Rectification-based View Interpolation and Extrapolation: R-D Analysis and Applications in Multiview Video/Image Coding

Jie Liang (jiel@sfu.ca)
School of Engineering Science, Simon Fraser University, Vancouver, BC, Canada
Workshop on Multimedia, Mathematics & Machine Learning II
Banff, Alberta, Canada, July 5-10, 2009

Outline

- Introduction
- R-D analysis
- Applications
- Summary

Introduction

- Multiview video coding (MVC) is an important component of 3DAV applications
- Disparity-compensated prediction:
  - Used by H.264 MVC
  - Not always efficient [Schwarz'06]
- View interpolation/synthesis-based prediction:
  - Exploits geometric redundancy
  - Also useful for free viewpoint TV

Introduction

- **Existing view interpolation-based MVC methods:**
  - Require parallel cameras [Yamamoto'07]
  - Or need camera parameters or depth map [Martinian'06, Shimizu'07]

- **Recently at PCS'09 we apply the projective rectification-based view interpolation in [Hartley'99] to MVC**
  - Allow flexible camera setup
  - Camera parameters not needed
  - Also used for rendering in [Farin'06, Kauff'07]
  - R-D performance not studied yet

- **This talk:**
  - Theoretical model for rectification-based view interpolation and the corresponding R-D model for MVC [ICME'09]
  - View extrapolation-based MVC [ICME'09]
  - Multiple description coding (MDC) of multiview images [Asilomar'09]

Rectification based View Interpolation

[Hartley'99]

Left and right views

↓

Estimate the fundamental matrix F:
- Corner detection
- Random Sample Consensus (RANSAC)
(Calibration not needed)

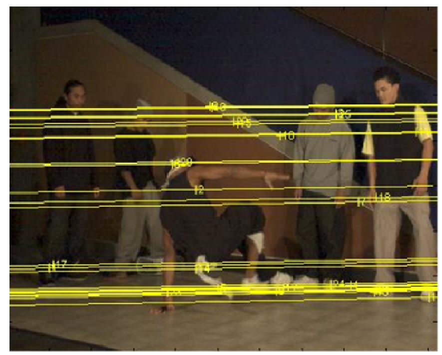
↓

Get projective rectification matrix:
- Find **epipoles** from F matrix
- Rotate the **epipoles** to infinity
(parallel cameras created)

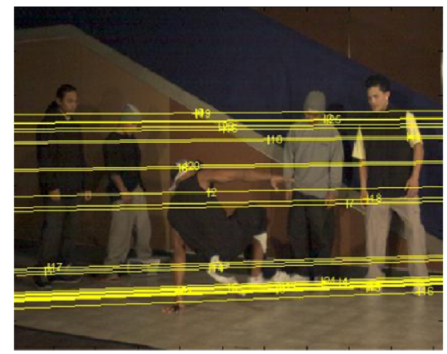
1-D disparity after rectification →

Left view

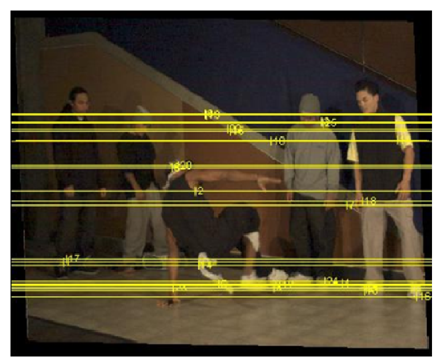
Right view



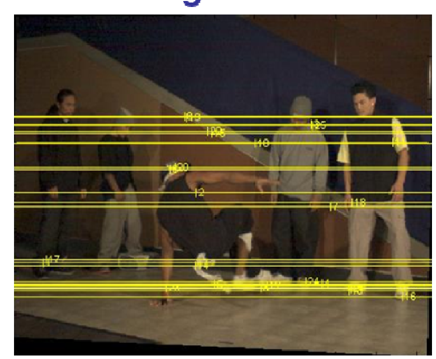
Left view



Right view



Rectified Left view



Rectified Right view

Rectification based View Interpolation

Rectified left & right views

↓

Perform disparity estimation

↓

Interpolate the middle view

↓

Un-rectify the interpolated view

Interpolated view

Unrectified interpolated view

- **Problem**: complexity

View Interpolation-based MVC

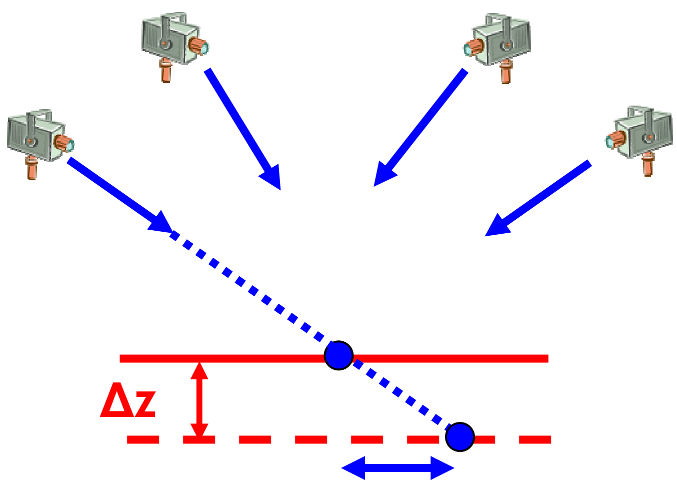
- **To use the view interpolation in H.264 based MVC, each frame can have 4 reference frames:**
  - The interpolated view (from rectification)
  - The temporal reference from the same view
  - The current frames in the left & right neighbouring views
- **Each block picks the best reference using R-D optimization.**

Existing Theoretical Models

- **Our R-D analysis is mainly based on the following papers:**
- [Girod *et al.*-'03,'06,'07]
- [Takahashi-ICIP'08]

- **1. MVC R-D Model in [Girod *et al.*-'03,'06,'07]:**
  - Object: 2D signal s(x,y) on a planar surface
  - Geometry error: an **offset Δz** from its true position

➔ Each view is a **shifted** version of s(x,y):



$$c_i(x,y) = s(x-\Delta_{x_i}, y-\Delta_{y_i}) + n_i(x,y)$$

- Other view-dependent effects are modeled as noise.
- The impact of Δz on MVC R-D performance can be derived from this.

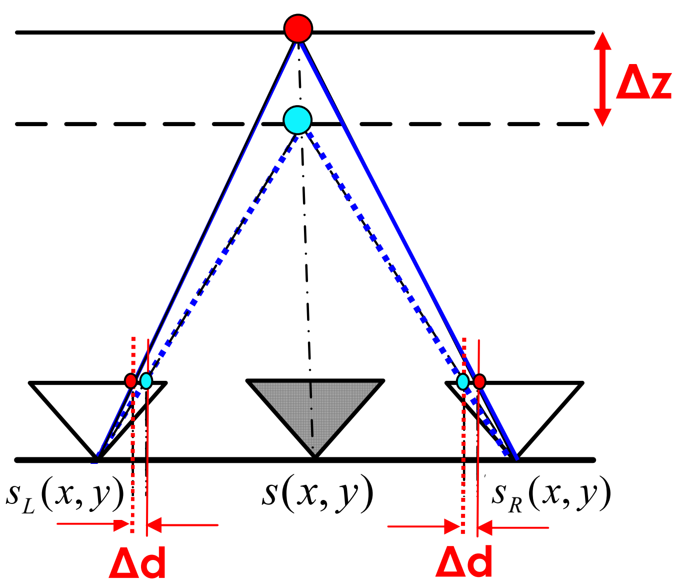
Existing Theoretical Models

- **2. View Interp. Model in [Takahashi-ICIP'08]:** s(x,y): middle view

$$s(x,y) = 1/2\{s_L(x+d,y) + s_R(x-d,y)\}$$

2d: disparity between left & right views

- **Assumption:** geometry error Δz causes some disparity error Δd



Interpolated view:

$$\hat{s}(x,y) = 1/2\{s_L(x+(d+\Delta d),y) + s_R(x-(d+\Delta d),y)\}$$
$$= 1/2\{s(x+\Delta d,y) + s(x-\Delta d,y)\}$$

Error: $$e(x,y) = s(x,y) - \hat{s}(x,y)$$

Power spectral density (PSD) of e(x,y):

$$\Phi_{ee}(\omega_x,\omega_y) = (1-\cos(\Delta d\cdot\omega_x))^2\Phi_{ss}(\omega_x,\omega_y).$$

The MSE of the interpolated view can be obtained.

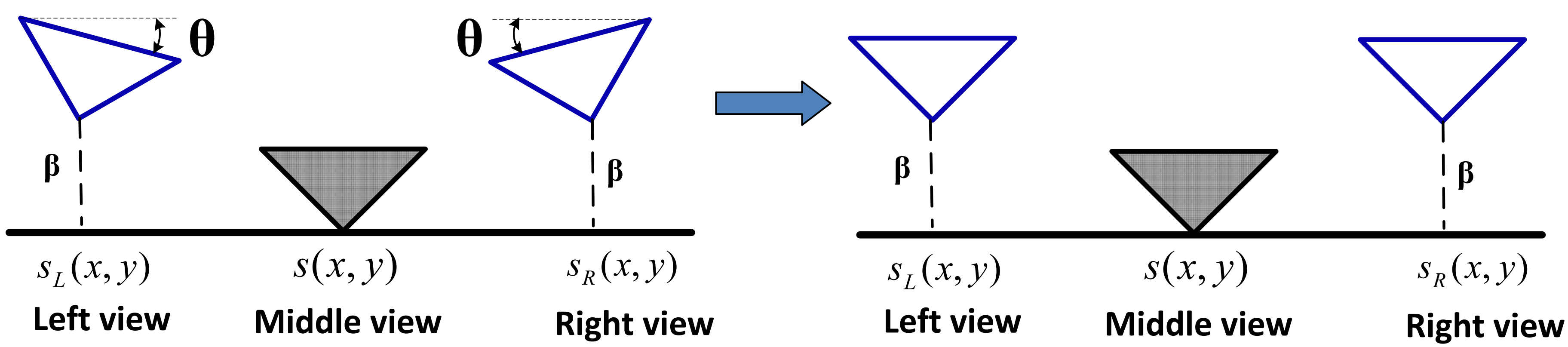
Limitations: 1. Parallel cameras; 2. R-D and MVC are not considered.

Proposed Model for Rectification-based View Interpolation

- **Our Goal: analyze the R-D performance of MVC when rectification-based view interpolation is used**
  - Assuming the orientations of the left & right views differ from the middle view by θ and are not in the same line as the middle view

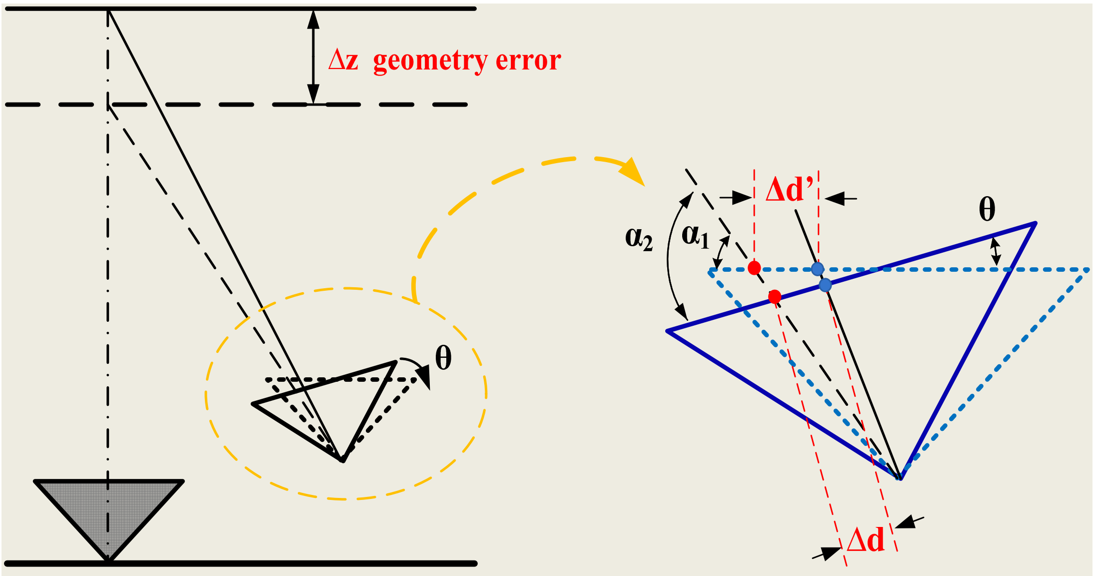
- **Rectification becomes:**

  1. Rotate left & right views to parallel positions.
  2. Shift them to the same line as the middle view.



Proposed Model for Rectification-based View Interpolation

- **Step 1:** After rotation, the disparity error **Δd** becomes **Δd'**:



$$\frac{\Delta d'}{\Delta d} \approx \cos\theta + \sin\theta \frac{\cos\alpha_1}{\sin\alpha_1}.$$

Avg. of Δd'/ Δd in the field of view:

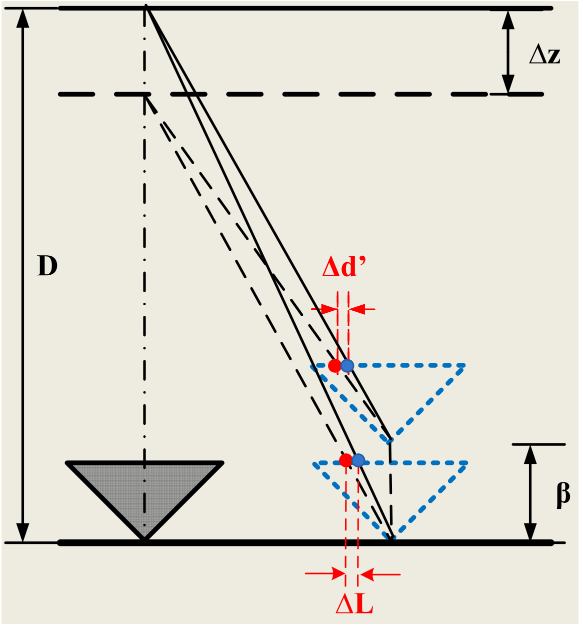
$$\frac{1}{\alpha_{FOV}} \int_{FOV} \left(\cos\theta + \sin\theta \cdot \frac{\cos\alpha_1}{\sin\alpha_1} \right) d\alpha_1 = \cos\theta.$$

➔ Rectification reduces the disparity error by a factor of cosθ on average.

Proposed Model for Rectification-based View Interpolation

- **Step 2:** If D>>β, the disparity error has little change after shifting by β ➔

$$\Delta L \approx \Delta d' \approx \Delta d \cdot \cos\theta.$$



Interpolated middle view after rectification:

$$\hat{s}(x,y) = 1/2\{s(x+\Delta d\cdot\cos\theta, y) + s(x-\Delta d\cdot\cos\theta, y)\}$$

Fourier Transform:

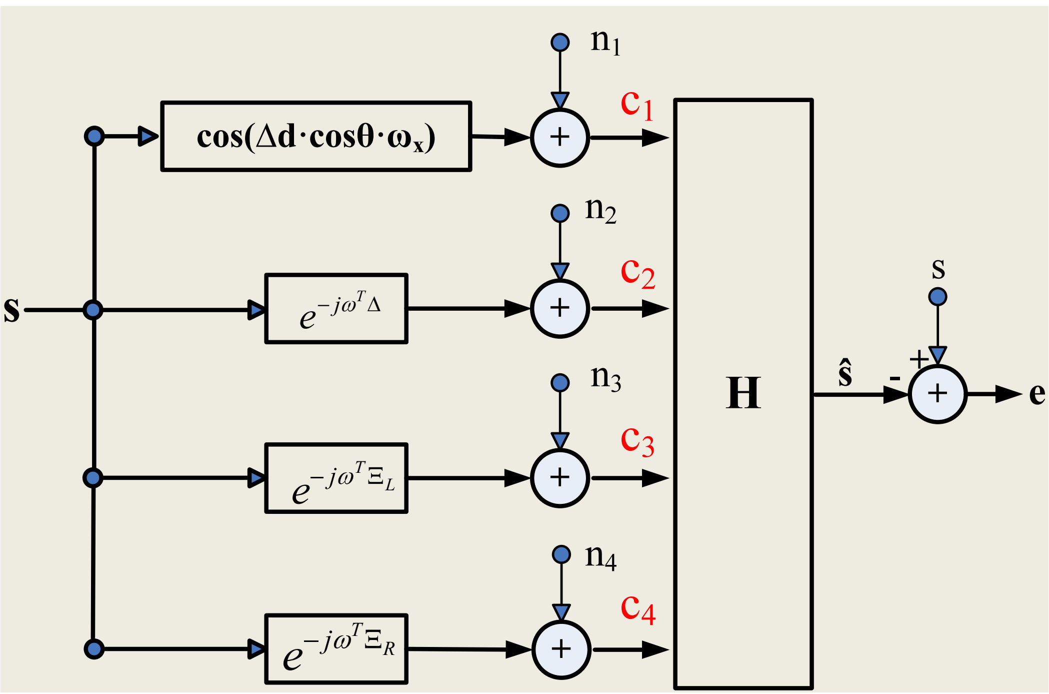
$$\hat{S}(\omega_x,\omega_y) = \cos(\Delta d\cdot\cos\theta\cdot\omega_x)\cdot S(\omega_x,\omega_y).$$

The model in [Takahashi-ICIP'08] is a special case of ours with θ = 0.

R-D Model for MVC with Rectification-based View Interpolation

- **To use view interp. In MVC, each frame has 4 reference frames:**
  - C_1: The interpolated view
  - C_2: The temporal reference with motion displacement Δ.
  - C_3, C_4: The references from left & right views, with disparity Ξ_L and Ξ_R.

- **The following model is obtained (based on [Girod *et al.*-'03,'06,'07]):**



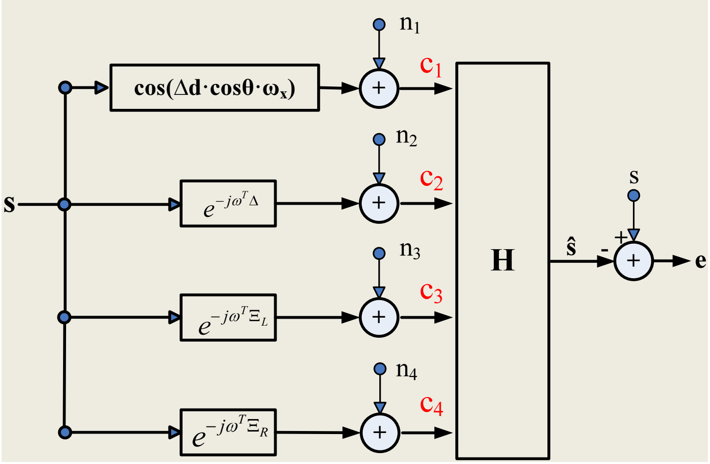
- The MVC can be modelled as a linear estimation problem.
- The prediction residual ***e*** is encoded.
- The RD performance can be obtained.
- **Another performance measure:** Rate difference w.r.t. direct coding.

R-D Model for MVC with Rectification-based View Interpolation

- **The LMMSE filter:** $$H = \Phi_{cs}^{*}\Phi_{cc}^{-1}$$

- **The PSD of the prediction residual:**

$$\Phi_{ee} = \Phi_{ss} - \Phi_{cs}^{*}\Phi_{cc}^{-1}\Phi_{cs}$$



$$\mathbf{c} = \begin{bmatrix} c_1 & c_2 & c_3 & c_4 \end{bmatrix}^T$$

- **R-D function of Gaussian signal x(x, y):**

$$R(\lambda) = \frac{1}{8\pi^2}\int_{\omega_x}\int_{\omega_y} \max\left(0, \log_2 \frac{\Phi_{xx}(\omega_x,\omega_y)}{\lambda}\right) d\omega_x d\omega_y$$

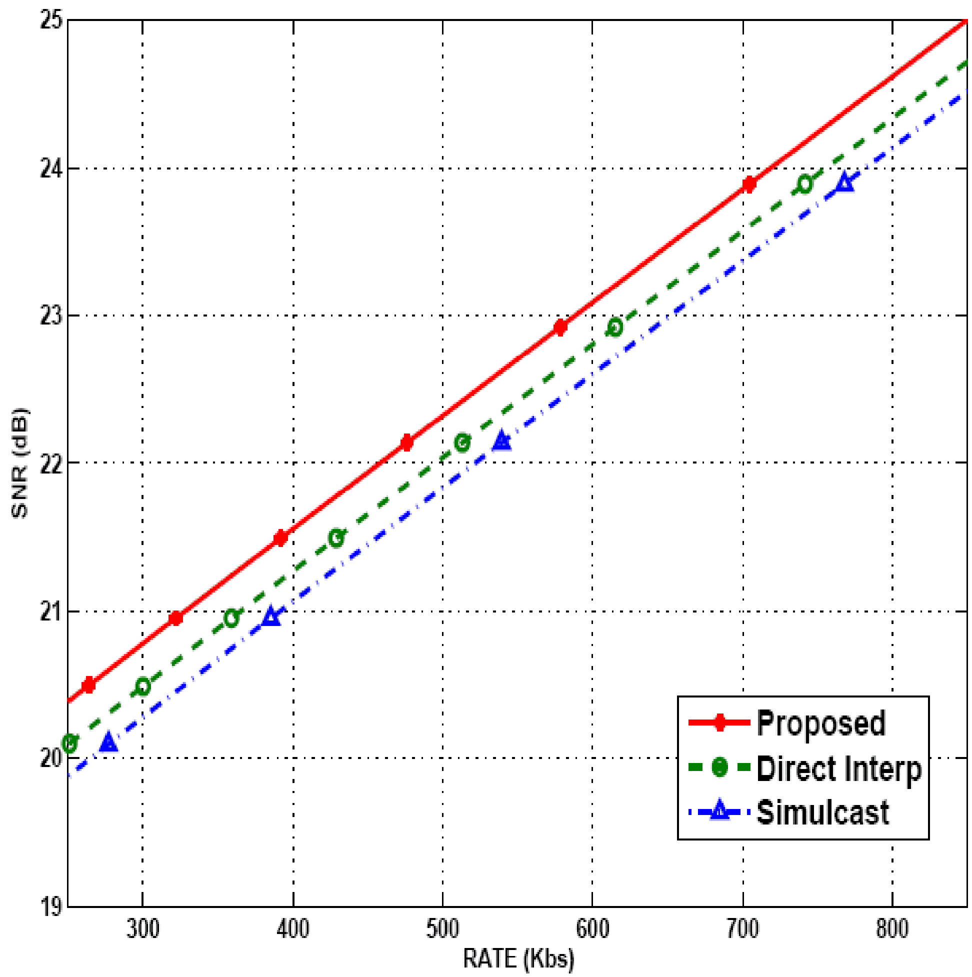
$$D(\lambda) = \frac{1}{4\pi^2}\int_{\omega_x}\int_{\omega_y} \min\left(\lambda, \Phi_{xx}(\omega_x,\omega_y)\right) d\omega_x d\omega_y$$

- **Rate difference between predictive coding and direct coding:**

$$\Delta R = R_e - R_s = \frac{1}{8\pi^2}\int_{\omega_x}\int_{\omega_y} \log_2 \frac{\Phi_{ee}(\omega_x,\omega_y)}{\Phi_{ss}(\omega_x,\omega_y)} d\omega_x d\omega_y$$

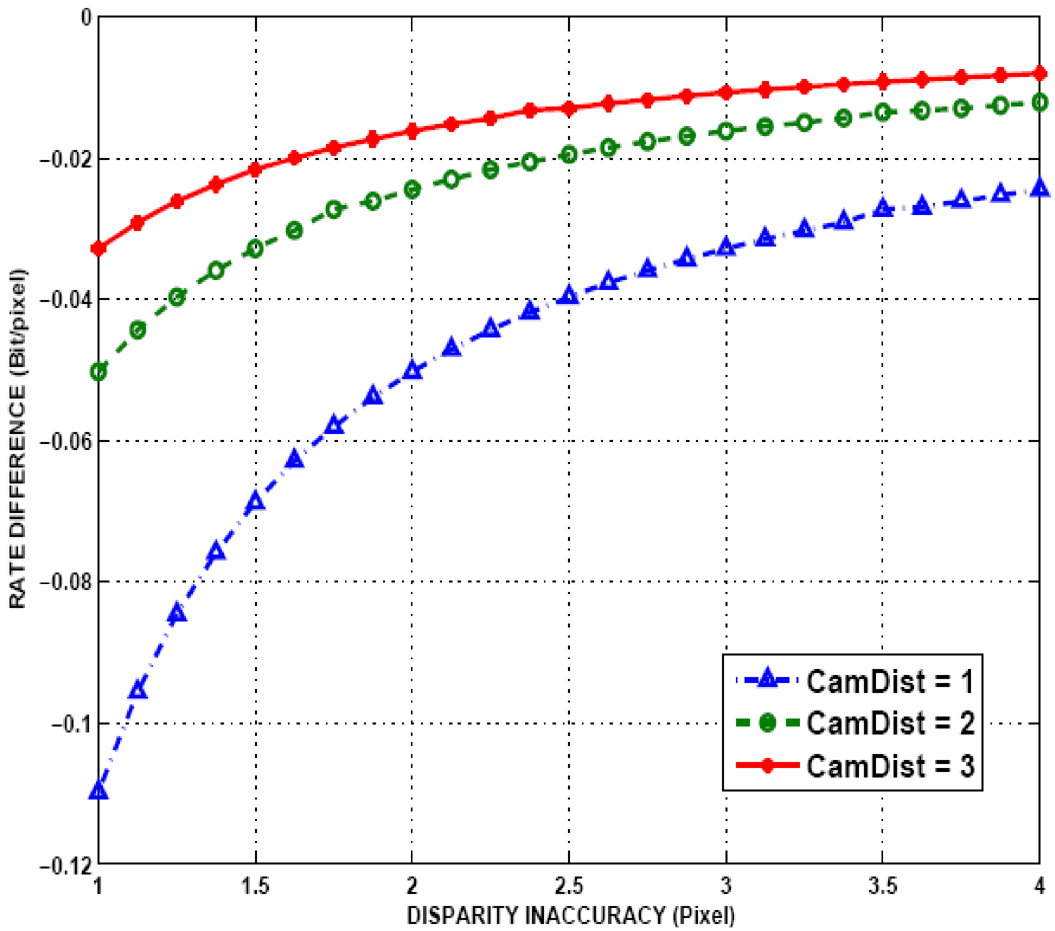
Simulation Results

- **R-D curves:** Gaussian source, $\theta = 30^{\circ}$
  (Direct interp.: view interp without rectification)



Simulation Results

- **Impact of disparity error Δd and camera distance on rate difference:**

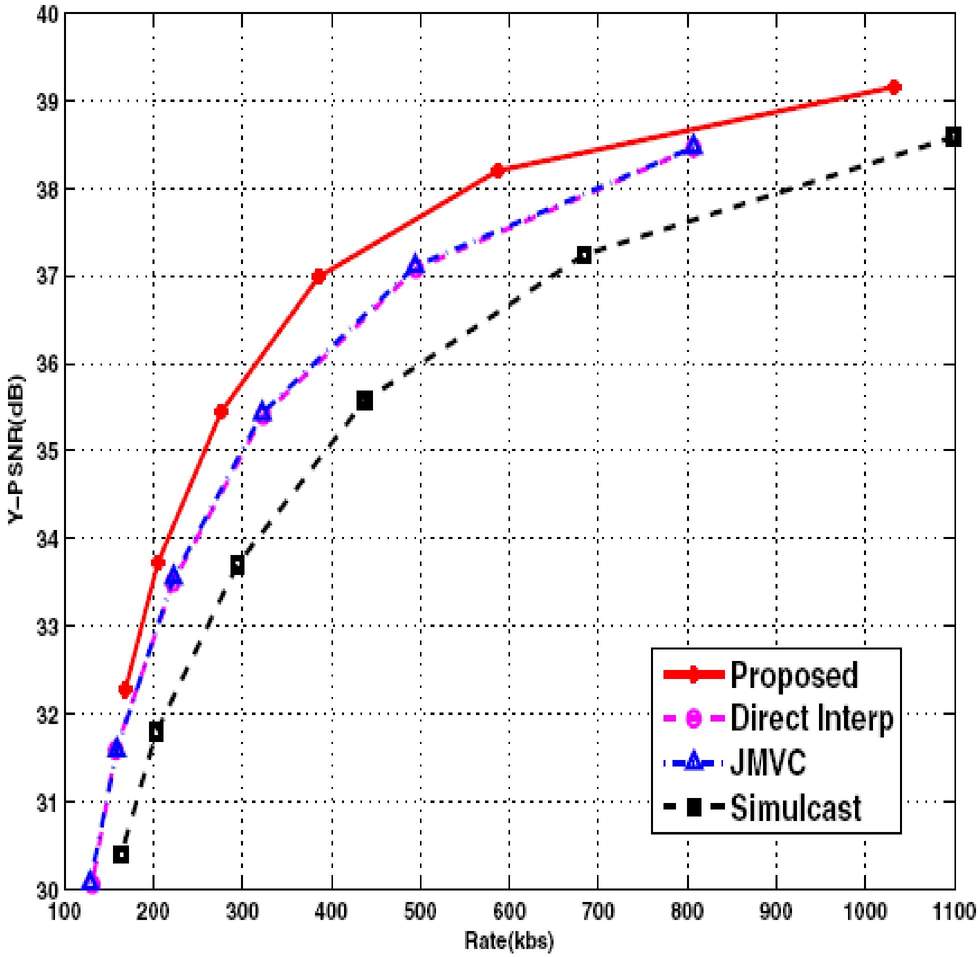


Outline

- Introduction
- R-D analysis
- Applications
  - View interpolation-based MVC
  - View extrapolation-based MVC
  - Multiple description coding of multiview images
- Summary

Interpolation-based MVC Results

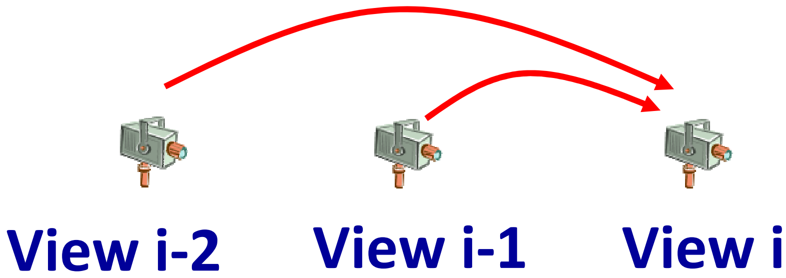
- **Coding Results: 2nd view of Breakdancers: up to 1 dB higher than JMVC**



- **Limitation: view interpolation can only be used in half of the views.**

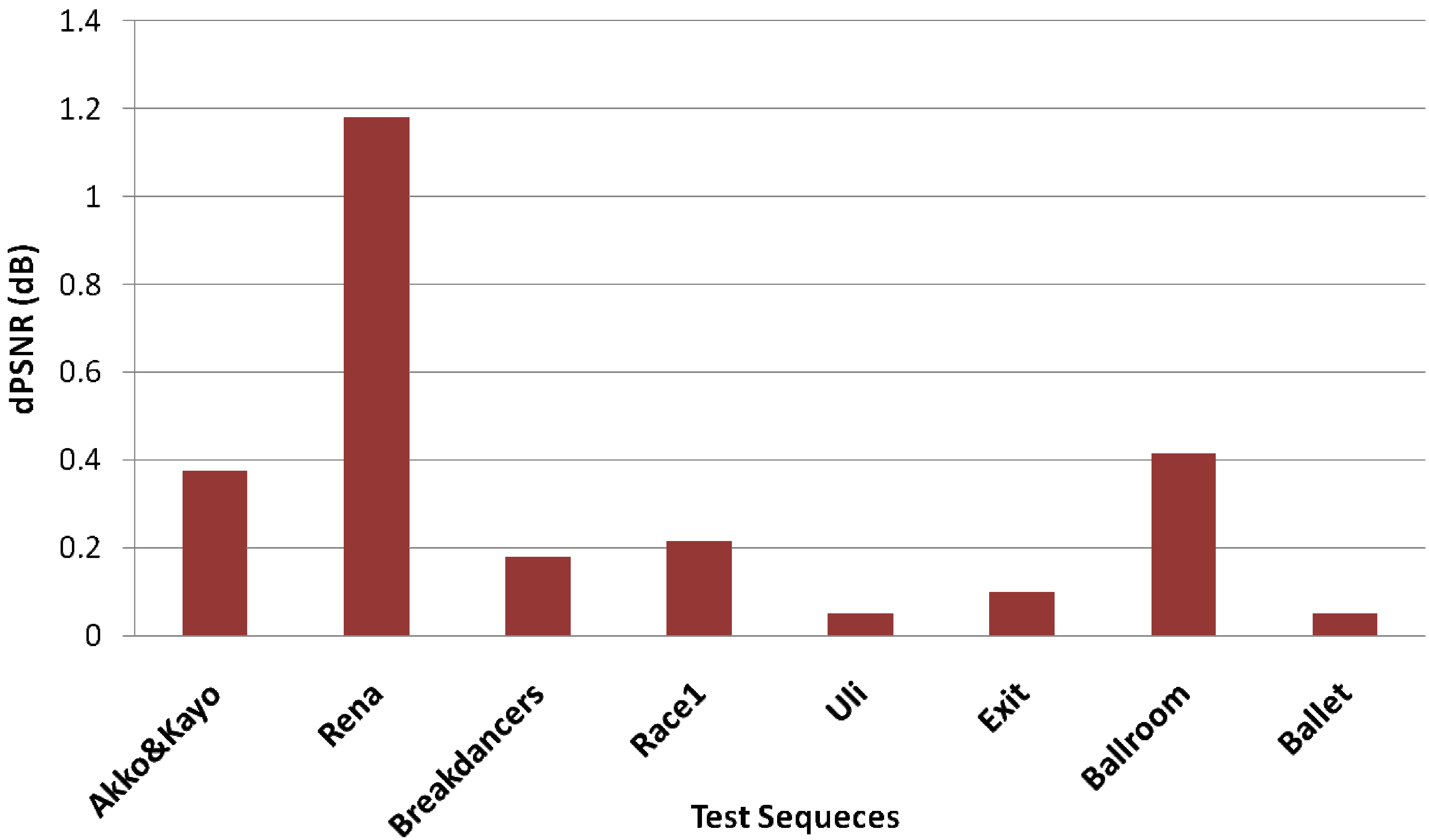
View Extrapolation

- **To apply view synthesis prediction to more views, we develop an rectification-based view extrapolation method:**
  - **Use two previous views**
  - Rectification → disparity estimation → view extrapolation → un-rectification

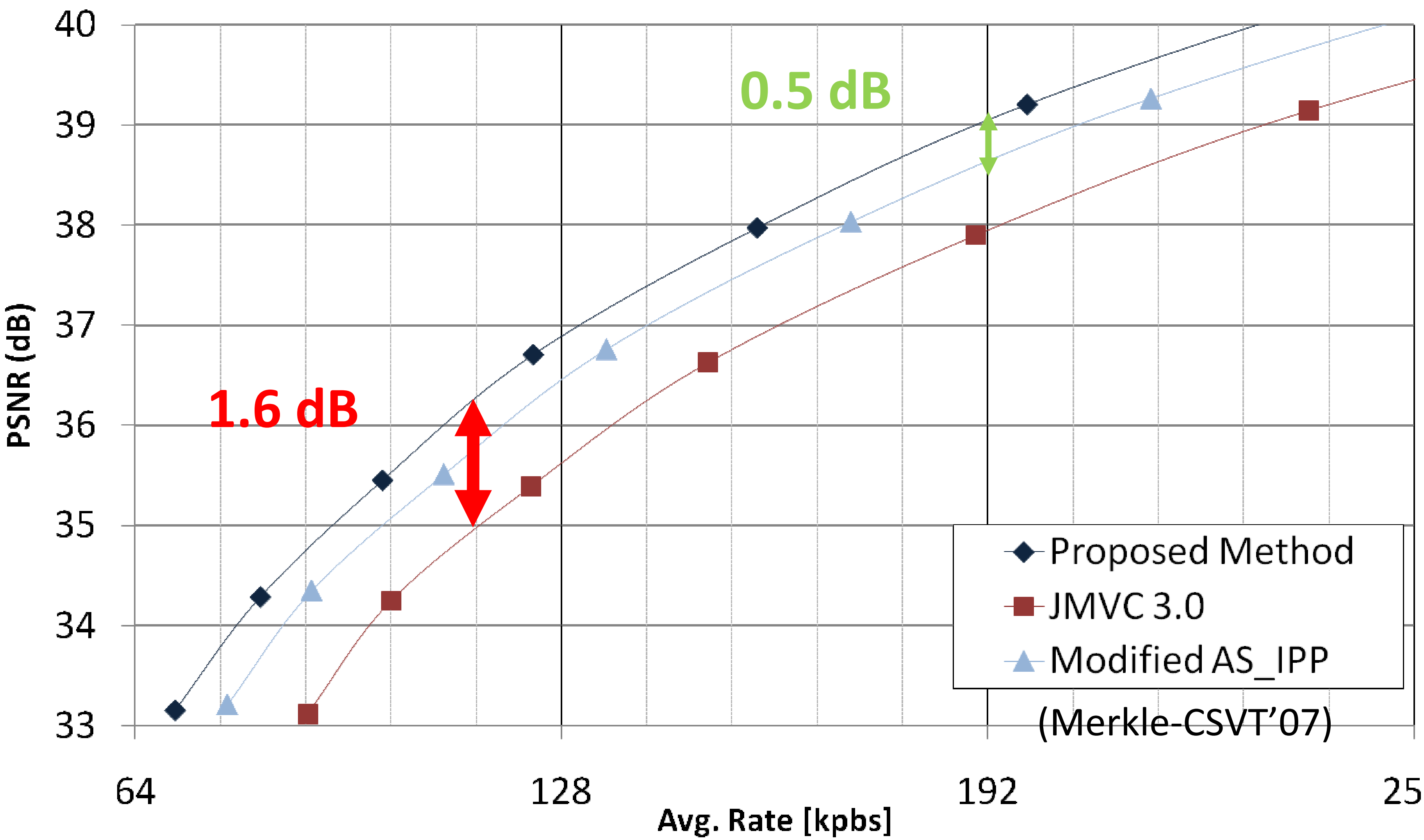


- **Extrapolation-based MVC:**
  - Encode the 1st and 2nd views using existing method
  - For other views, use the previous two views to extrapolate a synthesised view, which is used as an additional reference frame.

Average Gain of All Views Over JMVC



One View of Rena



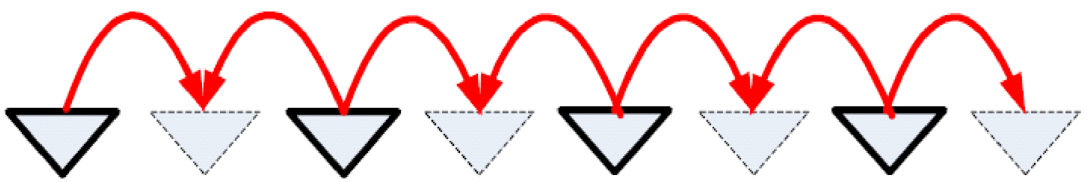
Outline

- Introduction
- R-D analysis
- Applications
  - View interpolation-based MVC
  - View extrapolation-based MVC
  - Multiple description coding of multiview images
- Summary

Multiple Description Coding of Multiview Images

- **Multiple description coding (MDC) of multiview images and videos have not been well studied.**

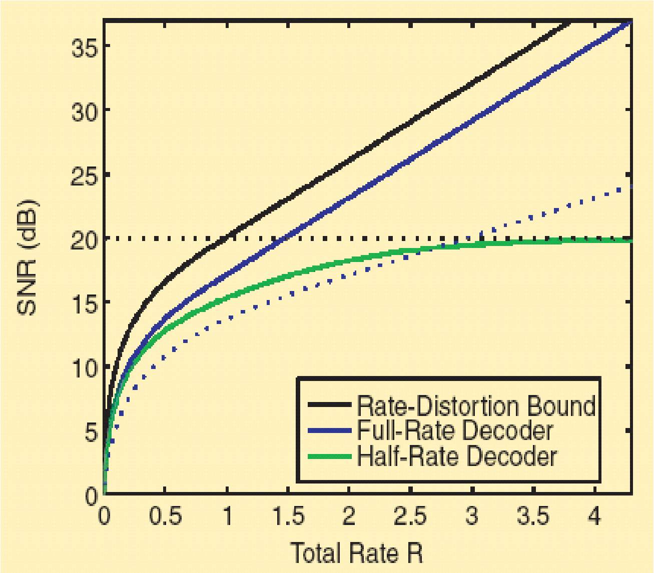
- **Related work: View subsampling [Ishii-ICIP08]:**
  - **Goal**: Reduce transmission bandwidth
  - **Approach**: Encode only half of the views (primary views).
  - **Decoder:** Missing (secondary) views are estimated from neighboring views using view interpolation.



- **Problem:**
  - The quality of the missing views cannot be improved even at high rates, due to prediction residual.

Multiple Description Coding of Multiview Images

- **This is equivalent to the even-odd spliting MDC scheme in** [Jayant'81]
- The problem of the half-rate decoder (side decoder) is also well known [Goyal'01]



[Goyal'01], pp. 78

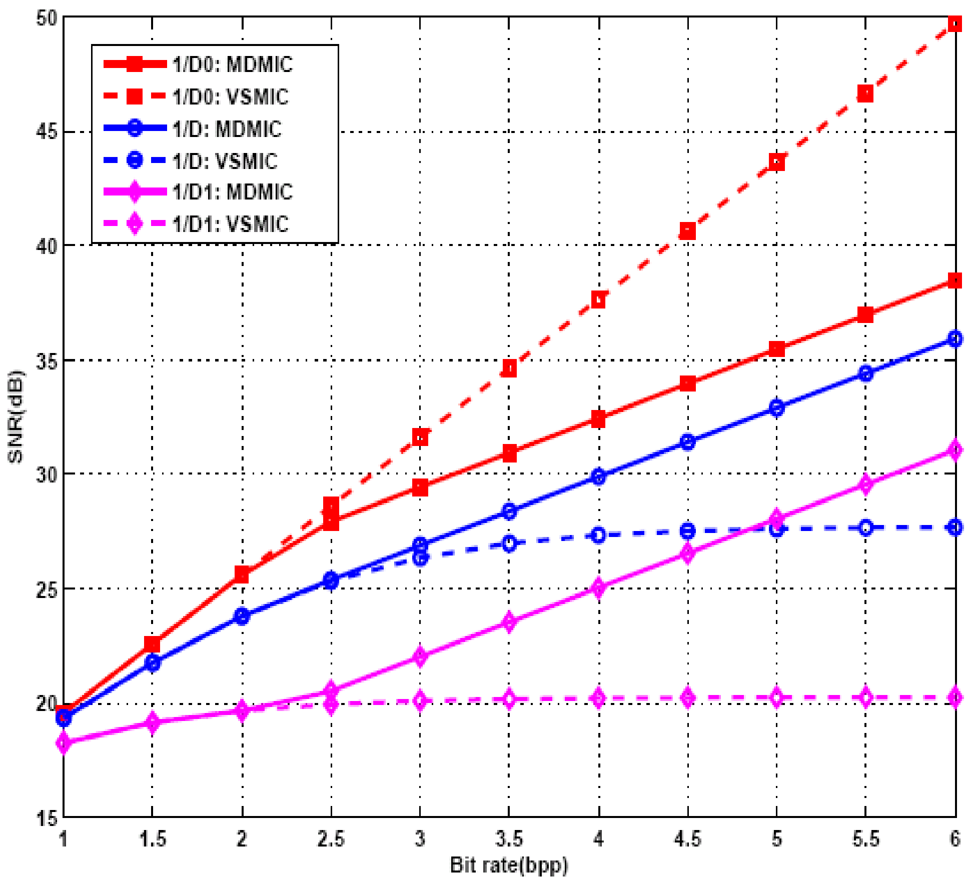
- **Possible solution: 2-layer coding [Wang'02], [Liang'07]:**
  - Base layer: primary views.
  - Enhancement layer: prediction residual of secondary views.

- **Rectification-based view interpolation is helpful here.**

Multiple Description Coding of Multiview Images

- **Theoretical R-D Analysis for MDC of multiview images:**

  Central/side/expected distortions vs bit rate



D0: **Central distortion** (two descriptions)

D1: **Side distortion** (one description)

D: **Expected distortion**

2-layer coding

View subsampling only

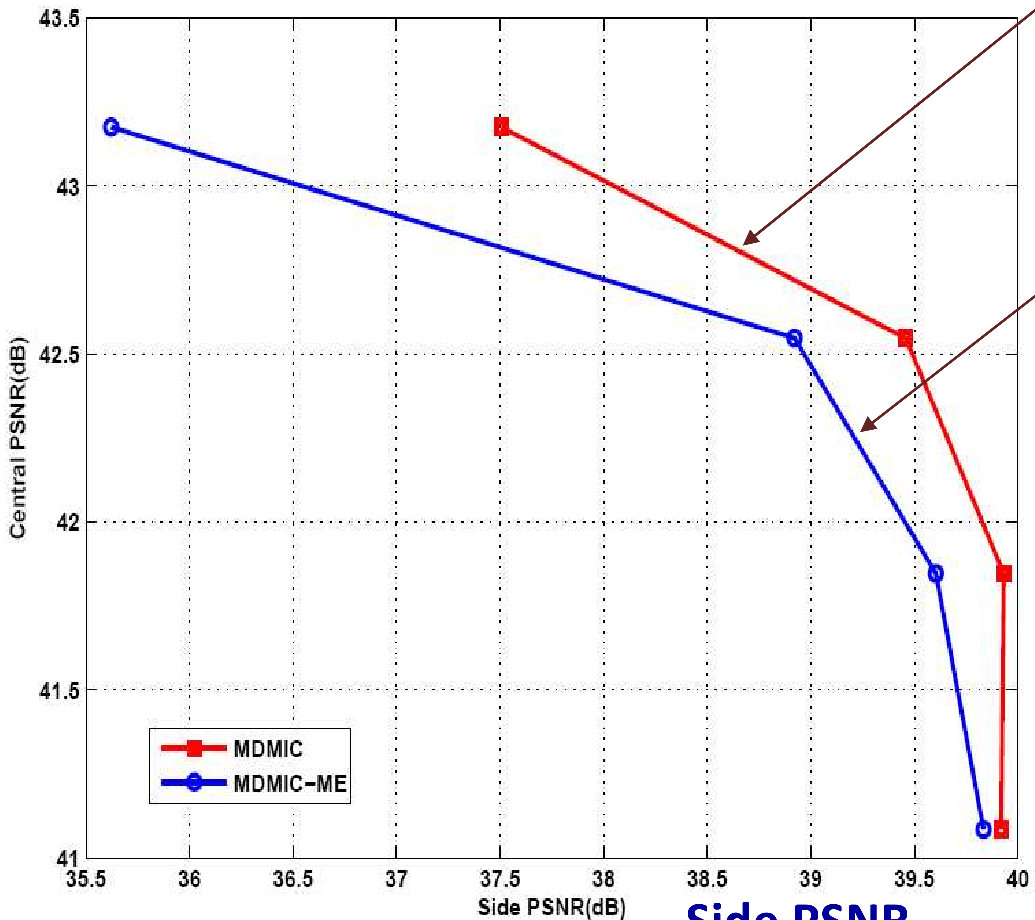
Multiple Description Coding of Multiview Images

- **MDC of multiview images with JMVC:**
  Ballroom sequence, 0.75 bpp per description

View interpolation
coded secondary views

JMVC inter frame coded
secondary views

Central PSNR



Side PSNR

Conclusions

- **Summary:**
  - Theoretical model for rectification-based view interpolation and the corresponding MVC
  - Performance in MVC with view interpolation and video extrapolation
  - Multiple description coding of multiview images

- **Current work:**
  - Refinement of the R-D model
  - Coding gain for all views

- **Acknowledgements:**
  - NSERC New Media Initiative Grant
  - Xiaoyu Xiu
  - Derek Pang

References

1. X. Xiu and J. Liang, "Projective Rectification-based View Interpolation for Multiview Video Coding and Free Viewpoint Generation," Proc. Picture Coding Symposium, Chicago, IL, USA, May 2009.
2. X. Xiu and J. Liang, "Rate-Distortion Analysis of Rectification-based View Interpolation for Multiview Video Coding," Proc. IEEE International Conference on Multimedia & Expo, New York, pp. 13-16, June, 2009.
3. D. Pang, X. Xiu and J. Liang, "Multiview Video Coding Using Projective Rectification-Based View Extrapolation and Synthesis Bias Correction," Proc. IEEE International Conference on Multimedia & Expo, New York, pp. 5-8, June, 2009.
4. X. Xiu, and J. Liang, "Multiple Description Coding of Multiview Images," 2009 IEEE Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, Nov. 2009, submitted, May 2009.

References

5. M. Flierl, A. Mavlankar, and B. Girod, “Motion and Disparity Compensated Coding for Multi-View Video,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 17, no. 11, pp. 1474 – 1484, November 2007.
6. P. Ramanathan and B. Girod, “Rate-Distortion Analysis for Light Field Coding and Streaming,” EURASIP Signal Processing: Image Communication, vol. 21, no. 6, pp. 462-475, July 2006.
7. M. Magnor, P. Ramanathan, and B. Girod, "Multiview Coding for Image-Based Rendering using 3-D Scene Geometry," IEEE Transactions on Circuits and Systems for Video Technology, vol. 13, no. 11, pp. 1092-1106, Nov. 2003.
8. K. Takahashi, and T. Naemura, ”Theoretical model and optimal prefilter for view interpolation,” IEEE International Conf. Image Processing, pp. 1528-1531, San Diego, Oct. 2008.
9. M. Ishii, K. Takahashi, T. Naemura, “Rate-distortion performance of multi-view image coding with subsampling of viewpoints,” IEEE International Conf. Image Processing, pp. 2428-2431, San Diego, Oct. 2008.