

VCIP 2007, San Jose, CA

Adaptive Filtering for Cross-View Prediction in Multi-View Video Coding

PoLin Lai^{a,b}, Yeping Su^a, Peng Yin^a, Cristina Gomila^a, and Antonio Ortega^b

^aThomson Corporate Research, Princeton, NJ

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

Outline of the presentation

- **Introduction**
 - Multi-view video coding: *Cross-view redundancy*
 - *Focus mismatch* across different views
 - Depths-dependent blurriness/sharpness changes
- **Review: Reference Frame Filtering**
 - Predefined filter set
 - Adaptive interpolation filters
- **The Proposed Approach**
 - Adaptive reference filtering design based on *disparity*
- **Simulation Results**
- **Summary**

Introduction: Multi-view video system

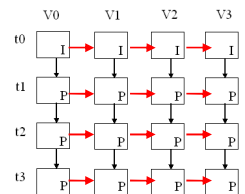
- Scene are captured simultaneously by multiple cameras
- Different camera arrangement: **1D line**, arc, 2D array



- **Possible applications:**
 - Interactive free view-point applications (TV-broadcasting)
 - Immersive virtual reality (entertainment, military training,.....)

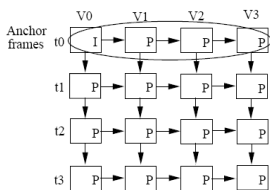
Introduction: Multi-view video coding (MVC)

- **Simulcast: Encode each view independently**
 - Data size proportional to the number of view
 - Cross-view redundancy not considered!!
- **MPEG Call for Proposals (Jul 2005)**
→ Moved to JVT (Jul 2006 Meeting)
- **Combine "Disparity Compensation" and motion compensation**
 - Higher coding efficiency
 - (Responses received to CfP, MPEG Document M12969)



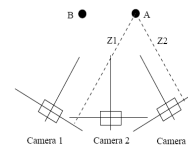
Introduction: Multi-view video coding (MVC)

- **Simulcast: Encode each view independently**
 - Data size proportional to the number of view
 - Cross-view redundancy not considered!!
- **MPEG Call for Proposals (Jul 2005)**
→ Moved to JVT (Jul 2006 Meeting)
- **Combine "Disparity Compensation" and motion compensation**
 - Higher coding efficiency
 - (Responses received to CfP, MPEG Document M12969)



Problem: Focus mismatches in cross-view prediction

- **Camera calibration**
 - Different focus settings, intrinsic parameters
- **Different shooting positions / orientations**
 - Same object appears at *different scene depths* across views



- **Localized focus mismatches: Depth-dependent!!!**

Problem: Design considerations

- **Generate better match to perform cross-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)**

Page N7



Outline of the presentation

- **Introduction**
 - Multi-view video coding: *Cross-view redundancy*
 - *Focus mismatch* across different views
 - Depths-dependent blurriness/sharpness changes
- **Review: Reference Frame Filtering**
 - Predefined filter set
 - Adaptive interpolation filters
- **The Proposed Approach**
 - Adaptive reference filtering design based on *disparity*
- **Simulation Results**
- **Summary**

Page N8



Review: Reference frame filtering

- **Fixed set of blurring filters (M. Budagavi: ICIP 2005)**
 - Sharpening filters not considered
 - Predefined set not adaptive to frame difference
- **Adaptive interpolation filters for sub-pixel reference**

T. Wedi (ICIP 2002, IEEE Trans. CSVT 2003 2006)
 Y. Vatis etc. (ICIP 2005, VCIP 2005, ICIP 2006) *Multiple adaptive filters*

 - Specifically designed to generate sub-pel reference
 - Filters are estimated based on the difference between Ref. and Cur. frame

Page N9



Review: Adaptive interpolation filtering by Y. Vatis etc (AIF)

- **Y. Vatis etc: ICIP 2005, VCIP 2005, ICIP 2006**
 - Initial motion estimation (with H.264/AVC 6-tap filters for sub-pixels)
 - Divide frames based on subpel portion (SP) of the motion vectors
 - One filter for each location: MMSE estimation
 - For example: (1/2, 23/4) and (45/2, 6/4) will be assigned the same filter
 - Generate subpel reference using filters
 - Final motion compensation

c	s	b	c
d	e	f	g
h	i	j	k
l	m	n	o

$$e^{SP}{}^2 = \sum_{x,y} \left(S_{x,y} - \sum_{i,j=1}^6 h_{i-1,j-1}^{SP} P_{x+i,y+j} \right)^2$$

Page N10



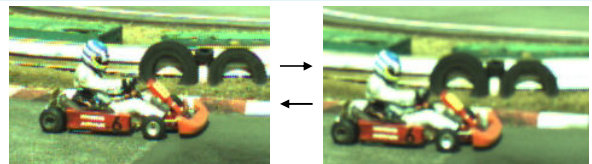
Outline of the presentation

- **Introduction**
 - Multi-view video coding: *Cross-view redundancy*
 - *Focus mismatch* across different views
 - Depths-dependent blurriness/sharpness changes
- **Review: Reference Frame Filtering**
 - Predefined filter set
 - Adaptive interpolation filters
- **The Proposed Approach**
 - Adaptive reference filtering design based on *disparity*
- **Simulation Results**
- **Summary**

Page N11

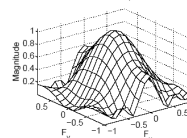


Adaptive reference filtering: An example (Race1)

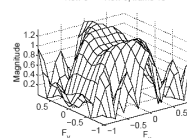


$$\min_{\psi} \sum_{x,y} (S_{x,y} - \psi * R_{x+dv_x,y+dv_y})^2$$

View 2 -> View 3, frame 10



View 5 -> View 6, frame 10



Page N12



Adaptive reference filtering (ARF) for cross-view prediction

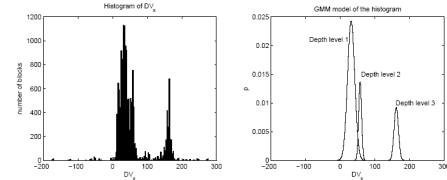
- **Filter association: Depth-dependent filter design**
 - Separate object into different scene-depth levels
 - Object depth Z and disparity d are reciprocals
 - Partition image blocks into different 'depth levels D_k ' based on disparity field
- **Calculate filters' coefficients**
 - For each depth level D_k :
$$\min_{\psi} \sum_{(x,y) \in D_k} \left(S_{x,y} - \sum_{j=-n}^n \sum_{i=-m}^m \psi_{ij} R_{x+mv_x+i,y+mv_y+j} \right)^2$$
- **Generate new references with these adaptive filters**
- **Final disparity compensation**

Page N°13



ARF: Filter association

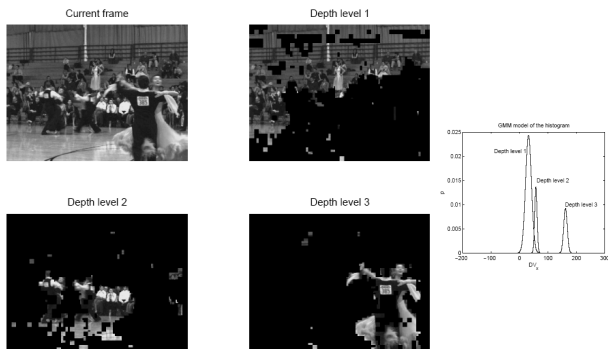
- **Partition frames into different depth levels**
 - Classify disparity vectors obtained by video encoders (x-component for 1D)
- **EM classification based on Gaussian Mixture Model (GMM)**
 - Unsupervised classification tool (**Bouman, Purdue Univ.**)
 - Specify the maximum number of Gaussian components: K
 - For $k = K, K-1, \dots, 1$: EM classification, estimate GMM with k components
 - Select model with lowest minimum description length (MDL)



Page N°14



ARF: Filter association (cont.)



Page N°15



ARF: Calculate filter coefficients, generate new references

- **Filter coefficients for each depth level D_k**

$$\min_{\psi} \sum_{(x,y) \in D_k} \left(S_{x,y} - \sum_{j=-n}^n \sum_{i=-m}^m \psi_{ij} R_{x+mv_x+i,y+mv_y+j} \right)^2$$
- **Focus changes typically model as isotropic \rightarrow 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**

Page N°16



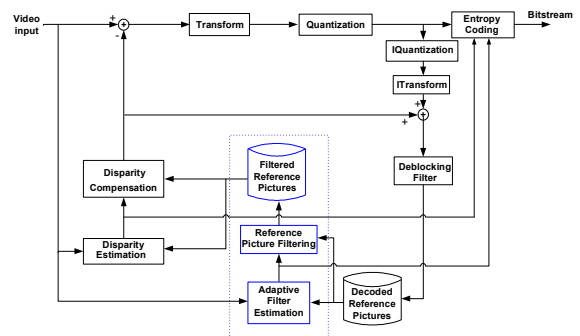
ARF: Encoding and signaling

- **Filter coefficients**
 - Quantized and encoded as frame-level overhead (Y. Vatis: VCIP 2005)
- **Filter selection**
 - Filtered references by adaptive filters
 - Utilize multiple-reference routine as H.264/AVC
 - Rate-distortion based reference selection
 - The selection is signaled by 'ref_idx' as in H.264/AVC

Page N°17



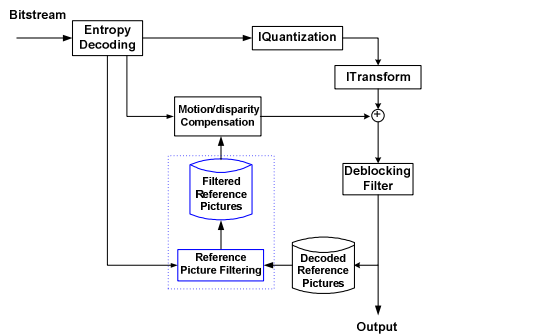
ARF: Encoder block diagram



Page N°18



ARF: Decoder block diagram



Page N°19

THOMSON

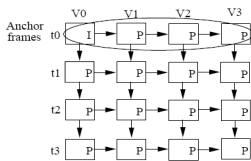
Outline of the presentation

- Introduction
 - Multi-view video coding: *Cross-view redundancy*
 - *Focus mismatch* across different views
 - *Depths-dependent blurriness/sharpness* changes
- Review: Reference Frame Filtering
 - Predefined filter set
 - Adaptive interpolation filters
- The Proposed Approach
 - Adaptive reference filtering design based on *disparity*
- Simulation Results
- Summary

Page N°20

THOMSON

Simulations: Cross-view prediction



- At a given time stamp
 - Feed frames from different views as time series
 - Disparity compensation is performed by motion compensation tools
- Implement code with H.264/AVC (Software: JM 10.2)
- Full search ± 64 pixels, QP:28 \rightarrow 32 \rightarrow 36 \rightarrow 40

Page N°21

THOMSON

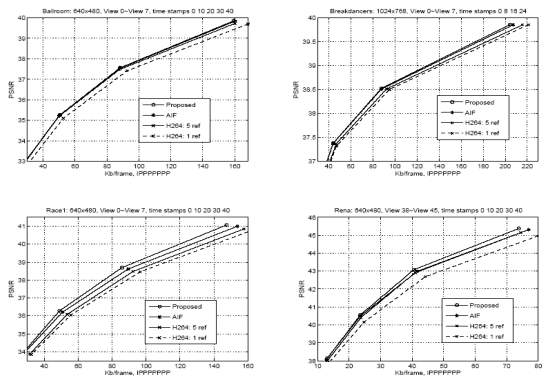
Simulations: Test sequences



Page N°22

THOMSON

Simulation results



Page N°23

THOMSON

Summary

- Consider *cross-view redundancy*: Higher coding efficiency
- Focus mismatch
 - Heterogeneous cameras, shooting positions / orientations
 - *Depth-dependent* focus changes across views
- Adaptive reference filtering based on disparity
 - Use *disparity* to estimate depth
 - Classify image blocks into different depth levels (EM algorithm for GMM)
 - Multiple adaptive filters for reference frame
- For cross-view prediction, simulation results outperform H.264/AVC

Page N°24

THOMSON