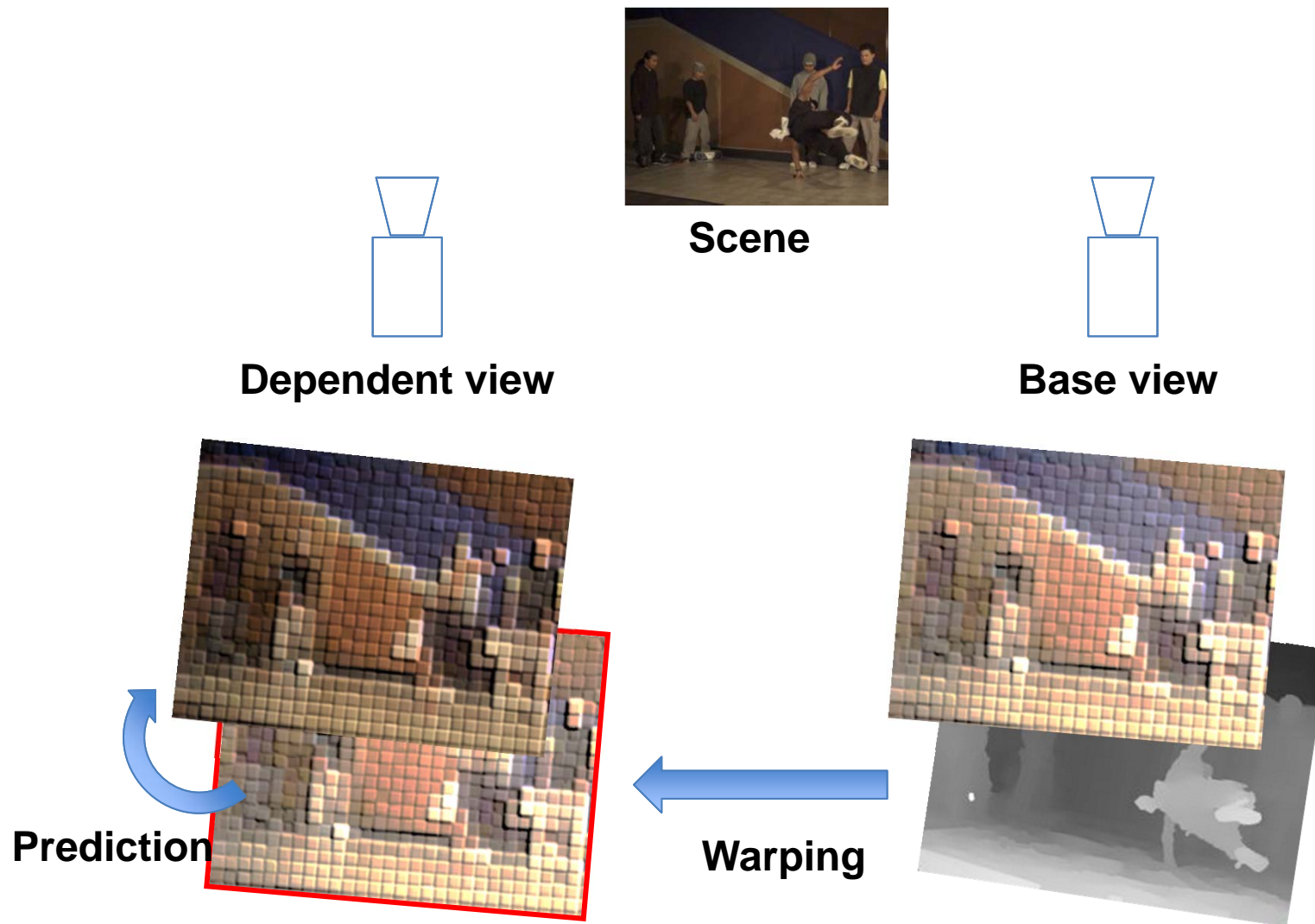JCT3V-C0152

CE1.h: Backward VSP using Neighboring Blocks

Dong Tian, Feng Zou, Anthony Vetro

Mitsubishi Electric Research Labs (MERL)

Cambridge, MA

3rd JCT-3V meeting, Geneva, CH, Jan 2013
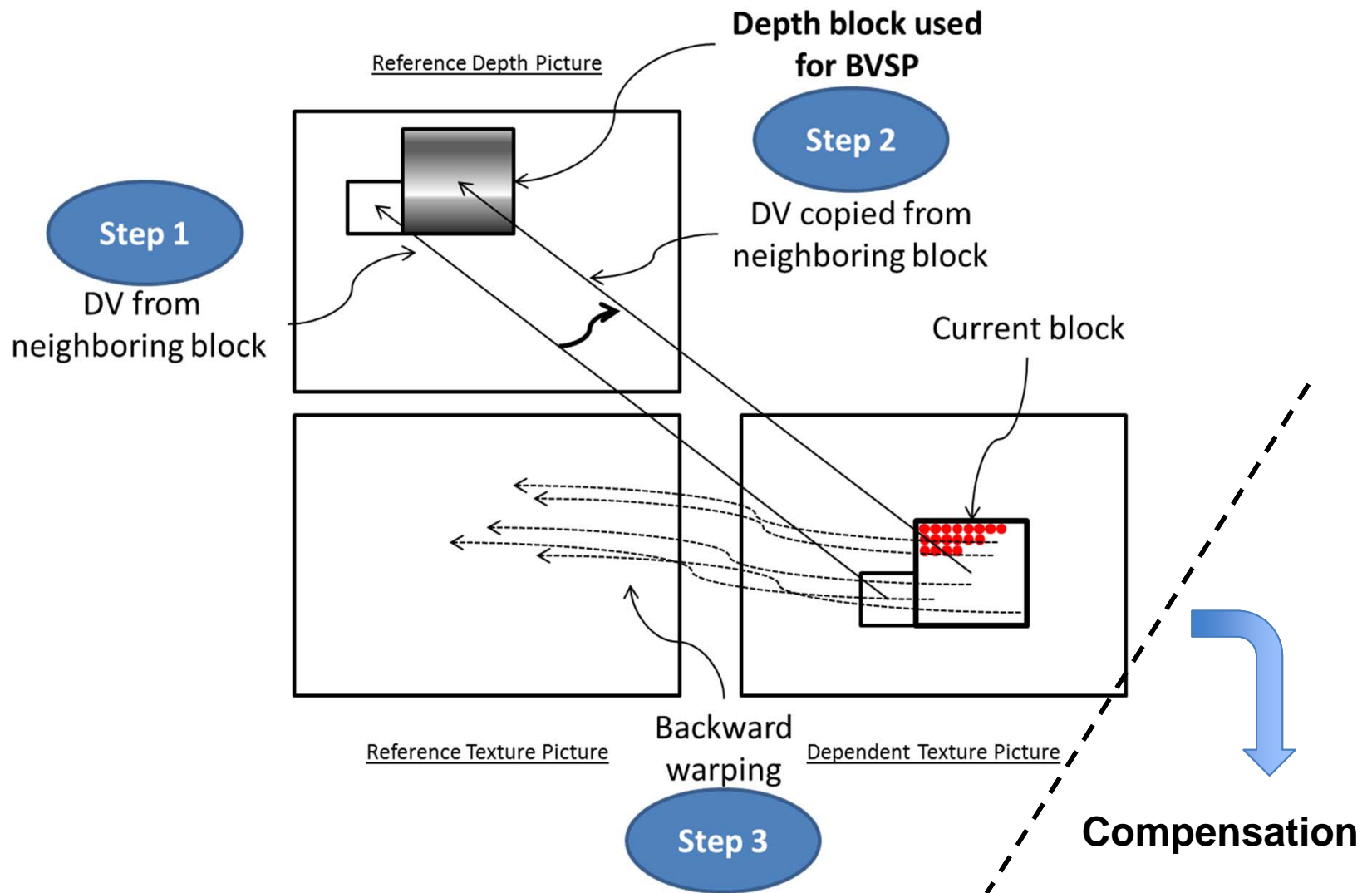
Introduction – View Synthesis Prediction

Scene

Dependent view

Base view

Prediction

Warping

Coding gains of the proposed BVSP

BVSP precision	Video PSNR vs. Video bitrate	Synthesis PSNR vs. Total bitrate	Coded & Synthesis PSNR vs. Total bitrate	Video PSNR vs. Total bitrate	Decoding time increase
1x1	1.2%	0.9%	1.0%	1.0%	3%
2x2	1.1%	0.9%	0.9%	0.9%	4%
4x4	0.9%	0.8%	0.8%	0.8%	3%

Scenario assumptions of the proposed BVSP

- Texture first coding
  - T0D0T1D1
  - Keep the depth coding tools and enhance synthesis quality

- Major challenge
  - BVSP requires depth map from the current viewpoint
  - The depth for the current view is not available

Proposed backward warping – 3 steps

BVSP Implementations

- Warping /3D geometric links pixel locations from one viewpoint to another

- Pixel values – Spatial domain
  - Motion/disparity compensation related function
  - Pixel level, 1x1
  - Sub-PU level, 2x2 and 4x4

- Motion field – Temporal domain
  - Interview motion prediction in HTM
    - NBDV function related
  - PU level

- Propose a new merge candidate to indicate BVSP mode

Simulations

- Proposal implemented into HTM 5.1
- Common test conditions followed

- Tested configurations
  - Different precisions
    - 1x1 pixel level
    - 2x2 sub-PU level
    - 4x4 sub-PU level

Simulation results – 1x1

	video 0	video 1	video 2	video only	synthesized only	coded & synthesized	enc time	dec time	ren time
Balloons	0.0%	-0.9%	-0.5%	-0.2%	-0.3%	-0.3%	101.7%	100.8%	100.2%
Kendo	0.0%	-1.6%	-1.8%	-0.5%	-0.6%	-0.5%	103.3%	99.5%	101.7%
Newspapercc	0.0%	-0.2%	-0.7%	-0.1%	-0.2%	-0.2%	103.7%	103.3%	101.1%
GhostTownFly	0.0%	-9.0%	-8.7%	-2.4%	-1.7%	-1.9%	101.2%	107.2%	102.2%
PoznanHall2	0.0%	-0.4%	-2.8%	-0.6%	-0.7%	-0.6%	98.5%	104.9%	101.8%
PoznanStreet	0.0%	-2.9%	-3.0%	-0.9%	-0.7%	-0.7%	102.0%	100.5%	105.6%
UndoDancer	0.0%	-12.4%	-10.8%	-3.4%	-2.2%	-2.5%	98.5%	106.8%	107.2%
1024x768	0.0%	-0.9%	-1.0%	-0.3%	-0.4%	-0.3%	102.9%	101.2%	101.0%
1920x1088	0.0%	-6.2%	-6.3%	-1.8%	-1.3%	-1.4%	100.0%	104.8%	104.2%
average	**0.0%**	**-3.9%**	**-4.1%**	**-1.2%**	**-0.9%**	**-1.0%**	**101.2%**	**103.3%**	**102.8%**

1.2%: video PSNR vs. video bitrate
0.9%: synthesis PSNR vs. total bitrate
1.0%: coded & synthesis PSNR vs. total bitrate

3%: decoding time increase

Simulation results – 2x2

	video 0	video 1	video 2	video only	synthesized only	coded & synthesized	enc time	dec time	ren time
Balloons	0.0%	-0.9%	-0.6%	-0.2%	-0.2%	-0.2%	101.0%	102.0%	101.4%
Kendo	0.0%	-1.6%	-1.7%	-0.5%	-0.6%	-0.5%	103.0%	101.5%	103.7%
Newspapercc	0.0%	-0.2%	-0.6%	-0.1%	-0.1%	-0.1%	105.1%	102.0%	105.1%
GhostTownFly	0.0%	-8.6%	-8.8%	-2.3%	-1.7%	-1.8%	102.5%	108.0%	105.5%
PoznanHall2	0.0%	-0.3%	-2.5%	-0.5%	-0.6%	-0.5%	99.6%	102.6%	102.9%
PoznanStreet	0.0%	-2.8%	-2.8%	-0.9%	-0.7%	-0.7%	103.2%	103.4%	101.2%
UndoDancer	0.0%	-10.9%	-9.7%	-3.0%	-2.3%	-2.5%	102.2%	105.1%	105.9%
1024x768	0.0%	-0.9%	-0.9%	-0.3%	-0.3%	-0.3%	103.0%	101.9%	103.4%
1920x1088	0.0%	-5.6%	-5.9%	-1.7%	-1.3%	-1.4%	101.8%	104.7%	103.9%
average	**0.0%**	**-3.6%**	**-3.8%**	**-1.1%**	**-0.9%**	**-0.9%**	**102.3%**	**103.5%**	**103.6%**

1.1%: video PSNR vs. video bitrate   Drop by 0.1%

0.9%: synthesis PSNR vs. total bitrate

0.9%: coded & synthesis PSNR vs. total bitrate

4%: decoding time increase

Simulation results – 4x4

	video 0	video 1	video 2	video only	synthesized only	coded & synthesized	enc time	dec time	ren time
Balloons	0.0%	-0.8%	-0.6%	-0.2%	-0.3%	-0.2%	102.1%	102.4%	100.2%
Kendo	0.0%	-1.4%	-1.7%	-0.5%	-0.6%	-0.5%	103.0%	102.0%	101.7%
Newspapercc	0.0%	-0.1%	-0.5%	0.0%	-0.2%	-0.1%	105.1%	102.0%	101.2%
GhostTownFly	0.0%	-8.0%	-8.1%	-2.1%	-1.6%	-1.7%	97.7%	108.1%	99.2%
PoznanHall2	0.0%	-0.1%	-2.4%	-0.5%	-0.5%	-0.5%	98.2%	102.4%	99.2%
PoznanStreet	0.0%	-2.4%	-2.8%	-0.8%	-0.6%	-0.7%	102.9%	102.9%	102.0%
UndoDancer	0.0%	-8.2%	-7.0%	-2.2%	-1.6%	-1.7%	100.5%	104.3%	102.8%
1024x768	0.0%	-0.8%	-0.9%	-0.2%	-0.3%	-0.3%	103.4%	102.2%	101.0%
1920x1088	0.0%	-4.7%	-5.0%	-1.4%	-1.1%	-1.1%	99.8%	104.4%	100.8%
average	**0.0%**	**-3.0%**	**-3.3%**	**-0.9%**	**-0.8%**	**-0.8%**	**101.3%**	**103.4%**	**100.9%**

0.9%: video PSNR vs. video bitrate   Drop by 0.2%
0.8%: synthesis PSNR vs. total bitrate
0.8%: coded & synthesis PSNR vs. total bitrate

3%: decoding time increase

Highlights on the achieved results

- Consistent individual coding gains
  - All sequences
  - Both coded and synthesis
  - Levels from pixel level to 2x2, 4x4 sub-PU level
- Overall gain is non-negligible
  - 1.0% for coded & synthesis vs. total bitrate
- High potential for sequences with high quality depth
  - Up to 12% for dependent view
  - Up to 3.4% including base view
- Additional complexity is limited
  - Decoding time increase is minimal, 3~4%
  - Percentage of pixels using BVSP mode, averagely, 3%

Recommendations

- Adopt the proposed BVSP method into HTM
  - Common test conditions
- Add a BVSP compensation function
  - 1x1 pixel level
  - Sub-PU level, 2x2 and 4x4
- Update NBDV function
  - PU level
  - On VSP neighbors and neighbors with other modes
- Add a new merge candidate
  - Indicating the use of BVSP mode
  - Placed right after the spatial neighboring blocks

Discussions

- BVSP precisions
  - Indicator using high level syntax
  - bvsp_precision
    - 0: pixel level
    - 1: 2x2 level
    - 2: 4x4 level
- Test model description to be available, *B1005-based*
- Depth first coding order can be easily supported
- Similar proposals in parts
  - NTT (JCT3V-C0100)
  - MediaTek (JCT3V-C0131)

THANKS

Especially to

Nokia

& **MediaTek**

for their valuable time in cross checking