

JVET-U0088: VVC software decoder implementation for mobile devices

J. Chen, L. Wang, R.-L. Liao, Y. Ye

Alibaba

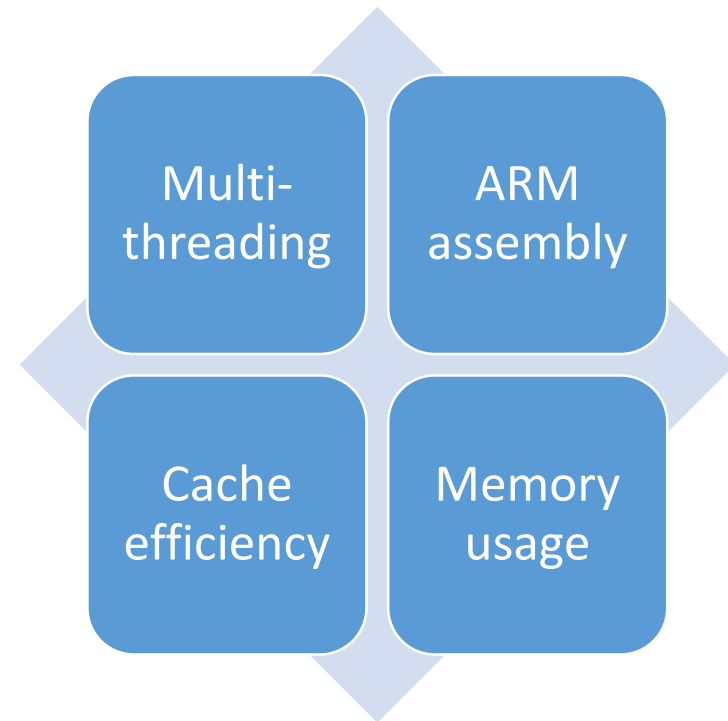
Background

- Early implementations of optimized VVC software decoder were presented at previous meetings
 - JVET-S0224, JVET-T0095, and JVET-T0099 all achieved real-time 4K decoding on desktop platforms
 - JVET-Q0386 presented a real-time 1080p decoder for mobile devices using 4 tiles to fit 4-core ARM architecture
- Mobile devices are the most important device category for e-commerce
- This contribution presents Ali266, a VVC software decoder implementation specifically optimized for mobile

Ali266 overview

- Designed with stability and maximal decoding speed in mind
- All functional modules **written from scratch based on the VVC spec**
 - bitstream parsing, CABAC decoding, inv. quantization and inv. transform, intra pred, inter pred, loop filtering, etc
- Thin decoder: binary size < 1MB

Four main aspects of Ali266 optimization



Ali266 optimizations

Multi-threaded decoding:

- picture level, CTU level, within a CTU
- Parallelization efficiency greater than 90%
- General purpose, **no need for tiles or WPP to be enabled**

ARM assembly

- Neon SIMD instructions to accelerate computation intensive modules
- Intra prediction, motion compensated prediction, inv. transform, SAO, deblocking filter, etc.
- Acceleration factor: ~ **2.5x**

Cache efficiency

- Data structure and memory storage strategy designed to maximize cache hit & avoid cache miss
- Minimizing the code size to store in cache entirely
- Data prefetching to avoid idling, and so on

Memory usage

- Need to cover a wide spectrum of mobile devices, including those with quite limited memory based on today's standard
- Strategies such as efficient reference picture management, memory sharing among different threads etc.
- **Small memory footprint: 30 MB to decode 720p**

Performance tests

- Two types of content:
 - JVET CTC content, classes A to D, QP = {22, 27, 32 37}
 - 20 Taobao e-commerce content, 720p resolution, bit rates {0.6, 1.0, 1.5} (Mbps)
- Bitstreams generated using VTM-11.0:
 - Random access, GOP size = 32
 - Internal bit depth = 8 bits
 - ALF and CCALF disabled
- Two categories of tests
 - Speed test: wide variety of both Android and iOS devices
 - Robustness test: 11,000+ corrupted bitstreams

Decoding speed: JVET CTC content

	VTM-11.0 *	Ali266							
		Huawei P40				iPhone 12 Pro Max			
# threads	1	1	2	4	8	1	2	4	6
Class A	2	8	15	27	32	18	31	42	50
Class B	9	31	60	110	128	77	140	179	213
Class C	38	136	255	456	503	313	577	683	761
Class D	137	446	797	1360	1210	966	1617	1726	2146

* VTM-11.0 speed was tested on Huawei P40

- Single thread: 3.5x speedup compared to VTM-11.0
- Maximum threads: the average 4K decoding speed exceeds 30fps

Decoding speed: 4K content

		Bitrate (Mbps)	VTM11	Huawei P40				iPhone 12 Pro Max			
				1	2	4	8	1	2	4	6
Class A1 4K	Tango2	16.62	1.97	7.08	13.73	24.85	28.83	17.17	28.73	37.94	44.50
		6.75	2.34	8.19	15.72	28.58	33.56	20.30	34.37	45.67	54.11
		3.45	2.55	8.89	17.01	30.87	36.00	21.91	37.39	49.57	59.49
		1.96	2.64	9.26	17.54	32.26	37.61	22.70	39.31	51.70	61.68
	FoodMarket4	14.66	2.02	7.10	13.78	24.83	29.16	17.54	29.26	39.54	45.70
		7.17	2.18	7.69	14.79	26.56	31.61	19.11	31.96	42.23	50.49
		3.68	2.34	8.15	15.60	28.36	33.43	20.23	34.21	46.14	53.97
		2.00	2.42	8.43	16.14	29.38	34.57	20.80	35.16	47.28	57.01
	Campfire	42.26	1.68	5.23	10.19	18.31	22.08	10.30	17.36	23.87	27.56
		12.36	2.35	8.43	16.35	29.30	33.69	18.56	31.65	41.40	46.61
		6.49	2.75	10.04	19.22	34.26	39.69	22.73	38.64	49.85	58.40
		3.46	3.02	11.00	20.97	37.64	43.65	25.43	43.98	56.29	65.53
Class A2 4K	CatRobot	17.46	1.99	6.96	13.46	24.40	28.10	16.85	28.65	37.86	44.66
		7.85	2.23	7.79	15.04	27.40	32.36	19.37	33.03	43.88	52.37
		3.92	2.39	8.25	15.79	29.13	34.02	20.66	35.28	47.55	57.18
		2.14	2.50	8.56	16.36	30.32	35.37	21.47	37.02	49.17	59.61
	DaylightRoad2	22.75	1.85	6.38	12.36	22.32	26.21	15.15	25.58	33.31	39.15
		7.84	2.16	7.59	14.62	26.66	31.64	18.85	32.17	43.16	51.73
		3.77	2.38	8.26	15.83	29.05	34.37	20.56	35.49	47.06	57.29
		1.99	2.50	8.66	16.52	30.50	35.95	21.41	37.20	49.46	60.20
	ParkRunning3	96.66	1.26	3.86	7.53	13.60	16.59	7.88	13.07	18.51	21.93
		37.96	1.51	5.08	9.91	18.01	21.66	11.32	18.85	26.12	31.33
		16.46	1.79	6.16	11.91	21.65	26.02	14.33	24.11	32.78	39.91
		7.43	2.06	7.19	13.81	25.26	30.04	17.01	28.83	39.11	47.78

- 4K content: 30fps decoding speed on both platforms up to 7 Mbps
- Worst case 4K decoding (ParkRunning, 96.66 Mbps @ QP 22): 17 fps on Huawei P40 and 22 fps on iPhone 12 Pro Max
- 1080p content: real-time decoding with only 2 threads on iOS and at most 4 threads on Android

Decoding speed: Taobao e-commerce content

- Content characteristics:
 - Thirteen portrait sequences and seven landscape sequences,
 - Selection matches the real-world popularity of portrait-mode mobile video
- 3 Android phones with wide-ranging hardware capabilities
- Observations:
 - ***All models achieve real-time decoding, even most affordable model needs only 2 threads***
 - Using fewer threads reduces CPU usage and lowers power consumption
- ***Software VVC decoder is suitable for deployment in the near future for e-commerce***

		Latest device model				Device a few years old				Affordable device a few years old			
# threads		1	2	4	8	1	2	4	8	1	2	4	8
Portrait (720x1280)	0.6 Mbps	109	201	355	386	49	90	154	193	21	40	74	123
	1.0 Mbps	100	184	323	351	45	83	140	177	20	38	69	112
	1.5 Mbps	88	163	287	312	40	73	123	159	18	34	62	101
Landscape (1280x720)	0.6 Mbps	83	155	279	308	39	72	123	164	17	32	60	105
	1.0 Mbps	78	146	262	288	36	67	116	152	16	31	57	98
	1.5 Mbps	72	136	243	268	34	63	108	143	15	29	54	91
Average		92	170	301	328	42	77	131	168	18	35	64	107

Robustness of Ali266

- 11,000+ erroneous / corrupted bitstreams:
 - Random packet losses
 - Random errors
 - Redundant and repeated coded NAL packets
 - Coded NAL packets out of order
- errors encountered at or above slice level: seek forward and resume decoding at the next intra refresh point
- errors encountered below slice level: terminate parsing but attempt to decode the available portion of coded data
- On-going work on more error concealment and recovery mechanisms

Summary

- Ali266 is a ***highly optimized VVC decoder implementation*** for mobile devices
 - Thin decoder, binary < 1MB, 720p decoding memory ~ 30MB
 - Single thread mode 3.5x faster than VTM-11.0
 - Multi-thread 30fps speed for 4K up to 7Mbps
 - Real-time decoding of e-commerce content with at most 2 threads on most affordable model
 - Proven robustness with 11,000+ corrupted bitstream
- ***Ready for real-world deployment in the near future***
- Future work includes completing optimized implementation of ALF and CCALF