

TE1: RWTH partner report on DMVD

Mathias Wien and Steffen Kamp

**Institut für Nachrichtentechnik
RWTH Aachen University**

- Introduction
- Overview on DMVD
 - Prediction
 - Candidates
 - Sup-pixel refinement
- Simulation results
 - Rate-distortion measurements
 - Complexity Assessment
- Conclusions

Introduction

- RWTH input to TE1 on
Decoder Side Motion Vector Derivation (DMVD)
- RWTH implementation of DMVD selected as the common platform
for TE1 on DMVD
- In SVN available to TE1 partners since May 5th, 2010
now revision 26
 - 12 partner logins provided
 - Branches created in TE:
 - Huawei
 - Intel
 - JVC
 - Mediatek

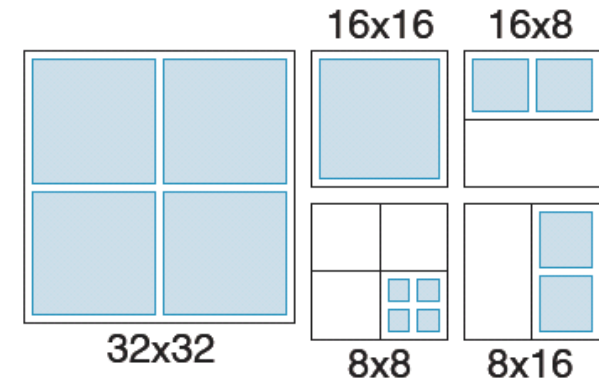
Decoder-side Motion Vector Derivation (DVMD)

- Decoder-side Motion Vector Derivation (DMVD) adds implicit motion vector coding in addition to explicit coding.
- DMVD prediction is available for:
 - P_L0_32x32, B_X_32x32,
 - P_L0_16x16, B_X_16x16,
 - P_L0_L0_16x8, B_X_X_16x8,
 - P_L0_L0_8x16, B_X_X_8x16,
 - P_L0_8x8
- One macroblock layer flag per partition: If `dmvd_flag[mbPartIdx]` is equal to 1, `mvd_IX` and `ref_idx_IX` are not present in the bitstream for this partition, but are derived during the decoding process.
- During bitstream parsing, `mvd_IX` and `ref_idx_IX` are assumed to be zero (for error resilient CABAC context derivation).
- During decoding, derived values are used for motion vector prediction (MVP) of subsequent (explicitly coded) partitions.

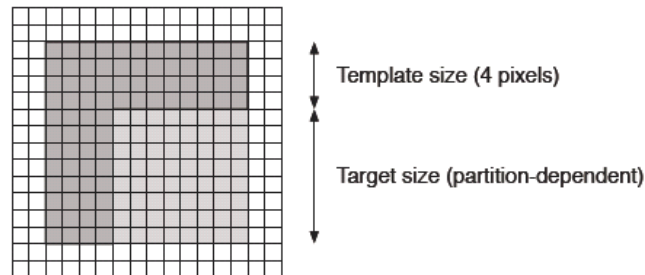
DMVD Prediction

- DMVD partitions are predicted using one or more targets in zig-zag scan order

Partition size	32x32	16x16	16x8	8x16	8x8
Number of targets	4	1	2	2	4
Target size	16x16	16x16	8x8	8x8	4x4



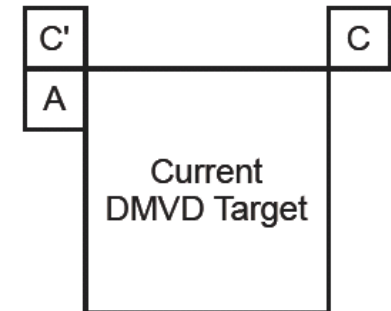
- For each target, a set of motion candidates is derived and the cost of each candidate is determined by template matching.



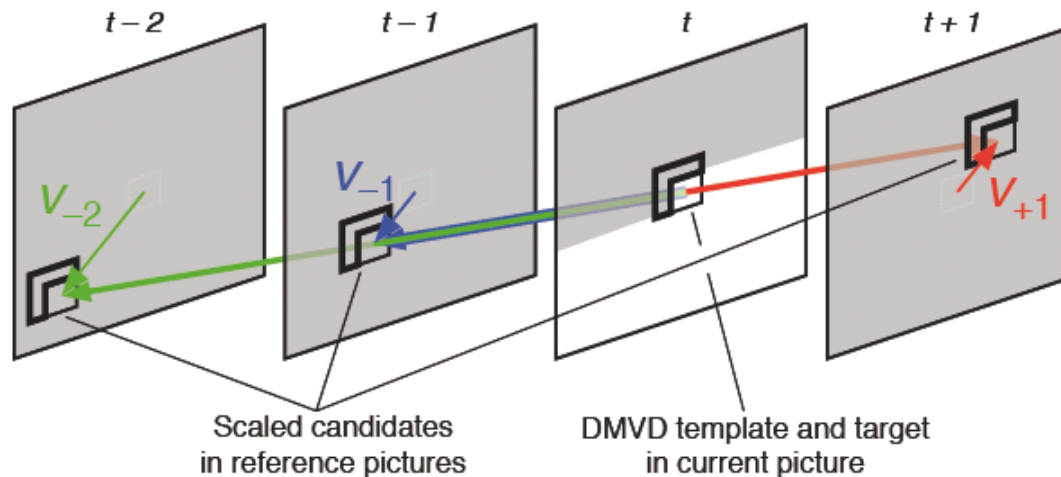
- The final prediction for each target is obtained by averaging the prediction signals of the 2 candidates with lowest cost (for both, uni- and bidirectional prediction)

- Template matching candidates are taken from the causal neighborhood of the current partition

- Motion from A and C
(if C is not available, C' is used instead)
- Max. candidates: 2 (P slice), 4 (B slice)

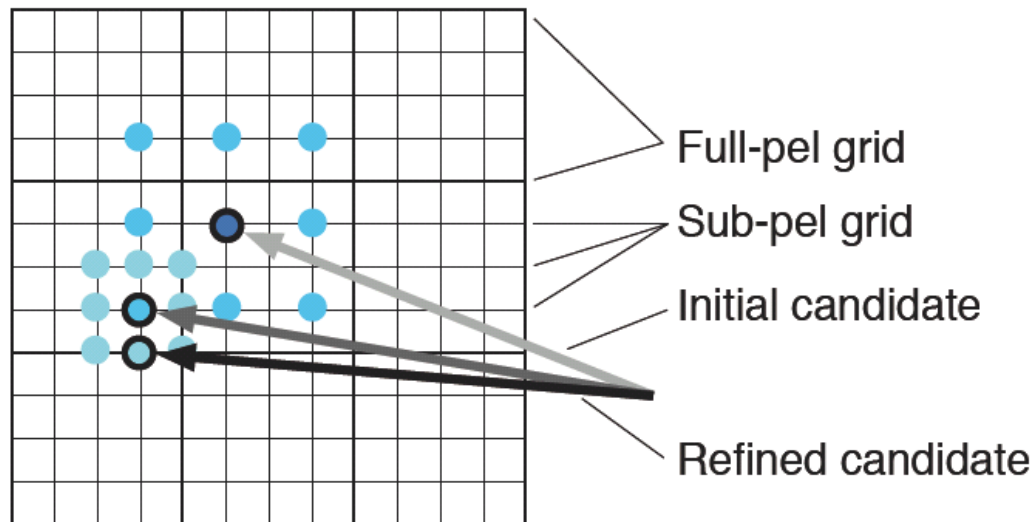


- Motion candidates are scaled to the reference picture based on temporal distances between the pictures.



DMVD Sub-pixel refinement

- A sub-pel refinement is performed on the 2 best candidates per reference picture:
 - Template matching for the 8 surrounding positions in horizontal and/or vertical half-pel distance,
 - followed by 8 positions in horizontal and/or vertical quarter-pel distance.



DMVD Complexity

- Max. count of DMVD cost calculations per target region: 144 (max. 36 per reference picture)
- Typical count for CS1: ~100, CS2: ~120
- If sub-pel refinement is switched off, max. count reduces to 16 (4 per reference picture)
- Estimate of required SAD calculations (difference between 2 samples and addition to sum):

Target Size	Target Pixels (K)	Template Pixels (L)	Max. SAD calculations per predicted pixel	
			Proposal $144 \cdot L / K$	no sub-pel refinement $16 \cdot L / K$
4x4	16	48	432	48
8x8	64	80	180	20
16x16	256	144	81	9

Simulation results

- Simulation results according to TE1 common simulation conditions
 - Constraint Set 1 and 2 from CfP
 - 100 frames per sequence (first frames for all but Kimono, where the last 100 frames were taken)
 - Fixed QP settings:
 - $QPI = 27, 30, 34, 38$
 - $QPP = QPI + 1, QPB = QPI + 2$
 - TE1 Anchor config: Single pass encoding, no WP, no RDOQ
ExtMB=1, BiPredMESubPel=1, UseAdaptiveFilter=0, MVC=0
- Simulation results provided for
 - Anchor conditions (with and w/o DMVD)
 - ExtMB=0
 - MVC
 - MVC+ExtMB=0

Summary RD results for Constraint Set 1

CS1

default

	BD Rate [%]	BD PSNR [dB]
Average Class A	-3,71	0,16
Average Class B	-4,34	0,14
Average Class C	-2,93	0,12
Average total	-4,11	0,14

ExtMB=0

	BD Rate [%]	BD PSNR [dB]
Average Class A	-6,29	0,29
Average Class B	-7,69	0,27
Average Class C	-5,05	0,21
Average total	-7,24	0,23

Sequence	BD Rate [%]	BD PSNR [dB]
BQMall_wvga	-3,73	0,16
BasketballDrill_wvga	-3,31	0,13
PartyScene_wvga	-2,28	0,10
RaceHorses_wvga	-2,40	0,09
PeopleOnStreet_2kcrop	-3,19	0,15
Traffic_2kcrop	-4,22	0,17
BQTerrace_1080p	-4,93	0,11
BasketballDrive_1080p	-5,61	0,16
Cactus_1080p	-3,75	0,12
Kimono_1080p	-6,86	0,23
ParkScene_1080p	-4,96	0,17

Sequence	BD Rate [%]	BD PSNR [dB]
BQMall_wvga	-5,96	0,27
BasketballDrill_wvga	-5,75	0,23
PartyScene_wvga	-3,89	0,17
RaceHorses_wvga	-4,59	0,19
PeopleOnStreet_2kcrop	-5,12	0,25
Traffic_2kcrop	-7,47	0,33
BQTerrace_1080p	-9,42	0,24
BasketballDrive_1080p	-9,95	0,33
Cactus_1080p	-6,47	0,22
Kimono_1080p	-11,63	0,43
ParkScene_1080p	-9,40	0,36

Summary RD results for Constraint Set 1

MVC=1

	BD Rate [%]	BD PSNR [dB]
Average Class A	-3,05	0,13
Average Class B	-3,69	0,12
Average Class C	-2,78	0,11
Average total	-3,49	0,12

ExtMB=0 MVC=1

	BD Rate [%]	BD PSNR [dB]
Average Class A	-4,58	0,21
Average Class B	-6,03	0,21
Average Class C	-4,56	0,19
Average total	-5,69	0,21

Sequence	BD Rate [%]	BD PSNR [dB]
BQMall_wvga	-3,27	0,14
BasketballDrill_wvga	-3,42	0,13
PartyScene_wvga	-2,20	0,09
RaceHorses_wvga	-2,24	0,09
PeopleOnStreet_2kcrop	-2,93	0,14
Traffic_2kcrop	-3,18	0,13
BQTerrace_1080p	-4,09	0,09
BasketballDrive_1080p	-5,09	0,15
Cactus_1080p	-3,15	0,10
Kimono_1080p	-4,88	0,16
ParkScene_1080p	-3,89	0,13

Sequence	BD Rate [%]	BD PSNR [dB]
BQMall_wvga	-5,14	0,23
BasketballDrill_wvga	-5,43	0,22
PartyScene_wvga	-3,40	0,15
RaceHorses_wvga	-4,26	0,18
PeopleOnStreet_2kcrop	-4,39	0,21
Traffic_2kcrop	-4,78	0,20
BQTerrace_1080p	-7,20	0,18
BasketballDrive_1080p	-8,90	0,29
Cactus_1080p	-4,87	0,16
Kimono_1080p	-7,32	0,26
ParkScene_1080p	-6,90	0,25

Summary RD results for Constraint Set 2

CS2

default

	BD Rate [%]	BD PSNR [dB]
Average Class B	-4,87	0,13
Average Class C	-3,25	0,13
Average total	-4,15	0,13

ExtMB=0

	BD Rate [%]	BD PSNR [dB]
Average Class B	-7,39	0,22
Average Class C	-3,76	0,15
Average total	-5,78	0,19

Sequence	BD Rate [%]	BD PSNR [dB]
BQMall_wvga	-2,87	0,12
BasketballDrill_wvga	-3,52	0,13
PartyScene_wvga	-3,57	0,15
RaceHorses_wvga	-3,06	0,13
BQTerrace_1080p	-12,65	0,29
BasketballDrive_1080p	-3,71	0,10
Cactus_1080p	-2,75	0,08
Kimono_1080p	-2,74	0,08
ParkScene_1080p	-2,53	0,08

Sequence	BD Rate [%]	BD PSNR [dB]
BQMall_wvga	-3,37	0,15
BasketballDrill_wvga	-4,22	0,16
PartyScene_wvga	-4,15	0,18
RaceHorses_wvga	-3,30	0,14
BQTerrace_1080p	-16,39	0,41
BasketballDrive_1080p	-5,09	0,15
Cactus_1080p	-3,20	0,10
Kimono_1080p	-7,31	0,26
ParkScene_1080p	-4,97	0,18

Summary RD results for Constraint Set 2

MVC=1

	BD Rate [%]	BD PSNR [dB]
Average Class B	-4,27	0,11
Average Class C	-3,17	0,13
Average total	-3,78	0,12

ExtMB=0 MVC=1

	BD Rate [%]	BD PSNR [dB]
Average Class B	-4,53	0,12
Average Class C	-3,36	0,14
Average total	-4,01	0,13

Sequence	BD Rate [%]	BD PSNR [dB]
BQMall_wvga	-2,67	0,11
BasketballDrill_wvga	-3,55	0,13
PartyScene_wvga	-3,49	0,14
RaceHorses_wvga	-2,97	0,12
BQTerrace_1080p	-11,51	0,25
BasketballDrive_1080p	-3,43	0,09
Cactus_1080p	-2,19	0,06
Kimono_1080p	-2,06	0,06
ParkScene_1080p	-2,14	0,07

Sequence	BD Rate [%]	BD PSNR [dB]
BQMall_wvga	-2,82	0,12
BasketballDrill_wvga	-3,70	0,14
PartyScene_wvga	-3,91	0,16
RaceHorses_wvga	-3,01	0,13
BQTerrace_1080p	-12,83	0,30
BasketballDrive_1080p	-3,08	0,09
Cactus_1080p	-2,31	0,07
Kimono_1080p	-2,01	0,06
ParkScene_1080p	-2,40	0,08

Complexity Assessment

- Data is collected from accounting of the Sun Grid Engine.
 - Memory: max vmem [in KB],
 - Simulation time: User time [s]

Default CS1

Average	Relative memory usage		User time	
	total	decoder	total	decoder
Class A	0,00%	0,01%	107,81%	238,43%
Class B	-0,05%	-0,48%	90,03%	192,65%
Class C	-0,07%	-1,29%	102,79%	242,20%

CS1 ExtMB=0

Average	Relative memory usage		User time	
	total	decoder	total	decoder
Class A	0,00%	-0,18%	106,78%	357,99%
Class B	-0,05%	-0,24%	105,25%	270,62%
Class C	-0,10%	-0,89%	85,70%	255,20%

CS1 MVC=1

Average	Relative memory usage		User time	
	total	decoder	total	decoder
Class A	0,00%	-0,06%	127,25%	213,65%
Class B	-0,03%	0,54%	114,33%	179,21%
Class C	0,11%	-1,22%	110,45%	229,59%

CS1 ExtMB=0 MVC=1

Average	Relative memory usage		User time	
	total	decoder	total	decoder
Class A	0,00%	0,06%	108,67%	305,91%
Class B	0,00%	-0,40%	99,19%	251,87%
Class C	-0,27%	-1,33%	91,37%	247,54%

Default CS2

Average	Relative memory usage		User time	
	total	decoder	total	decoder
Class B	0,00%	-0,17%	56,44%	399,58%
Class C	0,00%	-1,63%	49,87%	475,91%

CS2 ExtMB=0

Average	Relative memory usage		User time	
	total	decoder	total	decoder
Class B	0,00%	-0,28%	35,89%	485,02%
Class C	0,02%	-2,05%	39,29%	495,16%

CS2 MVC=1

Average	Relative memory usage		User time	
	total	decoder	total	decoder
Class B	0,00%	0,14%	42,59%	310,86%
Class C	0,00%	-1,22%	48,18%	448,31%

CS2 ExtMB=0 MVC=1

Average	Relative memory usage		User time	
	total	decoder	total	decoder
Class B	0,00%	-0,17%	26,80%	270,74%
Class C	0,01%	-1,73%	34,17%	430,38%

Conclusions

- Decoder-side motion vector derivation is proposed
- Good objective and subjective coding efficiency within KTA
- Flexible algorithm allowing a performance/complexity trade-off
- Complexity is considerably affected by KTA and should be evaluated within the new test model.
- Suggest the formation of an AhG on decoder-side parameter derivation in which the TE on DMVD and other activities are coordinated