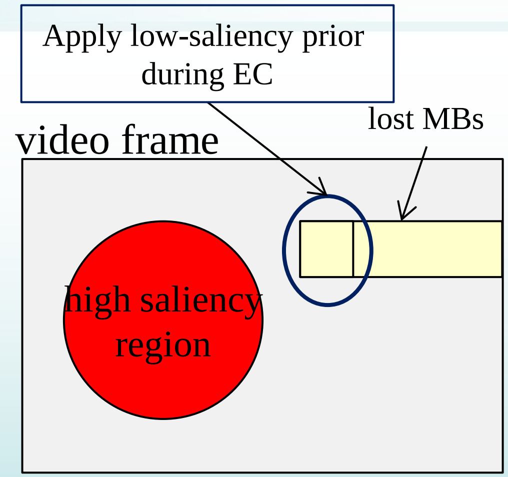
H. Hadizadeh, I. Bajic, G. Cheung, "**Saliency-cognizant Error Concealment in Loss-corrupted Streaming Video**", *ICME'2012* (**Best paper runner-up award**), "**Video Error Concealment Using a Computation-efficient Low Saliency Prior**," accepted to *TMM*, June 2013.

Saliency-based Error Concealment



Goal: Packets are dropped in network during video streaming. Reconstruct a missing pixel block ***b*** by minimizing some cost function:

$$\min_{\boldsymbol{b}} fit\_err(\boldsymbol{b})$$



Problem: The problem is *under-determined*.

Solution: Add a *convex* **saliency term** as follows:

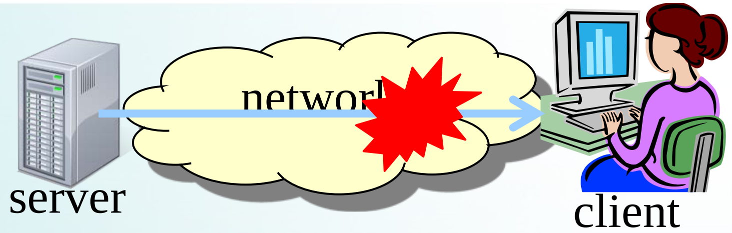
$$\min_{\boldsymbol{b}} \{ fit\_err(\boldsymbol{b}) + \lambda\, saliency(\boldsymbol{b}) \}$$

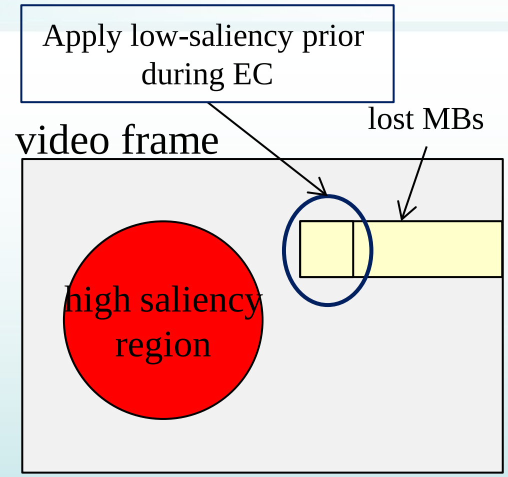
Advantages:

1. Potential wrong candidates become **less attention-grabbing**.
2. It serves as **a true prior** in an ROI-based streaming application.

H. Hadizadeh, I. Bajic, G. Cheung, "**Saliency-cognizant Error Concealment in Loss-corrupted Streaming Video**", *ICME'2012* (**Best paper runner-up award**), "**Video Error Concealment Using a Computation-efficient Low Saliency Prior**," accepted to *TMM*, June 2013.

Saliency-based Error Concealment





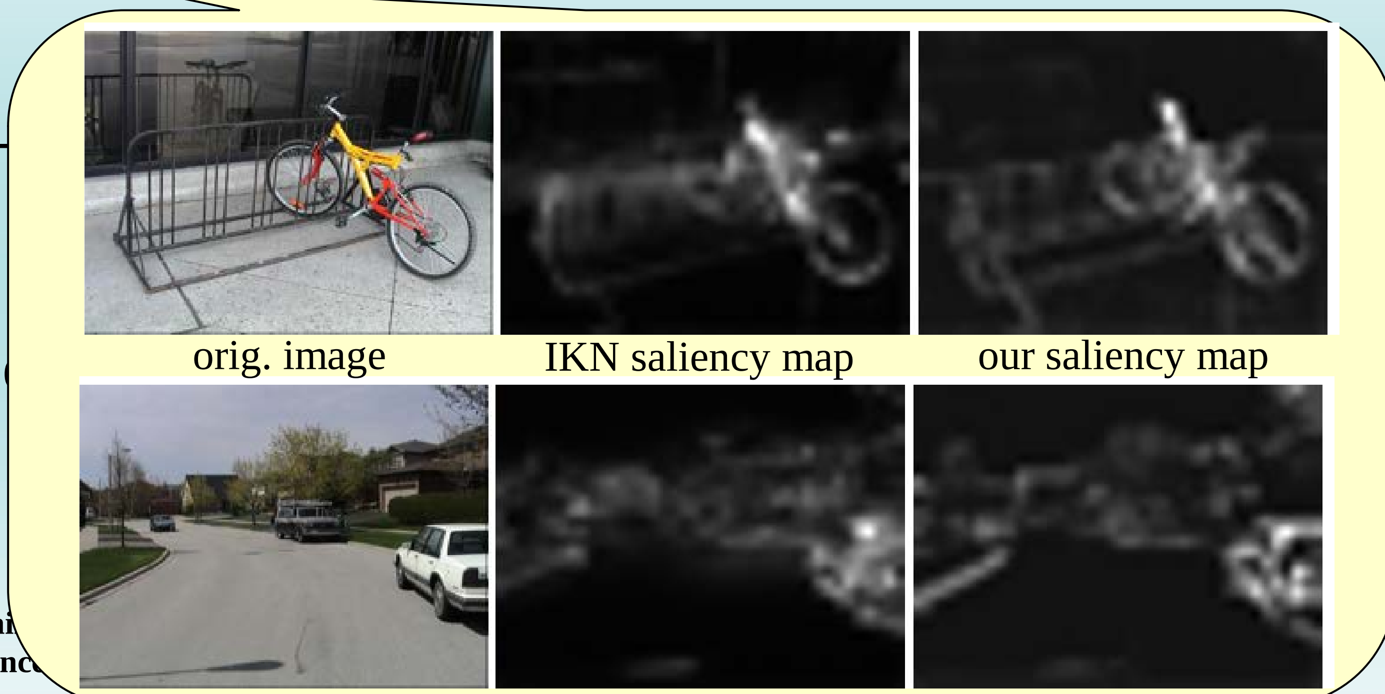
Goal: Packets are dropped in network during video streaming. Reconstruct a missing pixel block ***b*** by minimizing some cost function:

$$\min_{\boldsymbol{b}} fit\_err(\boldsymbol{b})$$

Problem: The problem is *under-determined*.

Solution: Add a *convex* **saliency term** as follows:

$$\min_{\boldsymbol{b}}\{fit\_err(\boldsymbol{b}) -$$

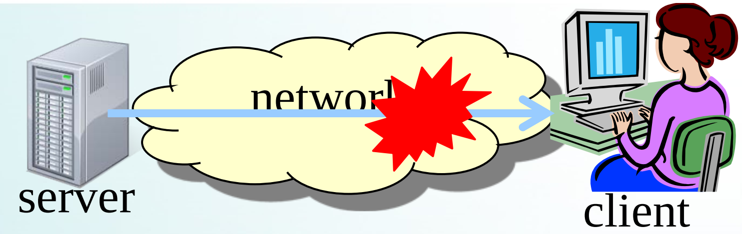


Advantages:

1. Potential wrong candi
2. It serves as **a true prior**

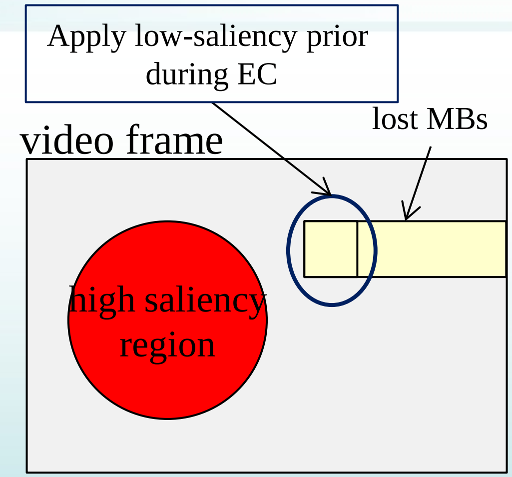
H. Hadizadeh, I. Bajic, G. Cheung, "**Saliency-cogni**
(**Best paper runner-up award**), "**Video Error Conc**

Saliency-based Error Concealment



Goal: Packets are dropped in network during video streaming. Reconstruct a missing pixel block ***b*** by minimizing some cost function:

$$\min_{\boldsymbol{b}} fit\_err(\boldsymbol{b})$$



Problem: The problem is *under-determined*.

Solution: Add a *convex* **saliency term** as follows:

$$\min_{\boldsymbol{b}} \{ fit\_err(\boldsymbol{b}) + \lambda \, saliency(\boldsymbol{b}) \}$$

Advantages:

1. Potential wrong candidates become **less attention-grabbing**.
2. It serves as **a true prior** in an ROI-based streaming application.

H. Hadizadeh, I. Bajic, G. Cheung, "**Saliency-cognizant Error Concealment in Loss-corrupted Streaming Video**", *ICME'2012* (**Best paper runner-up award**), "**Video Error Concealment Using a Computation-efficient Low Saliency Prior**," accepted to *TMM*, June 2013.

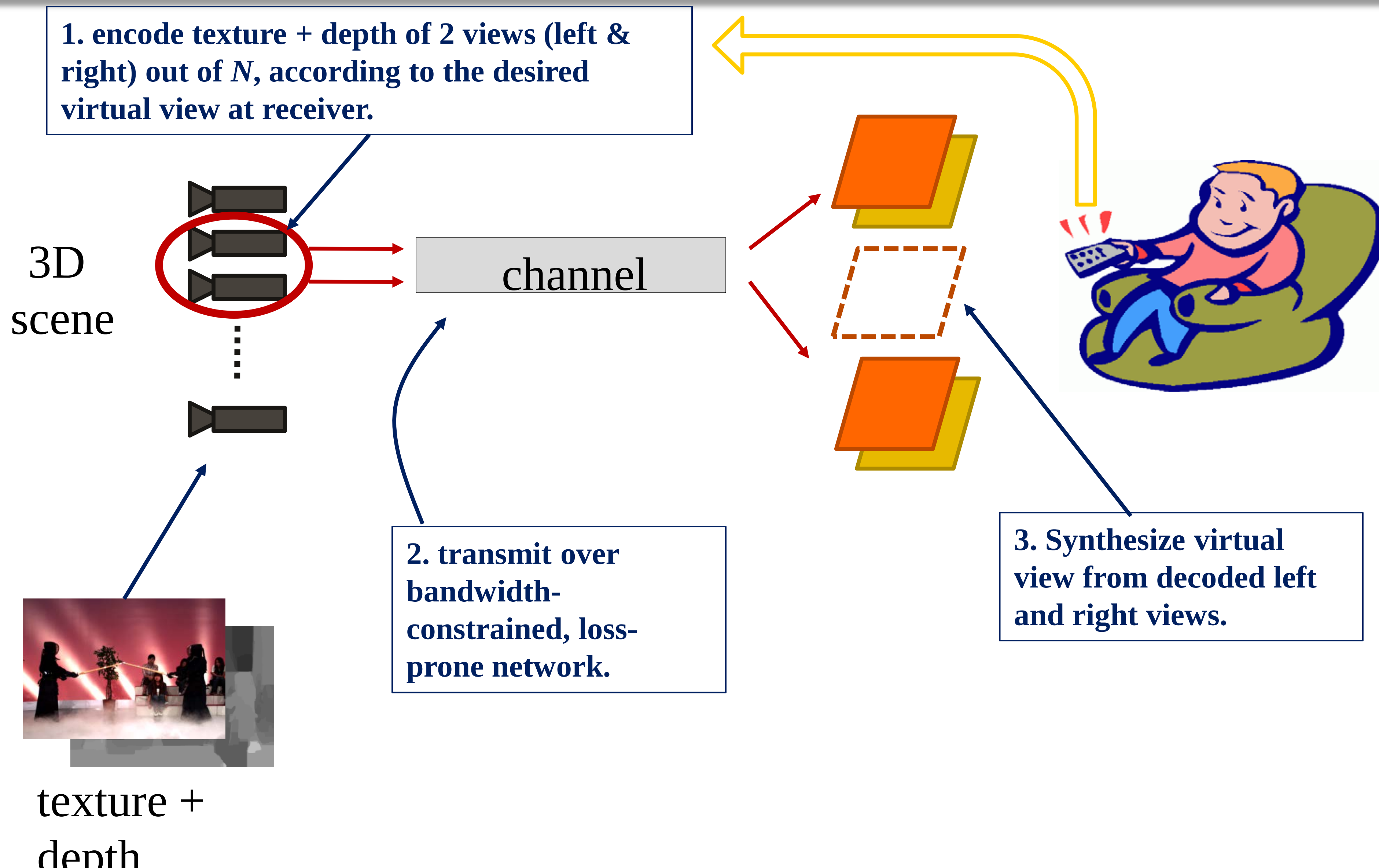
Saliency-based Error Concealment

Experiment: up to 3.6dB improvement in PSNR.

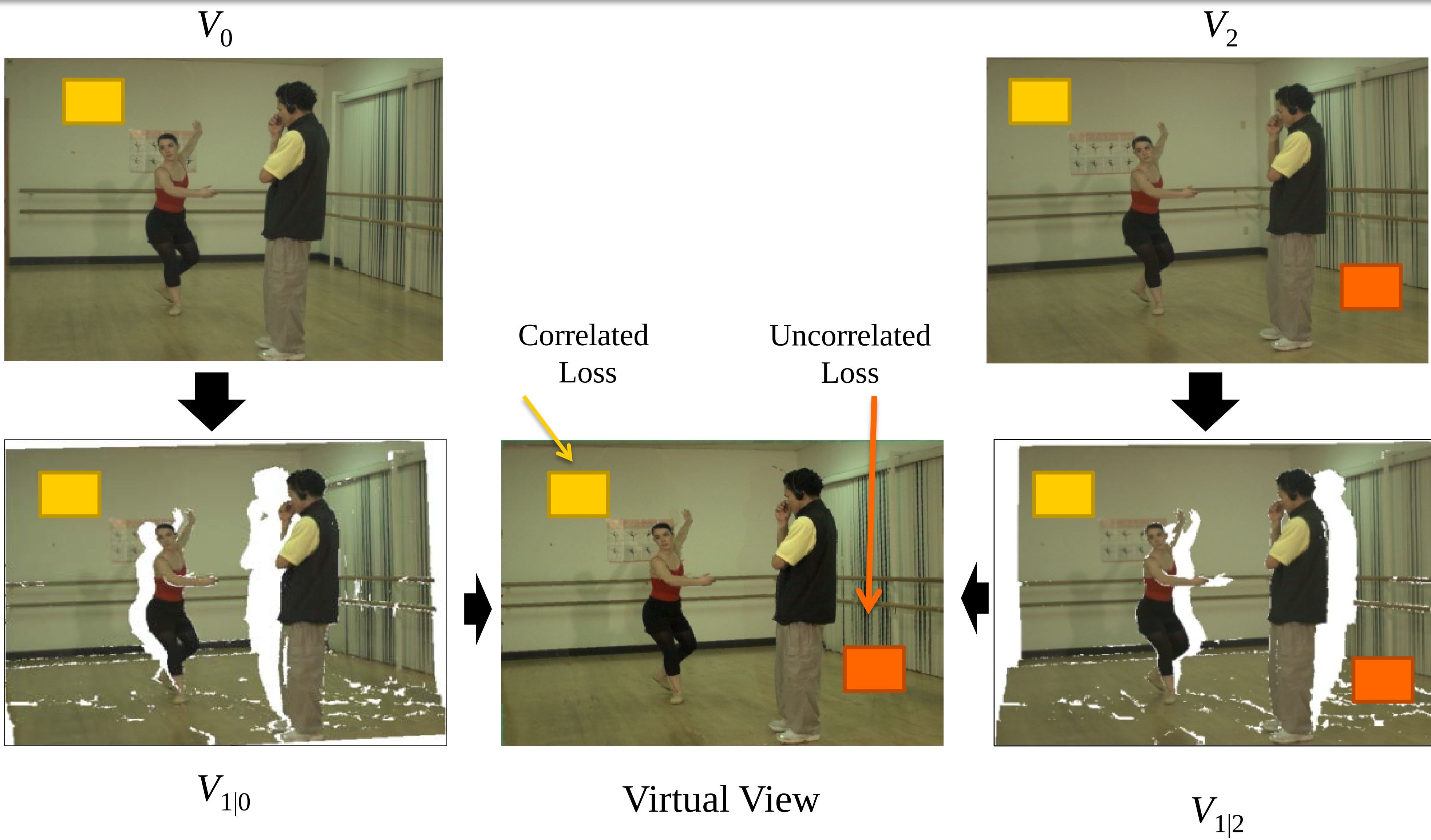
RECAP

Our Proposal

Background: Free Viewpoint Video Streaming

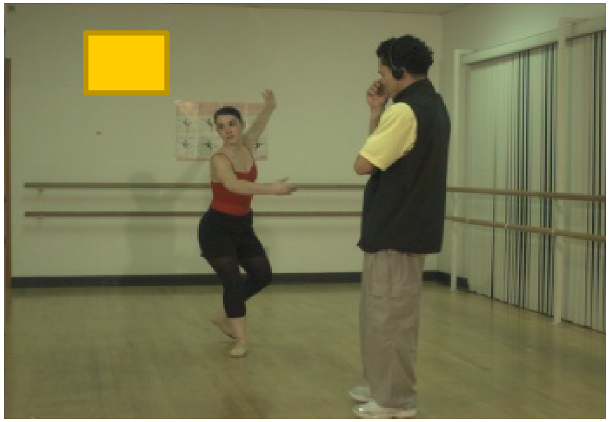


Background: Packet Loss



Background: Packet Loss

V_0



V_2

Q: What is a good view synthesis strategy given losses in reference views?

Correlated Loss

Uncorrelated Loss

$V_{1|0}$

Virtual View

$V_{1|2}$

System Assumption

- Retransmission of lost packets (ARQ) leads to **interactive delay**.
  - **Foward error correction** (FEC) code is used.

- **Unequal error protection** (UEP) is applied, where more important regions are protected more using FEC.

Low salient region: weak FEC

High salient region: strong FEC

Formulation

1. Identify lost pixels.

2. For each lost pixel patch *p*, construct two patch candidates:

- ***Weighted Pixel Blendng (WPB)*** $\rightarrow \psi_p^1$
- ***Examplar-Based Matching (EPM)*** $\rightarrow \psi_p^2$

3. Select between 2 candidates:

$$\min_{g \in \{1,2\}} D(\psi_p^g) + \lambda Z(\psi_p^g)$$

D – Expected Distortion
Z – Computed Saliency

Formulation

1. Identify lost pixels.

2. For each lost pixel patch p, construct two patch candidates:

- ***Weighted Pixel Blendng (WPB)*** $\rightarrow \psi_p^1$
- ***Examplar-Based Matching (EPM)*** $\rightarrow \psi_p^2$

Low-saliency prior

3. Select between 2 candidates:

$$\min_{g \in \{1,2\}} D(\psi_p^g) + \lambda Z(\psi_p^g)$$

D – Expected Distortion
Z – Computed Saliency

Weighted Pixel Blending

X_t^0

X_t^1

$$S_t^v(i, j) = (1 - v)X_t^0(i, j^0) + vX_t^1(i, j^1)$$

Weighted Pixel Blending

X_t^0

X_t^1

$$S_t^v(i,j) = (1-v)X_t^0(i,j^0) + vX_t^1(i,j^1)$$

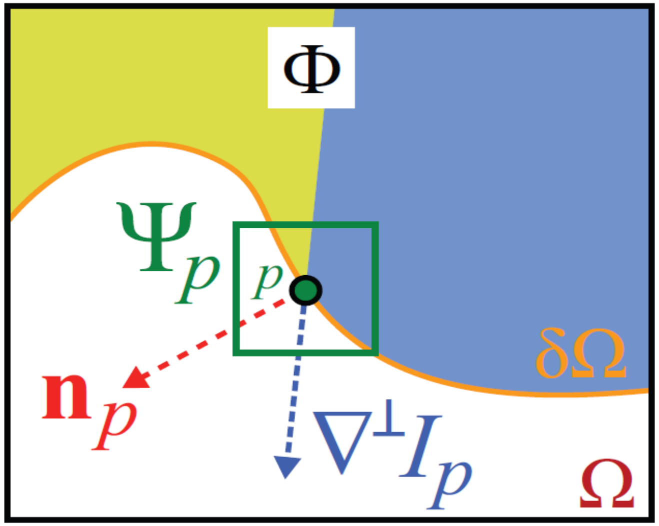
Key idea: adjust weights based on pixel reliability

Examplar-Based Patch Matching

- A similar algorithm as [8] is applied.

  [8] A. Criminisi, P. Perez and C. Gomila., "Region filling and object removal by examplar-based image inpaiting", in IEEE Transactions on Image Processing, September 2004, vol 13., no 9, pp 1-13.

- The order in which patches in the target region Ω is filled is done according to a ***priority factor*** *P(p)*.



$$P(p) = C(p)D(p)$$

C(p) denotes confidence

D(p) is the data term which is a function of the strenght of isophotes.

Low-Saliency Prior

- We determine the patch around a missing pixel with the highest priority.

- Then, the two possible candidates using WPB and EPM are seltected based on:

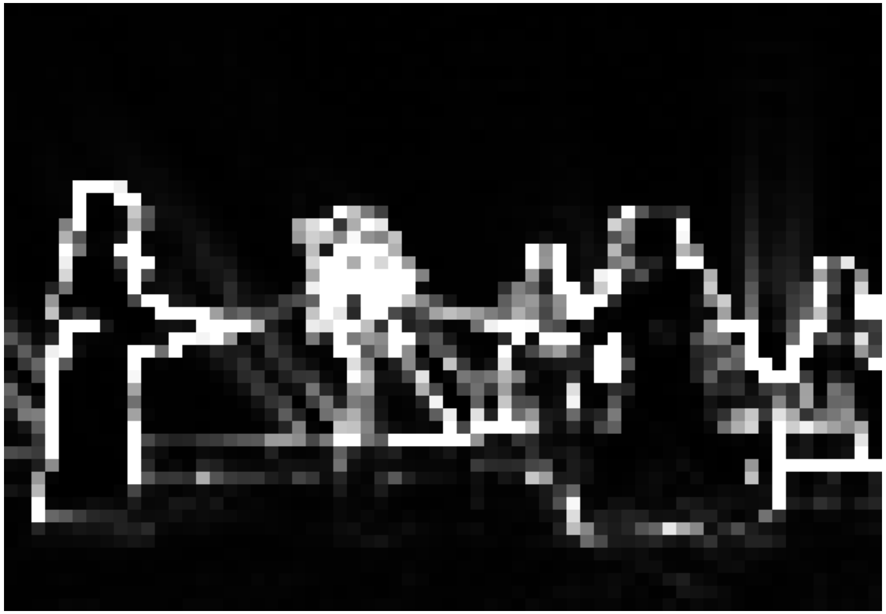
$$\min_{g \in \{1,2\}} D(\psi_p^g) + \lambda Z(\psi_p^g)$$

  - *D()* for WPB is the average estimated distortion of pixel in patch
  - *D()* for EPM is the average estimated distortion of the copied patch

Experimental Results

- Packet Losses manifest themselves as isolated MBs due to FMO.

- Packet Losses occur only in low-saliency regions (black regions in the image) due to UEP.



Results (uncorrelated losses)

Table 1. Uncorrelated losses

Kendo	5%	10%	20%	30%
Co-located	35.48 dB	35.33 dB	35.03 dB	34.72 dB
EPM	35.64 dB	35.55 dB	35.26 dB	35.09 dB
WPB	35.72 dB	35.62 dB	35.49 dB	35.35 dB
Proposed	35.74 dB	35.64 dB	35.68 dB	35.48 dB

Akko and Kayo	5%	10%	20%	30%
Co-located	28.70 dB	28.54 dB	28.12 dB	27.64 dB
EPM	28.88 dB	28.64 dB	28.25 dB	27.74 dB
WPB	28.87 dB	28.75 dB	28.35 dB	28.00 dB
Proposed	28.88 dB	28.78 dB	28.46 dB	28.22 dB

Results (correlated losses)

Table 2. Correlated losses

	Kendo			
	5%	**10%**	**20%**	**30%**
Co-located	35.50 dB	35.30 dB	35.07 dB	34.66 dB
EPM	35.68 dB	35.62 dB	35.48 dB	35.29 dB
WPB	35.57 dB	35.37 dB	35.12 dB	34.69 dB
Proposed	35.69 dB	35.70 dB	35.62 dB	35.42 dB
	Akko and Kayo			
	5%	**10%**	**20%**	**30%**
Co-located	28.70 dB	28.36 dB	27.92 dB	27.61 dB
EPM	28.77 dB	28.56 dB	28.30 dB	27.91 dB
WPB	28.69 dB	28.33 dB	27.99 dB	27.59 dB
Proposed	28.81 dB	28.59 dB	28.41 dB	27.99 dB

Experimental Results

Using Co-located Blocks.
PSNR 34.70 dB

Proposed
PSNR 35.39 dB

Presentation Outline

- **Background & Motivation** (3D, not your mother's 2D)
- **3D Video representation / coding:**
  - Depth map coding
    - HEVC tools for depth maps
    - Graph-based Transform (GBT) for depth maps
  - Depth map denoising
  - Denoising + compression?
  - Why code depth images?
- **3D Video streaming:**
  - Video compression with flexible decoding for interactive streaming
  - Loss-resilient texture-plus-depth video streaming (skip)
- **3D view synthesis:**
  - Robust view synthesis for free viewpoint video
  - Synthesized image interpolation for *z*-dimension camera movement

DIBR and its difficulty with z-movement

DIBR

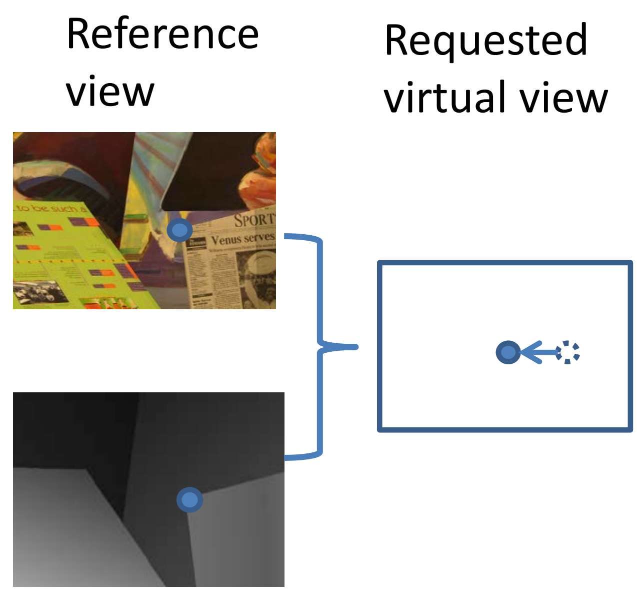
1. **Texture + Depth**
2. **DIBR** to project known pixels
3. Inpainting at decoder or intra-coded blocks sent from server to fill in pixels in disoccluded regions.

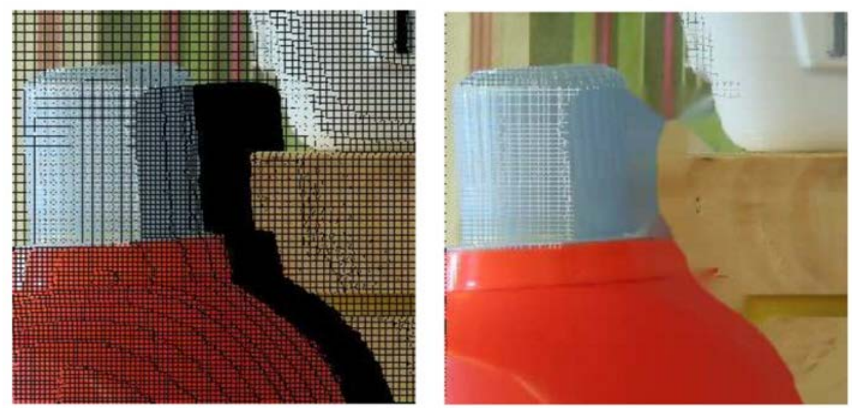
View-switch along the z-dimension is very natural, but it is missing in the current systems.

Difficulty:

Pixels get scattered far apart

P. Merkle, A. Smolic, K. Mueller, and T. Wiegand, "Multiview video plus depth representation and coding" in IEEE International Conference on Image Processing, San Antonio, TX,2007





(a) expansion holes (b) VSRS+

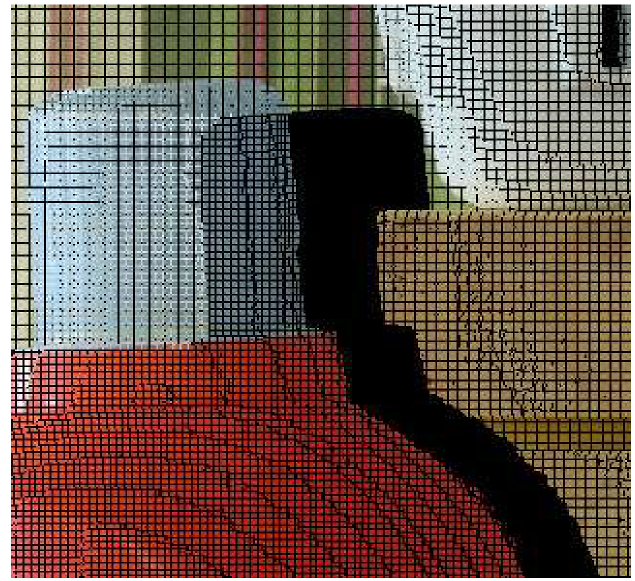
Our Work

- Goal:

Design a new interpolation method that supports z-dimension navigation

With better quality of interpolation, less information is need to be sent to enhance the quality

- Challenges:
  1. Distinguish between expansion holes and disocclusion holes
  2. How to interpolate the hole area



Example of expansion holes

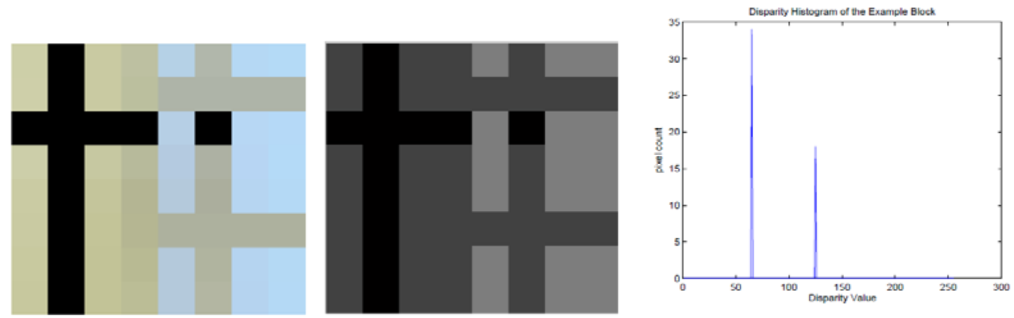
Disocclusion: region not visible in reference

Expansion: low sample rate

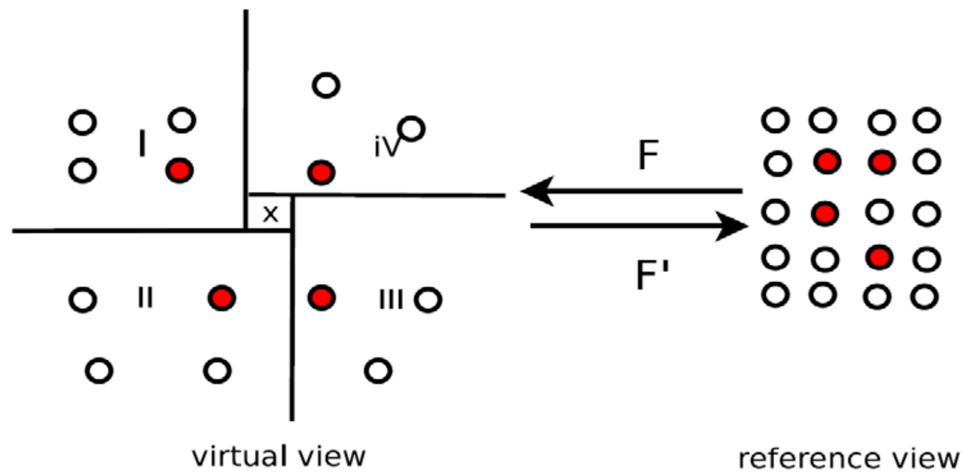
Distinguish between expansion holes and disocclusion holes

Block based processing :

1. construct a histogram of depth values of the synthesized pixels in the block,
2. separate depth pixels into layers
3. Convex set based identification



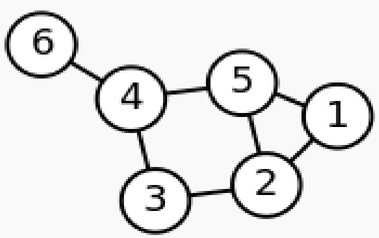
(a) texture block (b) disparity block (c) depth histogram



Expansion Hole Interpolation

Interpolation:

1. Construct a Graph G, with pixels as its vertices, and connect the vertices with weighted edges
2. Use the eigen-vectors of the Graph Laplacian as the transform matrix

Labeled graph	Degree matrix	Adjacency matrix	Laplacian matrix
	$\begin{pmatrix} 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 2 & -1 & 0 & 0 & -1 & 0 \\ -1 & 3 & -1 & 0 & -1 & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 \\ 0 & 0 & -1 & 3 & -1 & -1 \\ -1 & -1 & 0 & -1 & 3 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 \end{pmatrix}$

Calculation of Graph Laplacian

Expansion Hole Interpolation

Non-local means: exploit the self-similarity in the images

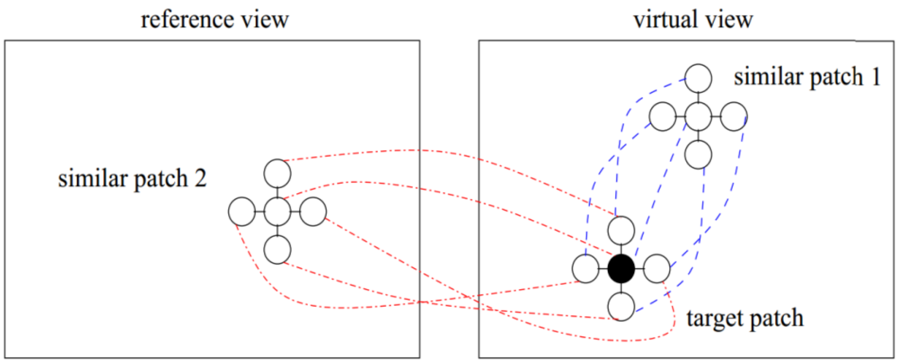


Fig. 3. Connecting pixels of similar patches to target patch

(a) expansion holes (b) VSRS+ (c) GBT

Edge weights $e_{i,j} = w_{i,j}\, u_{i,j}\, v_{i,j}$

Sparse signal recovery $\min_{\mathbf{w}} \|\sum_{i=1}^{N} \mathbf{u}_i^T \Phi \mathbf{w} - s_i\|_1 + \lambda \|\mathbf{w}\|_1$

Experiment results

PSNR COMPARISON FOR EXPANSION HOLE FILLING.

method	VSRS+	GBT	NLGBT
art PSNR(dB)	19.56	23.36	23.58
moebius PSNR(dB)	19.47	23.15	23.33

Summary:
3D View Synthesis

- Inverse 3D imaging problem
  - Not enough info for perfect reconstruction
  - Leverage on image interpolation, inpainting, super-resolution
- Co-design with signal representation at sender?

Presentation Summary

- **3D Video representation / coding:**
  - Depth map coding: standard + non-conventional coding tools.
  - Depth map denoising
  - Q: Denoising + compression?
  - Q: Why code depth images?
- **3D Video streaming:**
  - Video compression with flexible decoding for interactive streaming
  - Q: High-dimensional media navigation problem?
- **3D view synthesis:**
  - Robust view synthesis w/ low-saliency prior
  - Synthesized image interpolation using graph transform
  - Q: Inverse 3D imaging problem? Co-design w/ representation?

Q & A

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

- CfP for Special Issue on "**Interactive Media Processing for Immersive Communication**" in *IEEE Journal on Selected Topics in Signal Processing.*
  - Submission deadline: April 2nd, 2014
  - Guest Editors: Gene Cheung, Dinei Florencio, Patrick Le Callet, Chia-Wen Lin, Enrico Magli