Video coding technology proposal by RWTH Aachen University

JCTVC-A112

Steffen Kamp, Mathias Wien

Institut für Nachrichtentechnik
RWTH Aachen University

Proposal overview

- Based on KTA 2.6r1
- Proposed tool:
  Decoder-side Motion Vector Derivation

- Base configuration from VCEG-AJ10r1, including:

```
UseRDO_Q=1
UseNewOffset=1
```

- KTA-Tools that were set differently from AJ10r1:

```
UseExtMB=1
AdaptiveRounding=0
UseAdaptiveFilter=5
UseHPFilter=1
UseAdaptiveLoopFilter=1
UseIntraMDDT=1
MVCompetition=1
```

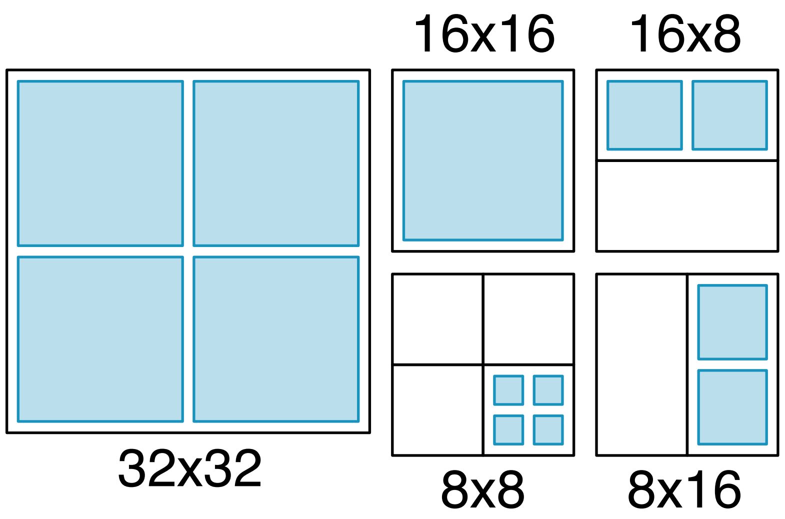
Decoder-side Motion Vector Derivation (DMVD)

- 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_lX and ref_idx_lX are not present in the bitstream for this partition, but are derived during the decoding process.
- During bitstream parsing, mvd_lX and ref_idx_lX 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.

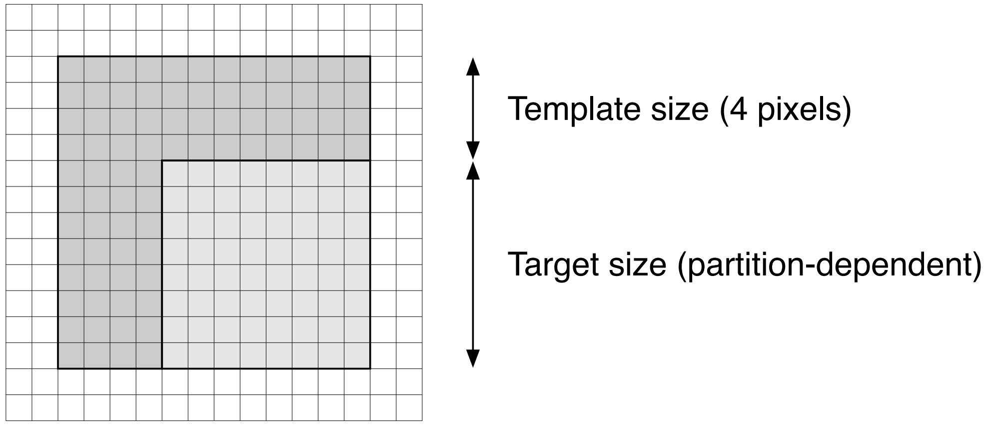
DMVD Prediction

- 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



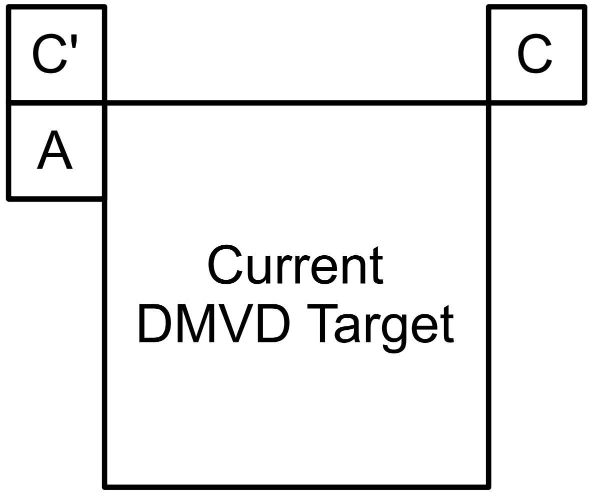
- For each target, a set of motion candidates is derived and the cost of each candidate is determined by template matching.



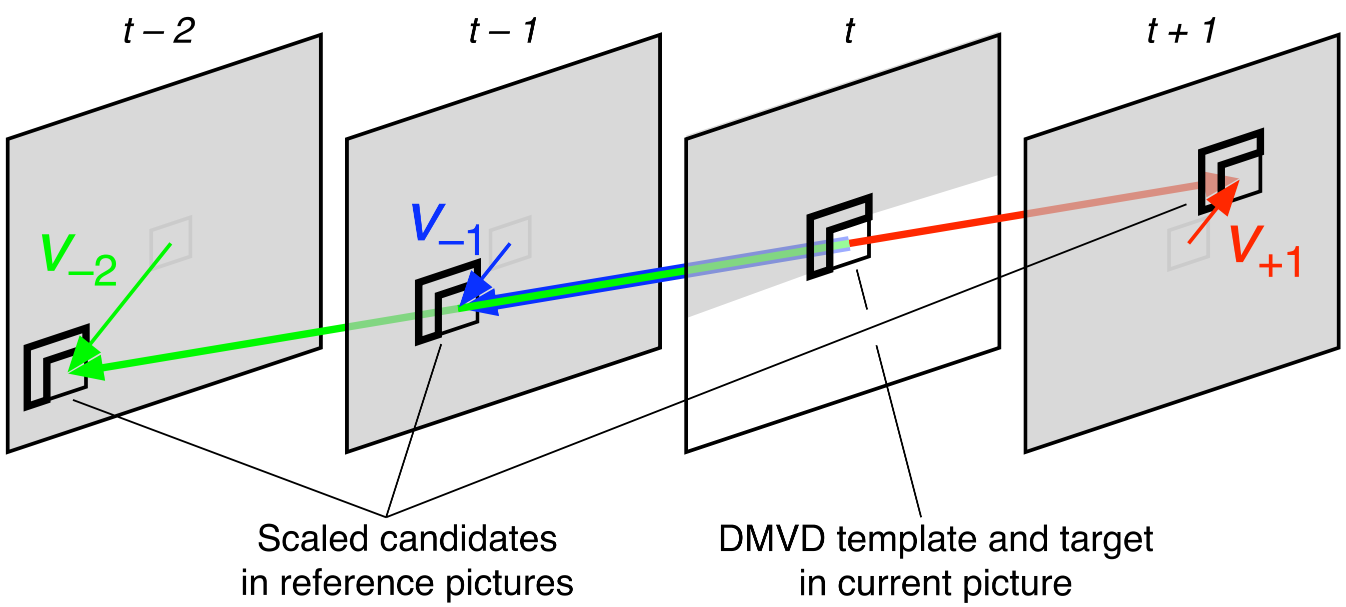
- The final prediction for each target is obtained by averaging the prediction signals of the 2 candidates with lowest cost (for both, uni- and bidirectional prediction)

DMVD Candidates

- 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)

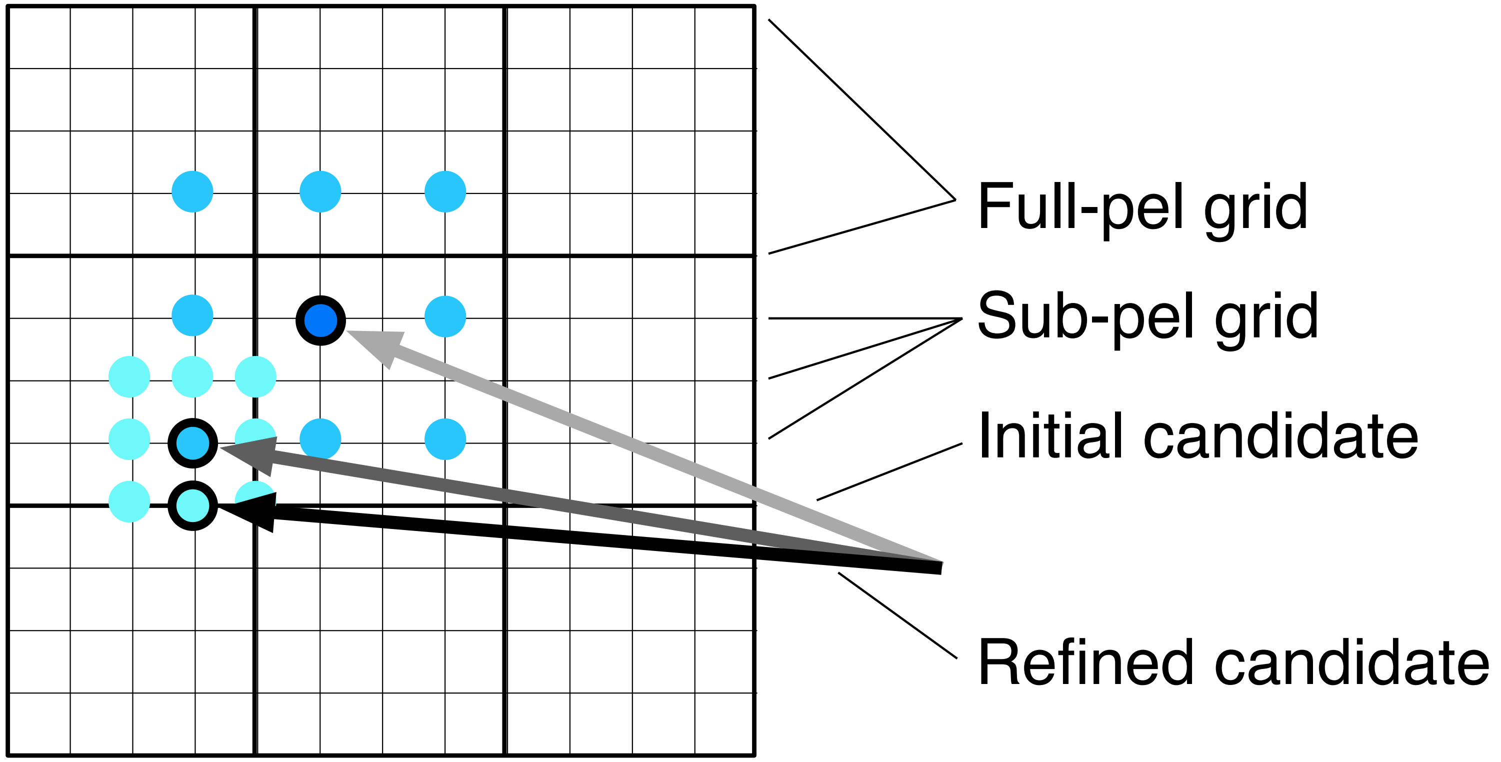


- Motion candidates are scaled to the reference picture based on temporal distances between the pictures.



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 • L / K	no sub-pel refinement 16 • L / K
4x4	16	48	432	48
8x8	64	80	180	20
16x16	256	144	81	9

Results for Constraint Set 1 (relative to Alpha anchor)

Sequence	BD-PSNR [dB]			BD-Bitrate [%]		
	Low	High	Overall	Low	High	Overall
Traffic_2kcrop	1.08	0.87	0.95	-25.34	-22.91	-23.84
PeopleOnStreet_2kcrop	1.17	1.02	1.07	-20.26	-18.32	-19.12
Kimono_1080p	1.40	1.14	1.28	-32.82	-30.65	-32.03
ParkScene_1080p	0.86	0.68	0.78	-21.06	-17.36	-19.45
Cactus_1080p	0.72	0.59	0.67	-20.63	-19.87	-20.48
BasketballDrive_1080p	1.10	0.87	1.00	-28.96	-27.04	-28.25
BQTerrace_1080p	0.62	0.54	0.59	-28.22	-30.67	-29.28
BasketballDrill_wvga	1.08	0.98	1.01	-23.75	-22.14	-22.58
BQMall_wvga	1.22	1.03	1.11	-22.67	-20.52	-21.38
PartyScene_wvga	0.74	0.74	0.74	-18.71	-18.13	-18.42
RaceHorses_wvga	1.27	1.16	1.21	-27.46	-25.40	-26.08
BasketballPass_wqvga	0.94	0.86	0.89	-18.62	-15.70	-16.76
BQSquare_wqvga	0.94	0.91	0.91	-23.91	-22.69	-22.96
BlowingBubbles_wqvga	0.46	0.57	0.53	-10.84	-12.65	-12.05
RaceHorses_wqvga	0.87	0.83	0.85	-16.78	-14.87	-15.66
average	0.96	0.85	0.91	-22.67	-21.26	-21.89

Results for Constraint Set 2 (relative to Beta anchor)

Sequence	BD-PSNR [dB]			BD-Bitrate [%]		
	Low	High	Overall	Low	High	Overall
Traffic_2kcrop	1.63	1.41	1.53	-35.87	-34.12	-35.17
PeopleOnStreet_2kcrop	0.48	0.36	0.44	-12.51	-9.72	-11.60
Kimono_1080p	0.55	0.46	0.52	-16.06	-15.15	-15.82
ParkScene_1080p	1.31	1.07	1.20	-32.31	-30.53	-31.54
Cactus_1080p	0.82	0.72	0.79	-33.30	-34.24	-34.04
BasketballDrive_1080p	0.72	0.68	0.70	-16.93	-16.61	-16.70
BQTerrace_1080p	1.14	0.92	1.01	-21.66	-18.60	-19.73
BasketballDrill_wvga	0.25	0.12	0.18	-6.59	-3.22	-4.54
BQMall_wvga	0.88	0.85	0.88	-20.84	-19.55	-20.26
PartyScene_wvga	0.65	0.64	0.64	-13.58	-12.24	-12.65
RaceHorses_wvga	0.09	-0.01	0.04	-2.46	0.44	-1.03
BasketballPass_wqvga	-0.23	-0.25	-0.23	6.16	6.34	6.02
BQSquare_wqvga	0.57	0.60	0.60	-11.63	-10.99	-11.45
BlowingBubbles_wqvga	1.43	1.22	1.30	-28.02	-28.85	-28.66
RaceHorses_wqvga	0.94	0.82	0.86	-19.02	-19.78	-19.49
average	1.28	0.97	1.10	-26.61	-25.80	-26.26

Results for Constraint Set 2 (relative to Gamma anchor)

Sequence	BD-PSNR [dB]			BD-Bitrate [%]		
	Low	High	Overall	Low	High	Overall
Traffic_2kcrop	2.62	2.38	2.50	-50.93	-50.03	-50.59
PeopleOnStreet_2kcrop	1.54	1.33	1.45	-34.88	-30.80	-33.04
Kimono_1080p	1.56	1.38	1.47	-38.87	-36.96	-37.98
ParkScene_1080p	2.25	1.88	2.09	-47.27	-45.79	-46.60
Cactus_1080p	2.08	1.68	1.91	-58.71	-55.81	-57.34
BasketballDrive_1080p	1.89	1.86	1.87	-38.60	-39.10	-38.93
BQTerrace_1080p	2.28	2.02	2.12	-38.77	-36.05	-37.08
BasketballDrill_wvga	1.42	1.42	1.43	-36.53	-33.80	-34.74
BQMall_wvga	1.36	1.34	1.35	-30.77	-28.90	-29.70
PartyScene_wvga	1.33	1.41	1.37	-26.32	-25.43	-25.69
RaceHorses_wvga	2.08	2.08	2.08	-50.55	-47.26	-48.47
BasketballPass_wqvga	1.10	1.15	1.12	-24.97	-24.53	-24.66
BQSquare_wqvga	1.04	1.06	1.05	-20.23	-18.33	-19.11
BlowingBubbles_wqvga	2.64	2.27	2.44	-44.69	-46.11	-45.52
RaceHorses_wqvga	2.26	1.97	2.08	-39.36	-39.27	-39.17
average	2.55	2.14	2.33	-45.53	-47.12	-46.43

Conclusion

- A 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.