



THOMSON
images & beyond

Adaptive Reference Filtering for MVC with Decoder Complexity Analysis

PoLin Lai^b, Antonio Ortega^b
Purvin Pandit^a, Peng Yin^a, Cristina Gomila^a

^aThomson Corporate Research, Princeton, NJ

^bSignal and Image Processing Institute, USC, Los Angeles, CA

Outline

- **Introduction**

- Multi-view video coding: *inter-view redundancy*
- *Focus mismatch* across different views
- Depths-dependent blurriness/sharpness changes

- **The Proposed Approach**

- Review JVT-W065 (5x5 horz and vert symmetric filter)
- New Proposal based on 3x3 symmetric filter

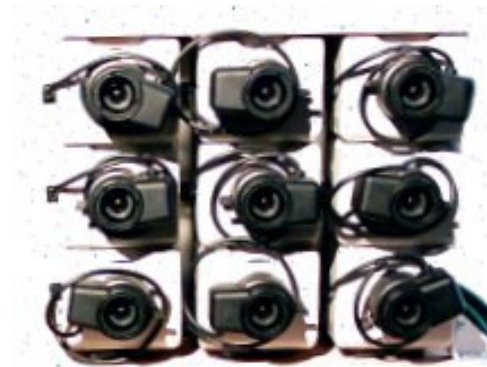
- **Complexity Analysis**

- **Simulation Results**

- **Summary**

Introduction: Multi-view video system

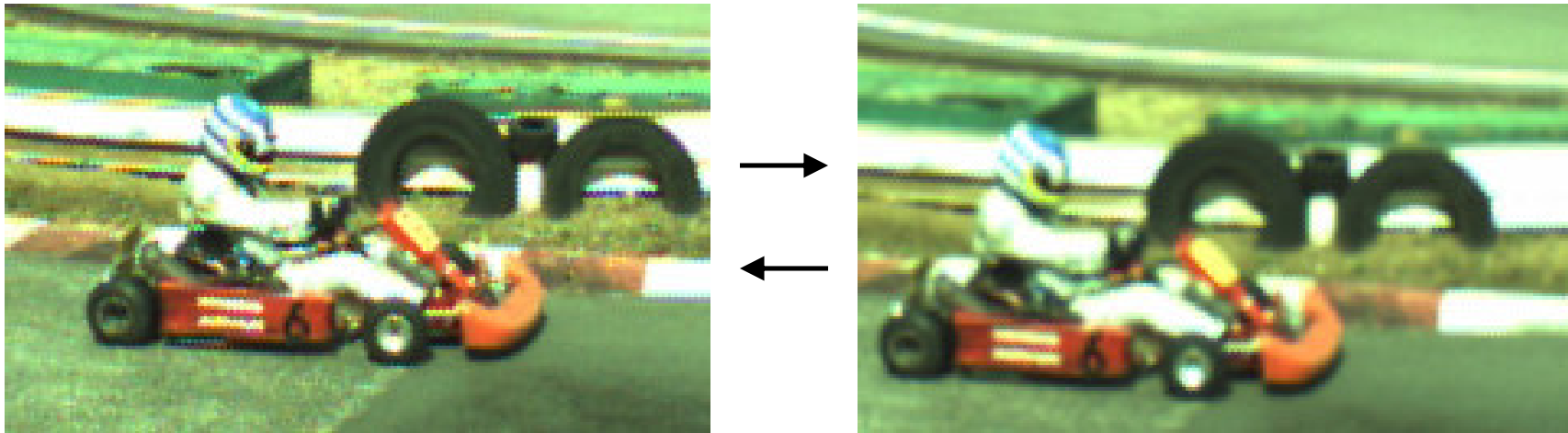
- Scene are captured simultaneously by multiple cameras
- Different camera arrangement: **1D line**, arc, 2D array



Problem: Design considerations

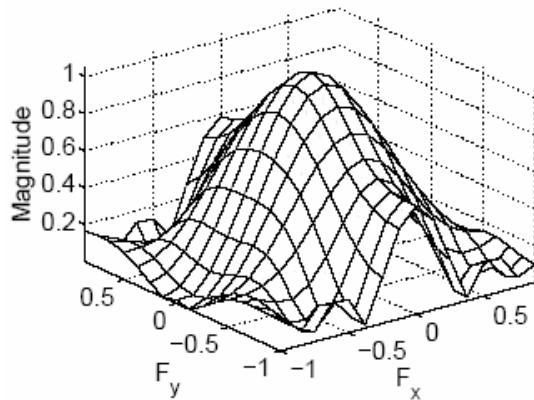
- Generate better match to perform inter-view prediction
- *Depth-dependent* local compensation
- *Adaptive* to the mismatches
 - Different types of focus changes across views
 - Scenes change over time
- Selection of compensation based on *rate-distortion* (RD)
- Low complexity design

Adaptive reference filtering: Example (Race1)

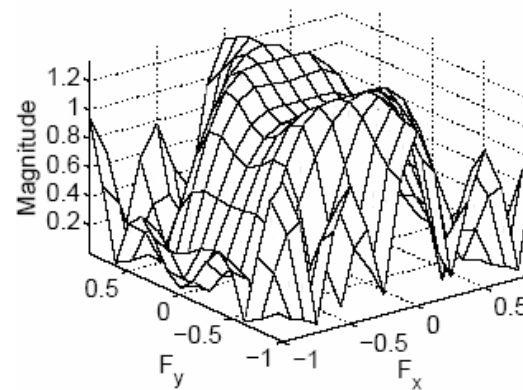


$$\min_{\psi} \sum_{x,y} (S_{x,y} - \psi * R_{x+dv_x,y+dv_y})^2$$

View 2 -> View 3, frame 10



View 5 -> View 6, frame 10



JVT-X060

ARF: Calculate filter coefficients, generate new references

- Filter coefficients for each depth level D_k

$$\min_{\psi} \sum_{(x,y) \in D_k} \left(S_{x,y} - \sum_{j=-n}^n \sum_{i=-m}^m \psi_{ij} R_{x+mv_x+i, y+mv_y+j} \right)^2$$

- Symmetric filters

$$\psi = \begin{pmatrix} a & b & c & b & a \\ d & e & f & e & d \\ g & h & j & h & g \\ d & e & f & e & d \\ a & b & c & b & a \end{pmatrix}$$

- Original unfiltered reference and multiple filtered references ψ^*R , each can also be interpolated for subpel disparity compensation

Complexity Analysis

- **Two stage filtering**
 - 2D Adaptive filters
 - AVC interpolation filters
 - Overall complexity needs to consider both
- **Consider both 4x4 and 8x8 blocks:**
 - MVC application is mainly high-resolution
 - 8x8 block-size is more likely selected for high res
- **Two filtering approaches**
 - **On-the-fly approach**
 - Filter whole frame approach

Complexity of AVC interpolation filter

- **Complexity of AVC interpolation (Vatis et. all)**
 - Num Ops for each sub-pel position evaluated
- **Dependent of the block size**
 - Average for all positions: 21.16 ops/pel for 4x4
 - Average for all positions: 19.20 ops/pel for 8x8
- **Profiling shows that interpolation takes about 20% of decode cycles**

Complexity Analysis of ARF filters

- Compare the number of ops/pel for ARF
 - 5x5 case (JVT-W065)
 - 3x3 case (Current Proposal)

$\psi =$	a_8 a_6 a_4 a_6 a_8 a_7 a_5 a_3 a_5 a_7 a_2 a_1 a_0 a_1 a_2 a_7 a_5 a_3 a_5 a_7 a_8 a_6 a_4 a_6 a_8	Pixels:	A B C D E F G H I J K L M N O P Q R S T U V W X Y	$\psi =$	a_3 a_2 a_3 a_1 a_0 a_1 a_3 a_2 a_3	Pixels:	A B C D E F G H I
Symmetric 5x5 filter				Symmetric 3x3 filter			

- O_{ARF} : 34 additions and 1 shift
 - Total 35 operations

- O_{ARF} : 13 additions and 1 shift
 - Total 14 operations

Overall Operations

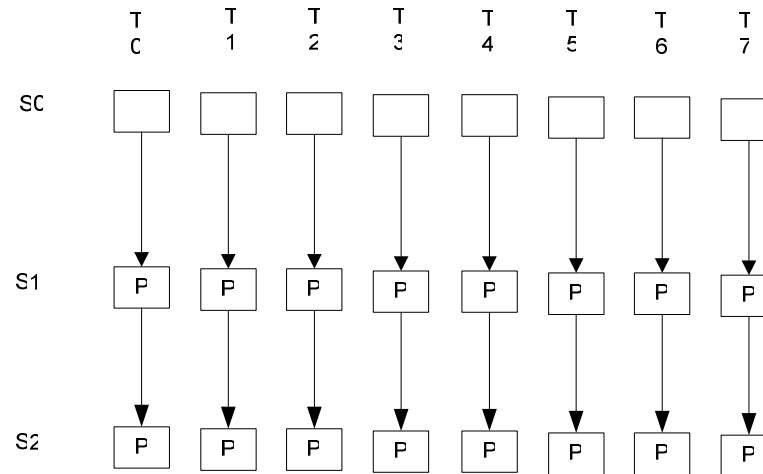
- **Average operation of subpel interpolation on ARF ref:**
 - $a \times O_{\text{ARF}} + b$
- **Consider two possible case of both ARF ref and non-ARF ref:**
 - $p \times (a \times O_{\text{ARF}} + b) + (1-p) \times b$
- **Observation of ARF on test sequences**
 - About $p=20\%-60\%$ blocks select ARF filtered references
 - Depends on bitrate and sequence

Overall Operation Summary

4x4 block size	Ratio of ARF on-the-fly over AVC Interp' (P = 20% ~ 60%)
Hor/ver 5x5 ARF	2.11 ~ 4.33
Hor/ver 3x3 ARF	1.44 ~ 2.33

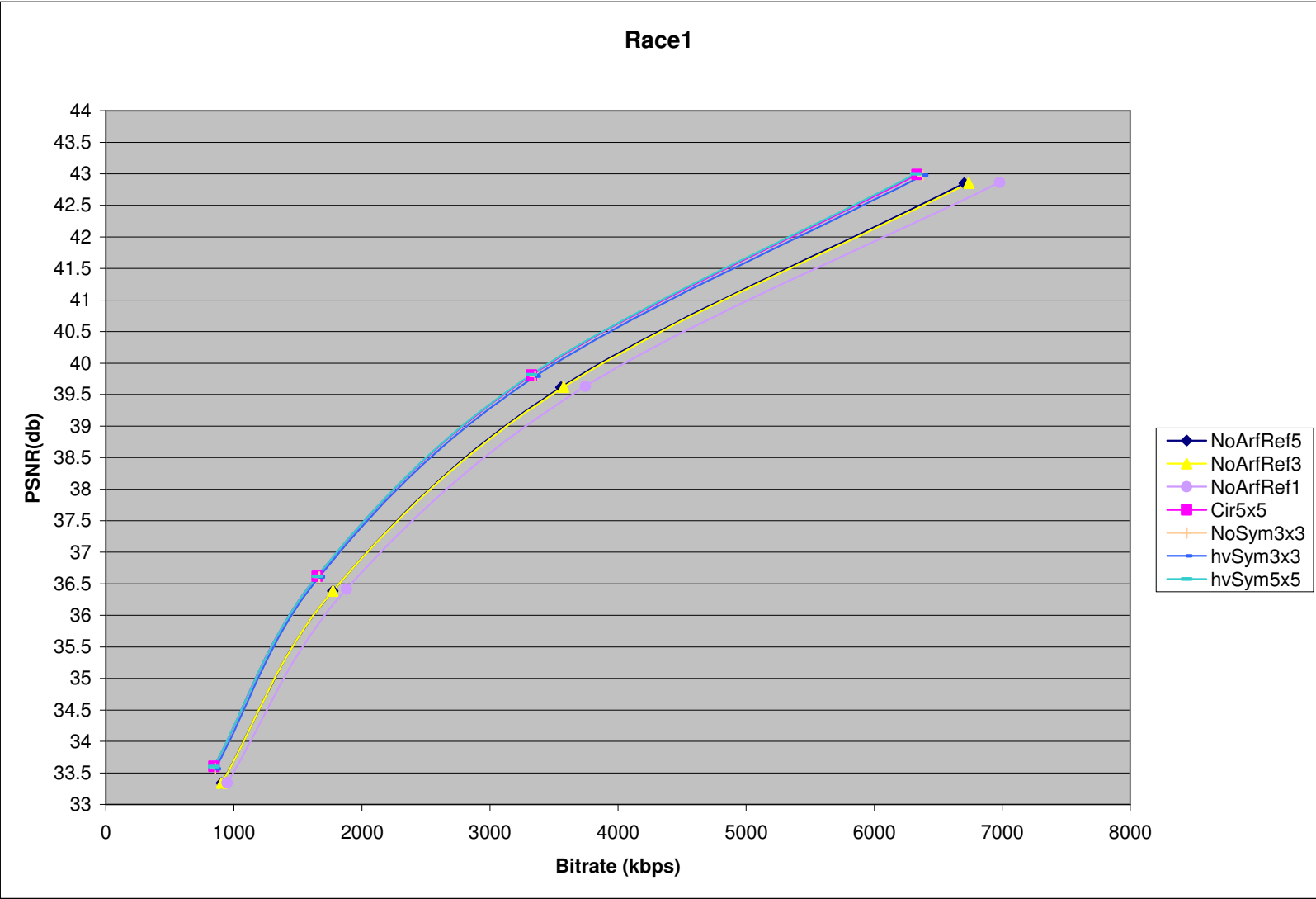
8x8 block size	Ratio of ARF on-the-fly over AVC Interp' (P = 20% ~ 60%)
Hor/ver 5x5 ARF	1.75~ 3.25
Hor/ver 3x3 ARF	1.3 ~ 1.9

Simulations: inter-view prediction

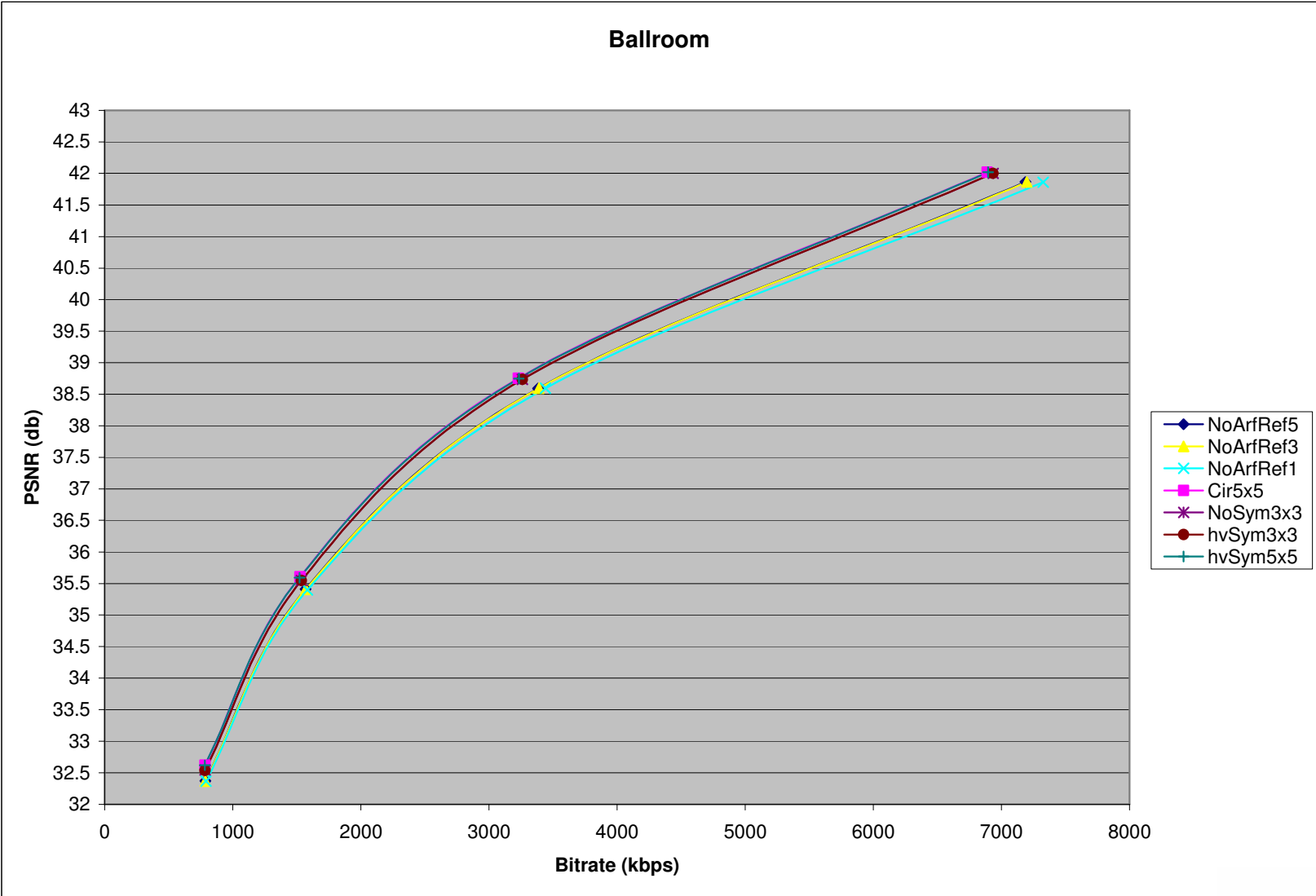


- **Anchor only encoding**
 - IPPPP structure
- **Implemented in JMVM_3_0**
- **Classification tool incorporated in JMVM**
- **IC turned off**
- **Multiple inter-view Reference enabled**
- **Comparison with noARF with Ref = 1, 3, 5**

Race 1



Ballroom



Results Summary

- **Average gains with Flamenco2**
 - Ref 1: 8.5% on BR, 0.385db on PSNR
 - Ref 3: 2.29% on BR, 0.083db on PSNR
 - Ref 5: 0.86% on BR, 0.015db on PSNR
- **Average gains without Flamenco2**
 - Ref 1: 7.73% on BR, 0.392db on PSNR
 - Ref 3: 4.63% on BR, 0.198db on PSNR
 - Ref 5: 3.78% on BR, 0.156db on PSNR
- **Min gain (without Flamenco2)**
 - 1.4% on BR, 0.06db on PSNR (Ref 1)
- **Max gain (without Flamenco2)**
 - 13.67% on BR, 0.69db on PSNR (Ref 1)
 - 9.71% on BR, 0.48dB on PSNR (Ref 5)

Summary

- **Consider Focus mismatch**
- **Adaptive reference filtering based on disparity**
 - Propose to use 3x3 filters to reduce complexity
 - Complexity of operation increases relative to AVC interpolation
 - $R = 1.4 \sim 2.3$ (4x4 block)
 - $R = 1.3 \sim 1.9$ (8x8 block)
- **For inter-view prediction, simulation results outperform JMVM in majority of the cases**
 - Result can be improved more if jointly considering ARF and multi-ref
- **Propose JVT to create a CE to study the tool**

Adaptive Reference Filtering for MVC

PoLin Lai^b, Antonio Ortega^b
Purvin Pandit^a, Peng Yin^a, Cristina Gomila^a

^aThomson Corporate Research, Princeton, NJ

^bSignal and Image Processing Institute, USC, Los Angeles, CA