USC Viterbi School of Engineering

VCIP 2008
San Jose, CA

Focus Mismatches in Multiview Systems and Efficient Adaptive Reference Filtering for Multiview Video Coding

PoLin Lai[*,+], Antonio Ortega[*], Purvin Pandit[+], Peng Yin[+], and Cristina Gomila[+]

[*]Signal and Image Processing Institute, USC, Los Angeles, CA
[+]Thomson Corporate Research, Princeton, NJ

USC

USC Viterbi School of Engineering

Presentation outline

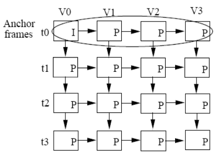
- **Introduction**
  - Multiview video coding: Inter-view redundancy, mismatches
  - Adaptive reference filtering (ARF) for MVC Inter-view
- **Analysis of focus mismatches**
  - Characteristics of images captured with a lens
  - Focus mismatches in multiview system
- **Efficient "adaptive reference filtering" techniques**
  - View-wise adaptation based on rate-distortion prediction
  - Filter updating based on scene-depth composition
- **Simulation Results**
- **Summary**

USC

USC Viterbi School of Engineering

Introduction: Multiview Video Coding (MVC)

- **Multiple cameras, different viewpoints**
  - 1D horizontal: parallel, arc; 2D camera array
  - Free view point TV applications, immersive virtual reality
- **Redundancy across different views**
  - *Inter-view prediction*: Block-based disparity compensation
  - Inter-view + temporal → Higher coding efficiency
    (CSVT special issue on MVC, Nov 2007)
  - "Anchor frames": Encoded with *inter-view prediction only*



USC

USC Viterbi School of Engineering

Introduction: Inter-view coding in MVC

- **Mismatches beyond simple displacement**
  - Multiple cameras: Settings, intrinsic parameters
  - Different viewpoint: Object depth, light radiation

  → **Affine projection, illumination, color, focus mismatches**……

USC

USC Viterbi School of Engineering

Introduction: Adaptive reference filtering (ARF)

- **Non-translational transforms in inter-view prediction**
  - J.-H. Kim, P. Lai, J. Lopez, A. Ortega, Y. Su, P. Yin, and C. Gomila *Tans. CSVT 2007* (Special Issue MVC)
  - P. Lai, Y. Su, P. Yin, C. Gomila, and A, Ortega. *Proc. VCIP 2007*
- **ARF: 2-pass coding scheme**
  - **Identify regions with possibly different mismatches**

    Initial disparity search

    Frame partition: S → $S^1, S^2, \ldots S^n$
  - **Design compensation kernel for each region**

$$\min_{\psi^i} \sum_{x,y} \left(S^i_{x,y} - \psi^i * R_{x+dv_x, y+dv_y}\right)^2$$

  - **Final encoding with new filtered references**

USC

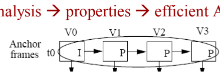
USC Viterbi School of Engineering

Introduction: Questions regarding ARF

- **ARF: 2-pass coding scheme**

  Data driven (*No prior knowledge)*

  → Higher complexity
- **Knowledge about the mismatch!!**
  - How to identify regions?
  - Type of compensation kernel? ψ^i
  - When to use ARF?
  - When to update filters?
- **Target of our work: Focus mismatches**
  - Analysis → properties → efficient ARF



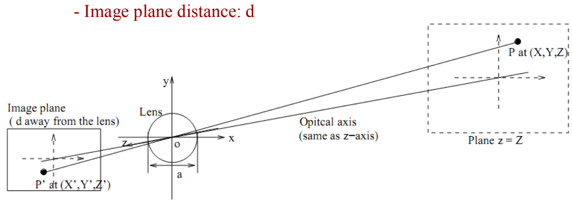
USC

Presentation outline

- Introduction
  - Multiview video coding: Inter-view redundancy
  - Inter-view coding with focus mismatches
- **Analysis of focus mismatches**
  - Characteristics of images captured with a lens
  - Focus mismatches in multiview system
- Efficient "adaptive reference filtering" techniques
  - View-wise adaptation based on rate-distortion prediction
  - Filter updating based on scene-depth composition
- Simulation Results
- Summary

Characteristics of lens imaging (1)

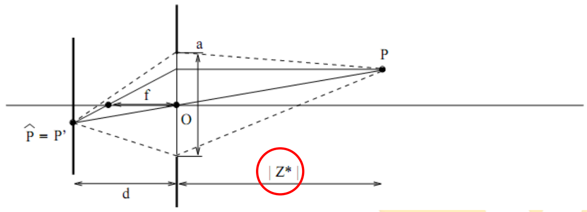
- Focal length: f
- Aperture diameter: a
- Image plane distance: d



$$(X', Y', Z') = \left(\frac{d}{Z}X, \frac{d}{Z}Y, d\right)$$

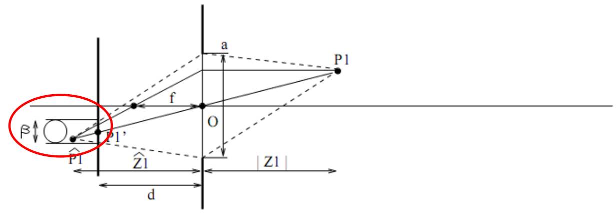
Characteristics of lens imaging (2)

- **Image plane with d**
  - Perfect in-focus (point-projection): $\frac{1}{|Z^*|} + \frac{1}{d} = \frac{1}{f} \Rightarrow |Z^*| = \frac{d \cdot f}{d - f}$
  - Other than |Z*| → ***Uniform circle***, diameter β



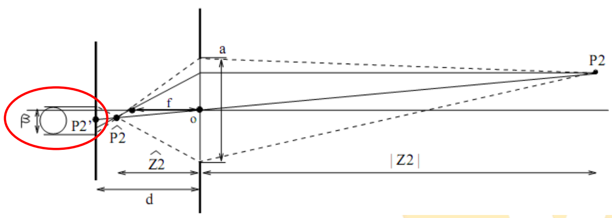
Characteristics of lens imaging (2)

- **Image plane with d**
  - Perfect in-focus (point-projection): $\frac{1}{|Z^*|} + \frac{1}{d} = \frac{1}{f} \Rightarrow |Z^*| = \frac{d \cdot f}{d - f}$
  - Other than |Z*| → ***Uniform circle***, diameter β



Characteristics of lens imaging (2)

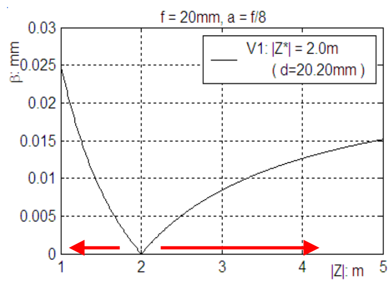
- **Image plane with d**
  - Perfect in-focus (point-projection): $\frac{1}{|Z^*|} + \frac{1}{d} = \frac{1}{f} \Rightarrow |Z^*| = \frac{d \cdot f}{d - f}$
  - Other than |Z*| → ***Uniform circle***, diameter β



Characteristics of lens imaging (3)

- **β as a function of "object depth Z"**

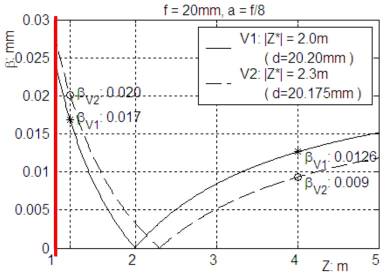
$$\beta = \frac{af(|Z - Z^*|)}{|Z|(|Z^*| - f)}$$



Focus mismatch in multiview system (1)

- **Focus setting difference ($|Z^*_1| \neq |Z^*_2|$)**

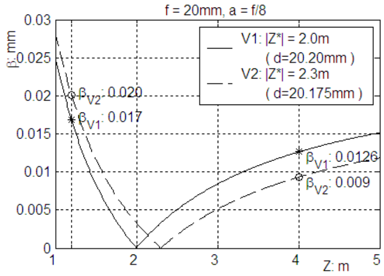
$$\beta = \frac{af\,(|Z - Z^*|)}{|Z|\,(|Z^*| - f)}$$



Focus mismatch in multiview system (1)

- **Focus setting difference ($|Z^*_1| \neq |Z^*_2|$)**

$$PSF_Z(r) = \begin{cases} 4/(\pi\beta^2), & \text{if } r^2 \le (\beta/2)^2 \\ 0, & \text{otherwise} \end{cases} \rightarrow OTF_Z(q) = \frac{2J_1(\pi\beta q)}{\pi\beta q}$$

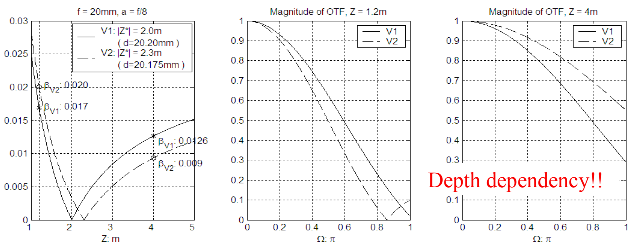


Focus mismatch in multiview system (2)

- **Difference in β curves:**

$$PSF_Z(r) = \begin{cases} 4/(\pi\beta^2), & \text{if } r^2 \le (\beta/2)^2 \\ 0, & \text{otherwise} \end{cases} \rightarrow OTF_Z(q) = \frac{2J_1(\pi\beta q)}{\pi\beta q}$$

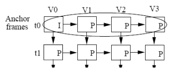
Circular symmetric!!

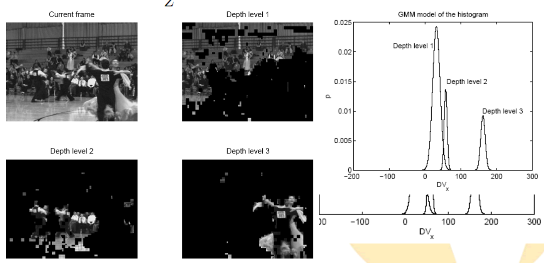


Properties of inter-view focus mismatches

- **Depth-dependent mismatches**
  - Design compensation ***kernel for each depth-level***
  - 1D parallel, spacing ***b*** between cameras

  → Disparity $\delta_Z = \frac{d}{Z}(-b)$ → Exploit disparity to partition frames





Properties of inter-view focus mismatches

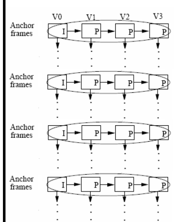
- **Circular symmetric mismatch (isotropic)**

  → Impose ***constraints*** to the compensation kernel

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

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

Properties of inter-view focus mismatches



- **View dependency:** $\beta = \frac{af\,(|Z - Z^*|)}{|Z|\,(|Z^*| - f)}$
  - Exhibited mismatches dependents on difference in β curves
  - ***Avoid*** applying 2-pass ARF to views with no focus mismatches

- **For a given view-pair:**
  - At different time, an object at $|Z|$ results in *same type* of mismatch
  - Across time, when scene is composed with similar depth-levels

    → No need to update filters

    → **Update filters only when *depth-composition* changes**

USC Viterbi School of Engineering

Presentation outline

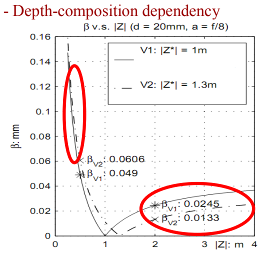
- Introduction
  - Multiview video coding: Inter-view redundancy
  - Inter-view coding with focus mismatches
- Analysis of focus mismatches
  - Characteristics of images captured with a lens
  - Focus mismatches in multiview system
- **Efficient "adaptive reference filtering" techniques**
  - View-wise adaptation based on rate-distortion prediction
  - Filter updating based on scene-depth composition
- Simulation Results
- Summary

USC

USC Viterbi School of Engineering

View wise: ARF performance in inter-view coding (1)

- **From analysis, focus mismatches:**
  - View dependency
  - Depth-composition dependency



→ Expect ARF performance also affected by these factors

USC

USC Viterbi School of Engineering

View wise: ARF performance in inter-view coding (1)

- **From analysis, focus mismatches:**
  - View dependency
  - Depth-composition dependency

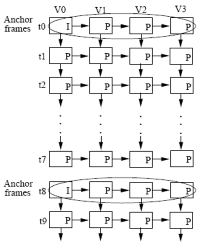
→ Expect ARF performance also affected by these

- **Rate-distortion performance of ARF**
  - Anchor frames with 0.5 second interval
  - RD cost in initial search

    v.s.

    RD cost in the final search (filtered refs)

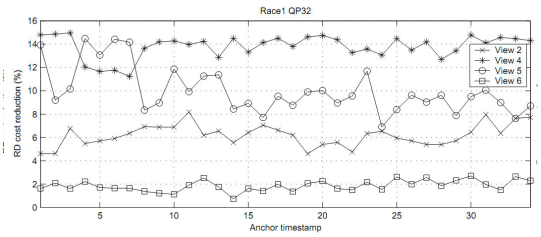
→RD cost reduction (%) = (RD1-RD2) / RD1



USC

USC Viterbi School of Engineering

View wise: ARF performance in inter-view coding (2)



Across view:

- Stronger focus mismatches → larger RD-cost reduction

Across time:

- View with larger gain, larger variation (depth-composition)

USC

USC Viterbi School of Engineering

View-wise ARF adaptation

- **Skip ARF for some views**
  - Based on **observed RD cost reduction**
  - Consistently small reduction, over multiple timestamps?

USC

USC Viterbi School of Engineering

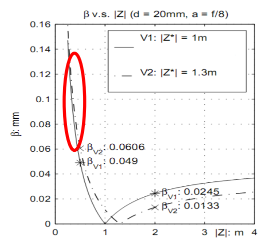
View-wise ARF adaptation

- **Skip ARF for some views**
  - Consistently small reduction, over multiple timestamps?

No focus mismatch

OR

Depth-composition does not reflect the mismatch



USC

View-wise ARF adaptation - Algorithm

- Given view V, for a **period of *N*** anchor timestamps ($cN+1 \sim c(N+1)$)
  - Initially apply ARF to the first *T* timestamps
  - Calculate mean and std of RD cost reductions: $\mu^V_{(cN+1,cN+T)}$ $\sigma^V_{(cN+1,cN+T)}$
  - Apply ARF to the remaining anchor timestamp *i* ?

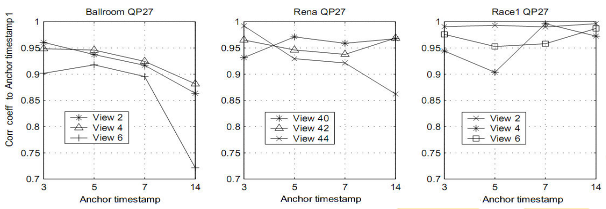
$$\begin{cases} \text{NO,} & \text{if } \mu^V_{(cN+1,cN+T)} < \kappa \text{ and } \sigma^V_{(cN+1,cN+T)} < \epsilon \\ \text{YES,} & \text{otherwise} \end{cases}$$

Across time: Correlation among ARF filters

- **From analysis, for given view pair:**
  - At different time, same depth Z → same type focus mismatch
- **Correlation analysis for ARF filters**

Concatenate filter coefficients estimated at time t_1 to form a coefficient vector $\mathbf{A}_{t1}$, and compare it with the corresponding filter coefficient vector $\mathbf{A}_{t2}$ at another time t_2

Across time: Correlation among ARF filters

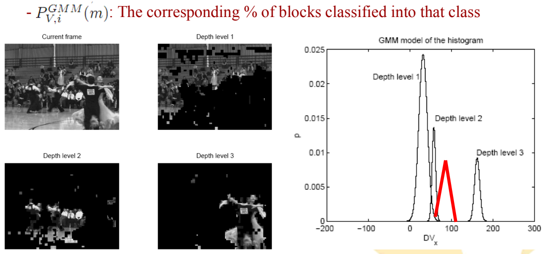


- **Highly correlated when *depth-composition* remain similar**

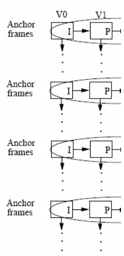
Across time: Correlation among ARF filters

Filter updating based on depth-composition

- Estimate (update) filter only if
  Depth-level composition changes → New depth-level introduced
- Compare GMM of DV at different timestamps
  - $\mu^{GMM}_{V,i}(m)$: Mean of Gaussian component *m*, at time *i* in view *V*
  - $P^{GMM}_{V,i}(m)$: The corresponding % of blocks classified into that class



Filter updating - Algorithm

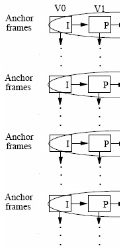


- Compare GMM of DV at different timestamps
  - $\mu^{GMM}_{V,i}(m)$: Mean of Gaussian component *m*, at time *i* in view *V*
  - $P^{GMM}_{V,i}(m)$: The corresponding % of blocks classified into that class

$\mu^{GMM}_{V,i}(m)$ **"not covered" (new depth-level)** in reference time *r*, if :

$$\forall n\ |\mu^{GMM}_{V,i}(m) - \mu^{GMM}_{V,r}(n)| > D$$

Filter updating - Algorithm



- Compare GMM of DV at different timestamps
  - $\mu_{V,i}^{GMM}(m)$: Mean of Gaussian component *m*, at time *i* in view *V*
  - $P_{V,i}^{GMM}(m)$: The corresponding % of blocks classified into that class

$\mu_{V,i}^{GMM}(m)$ **"not covered" (new depth-level)** in reference time *r*, if :

$$W_{V,i}(m) = \begin{cases} 1, & \text{if } \forall n\ |\mu_{V,i}^{GMM}(m) - \mu_{V,r}^{GMM}(n)| > D \\ 0, & \text{otherwise} \end{cases}$$

$$\text{If } \sum_m W_{V,i}(m) \cdot P_{V,i}^{GMM}(m) > P \rightarrow \text{apply two-pass ARF}$$

$$\text{Otherwise} \rightarrow \text{re-use filters}$$

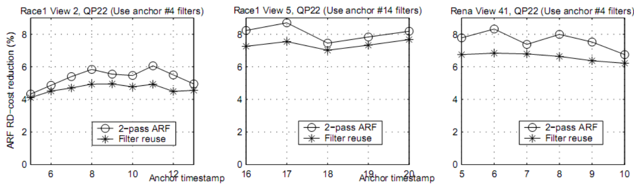
Filter updating - Algorithm

$$W_{V,i}(m) = \begin{cases} 1, & \text{if } \forall n\ |\mu_{V,i}^{GMM}(m) - \mu_{V,r}^{GMM}(n)| > D \\ 0, & \text{otherwise} \end{cases}$$

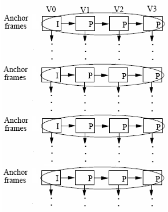
$$\text{If } \sum_m W_{V,i}(m) \cdot P_{V,i}^{GMM}(m) > P \rightarrow \text{apply two-pass ARF}$$

$$\text{Otherwise} \rightarrow \text{re-use filters}$$

We set D = 5 pixels, P = 15% *(Paired)*



Efficient ARF for focus mismatches

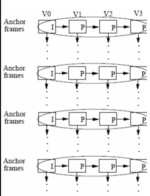


- **Given view, time:**
  - Depth dependency → Partition based on DV
  - Symmetry → Post constraints on filters
- **When to apply?**
  - View-wise: Only to views with ***high gain***
  - → Algorithm: Periodically observe ***RD cost reduction***
- **When to update filters**
  - For given view: Only when ***depth-composition*** changes
  - → Algorithm: Compare GMM of disparity

Presentation outline

- Introduction
  - Multiview video coding: Inter-view redundancy
  - Inter-view coding with focus mismatches
- Analysis of focus mismatches
  - Characteristics of images captured with a lens
  - Focus mismatches in multiview system
- Efficient "adaptive reference filtering" techniques
  - View-wise adaptation based on rate-distortion prediction
  - Filter updating based on scene-depth composition
- **Simulation Results**
- Summary

Simulation Setting



- Anchor timestamps with **0.5 sec intervals**
- View-wise adaptation: **N = 20, T = 4**
  - → Periods of 20 frames, evaluate ARF the first 4 frames
  - Apply ARF to the remaining anchor timestamp *i* ? ($cN+T < i < c(N+1)$)

$$\begin{cases} \text{NO}, & \text{if } \mu^{V}_{(cN+1,cN+T)} < \kappa \text{ and } \sigma^{V}_{(cN+1,cN+T)} < \epsilon \\ \text{YES}, & \text{otherwise} \end{cases}$$

k = 2%, ϵ =1% *(Uli, Exit……)*

- Filter updating: **D = 5 pixels, P = 15%**

$$W_{V,i}(m) = \begin{cases} 1, & \text{if } \forall n\ |\mu_{V,i}^{GMM}(m) - \mu_{V,r}^{GMM}(n)| > D \\ 0, & \text{otherwise} \end{cases}$$

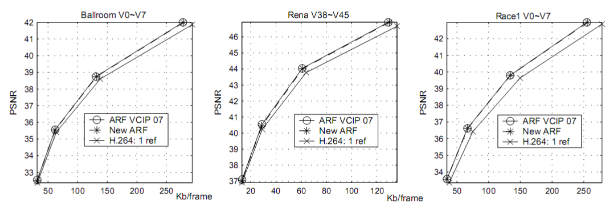
$$\text{If } \sum_m W_{V,i}(m) \cdot P_{V,i}^{GMM}(m) > P \rightarrow \text{apply two-pass ARF}$$

$$\text{Otherwise} \rightarrow \text{re-use filters}$$

Simulation Results

- **20% ~ 35% frames with 2-pass ARF → About 0.05 dB degradation**

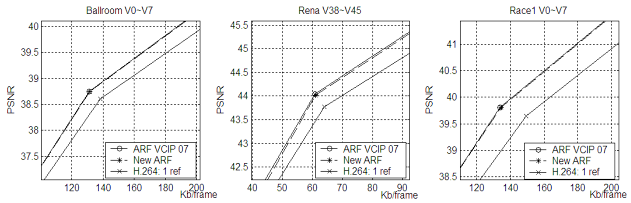
	Ballroom(20 anchors)							Rena(20 anchors)							Race1(35 anchors)						
QP	1	2	3	4	5	6	7	39	40	41	42	43	44	45	1	2	3	4	5	6	7
22			X				X	X				X		X						X	
27			X				X	X						X						X	
32			X	X		X	X	X		X		X	X	X						X	
37		X	X	X	X	X	X	X		X		X	X	X						X	



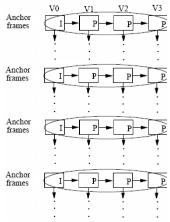
Simulation Results

- **20% ~ 35% frames with 2-pass ARF →About 0.05 dB degradation**

	Ballroom(20 anchors)							Rena(20 anchors)							Race1(35 anchors)						
QP	1	2	3	4	5	6	7	39	40	41	42	43	44	45	1	2	3	4	5	6	7
22			X				X	X				X		X						X	
27			X				X	X						X						X	
32			X	X		X	X	X		X		X	X	X						X	
37		X	X	X	X	X	X	X		X		X	X	X						X	



Summary



- **Two-pass ARF**
  - Data-driven, high complexity → ***Need knowledge!***
- **Focus mismatches in inter-view coding**
  - View-dependency: Setting difference
  - Depth-dependency: Type/degree function of Z
- **View-wise ARF adaptation**
  - Periodically evaluate ARF performance
- **Filter updating based on changes in depth-composition**
  - Compare depth levels among timestamps
- **Negligible degradation (0.05dB) with 20%~35% 2-pass ARF**

Questions?