Fast Multi-Rate Encoding for Adaptive HTTP Streaming

Hadi Amirpour [1], Ekrem Çetinkaya [1], Christian Timmerer [1,2], and Mohammad Ghanbari [3,4]

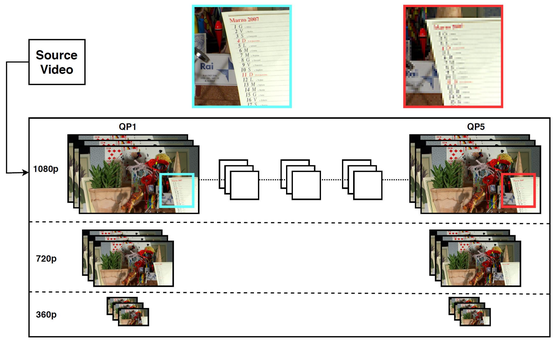
[1] Alpen-Adria-Universität Klagenfurt, Klagenfurt, Austria
[2] Bitmovin, Klagenfurt, Austria
[3] School of Electrical and Computer Engineering, University of Tehran, Tehran, Iran
[4] School of Computer Science and Electronic Engineering, University of Essex, UK

ATHENA Christian Doppler (CD) Laboratory

Introduction

MPEG's Dynamic Adaptive Streaming over HTTP (DASH) is a standardized solution in place enabling **HTTP adaptive streaming (HAS).**

In HAS, videos are divided into segments and each segment is encoded at different bit-rates and resolutions referred as **representations**.



The compression of multiple representations requires a highly efficient video encoding ecosystem.

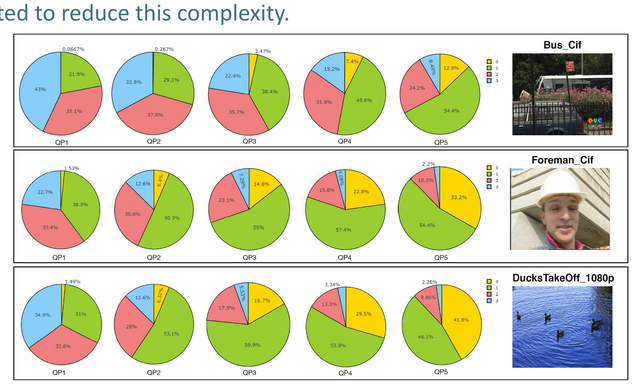
High Efficiency Video Coding (HEVC) allows for a high encoding efficiency by using sophisticated tools at the cost of increasing encoding time-complexity.

In HEVC, each frame is divided into tiles or slices which are further divided into **Coding Tree Units (CTUs)** with a predetermined size of up to 64 x 64 pixels.

Each CTU then can be divided recursively into multiple equal-sized square regions called **coding units (CUs)**. In inter coding mode, multiple **reference frames** are used for motion estimation.

Finding optimal CTU structure and choosing correct reference frame using exhaustive search takes the largest amount of time in the encoding process.

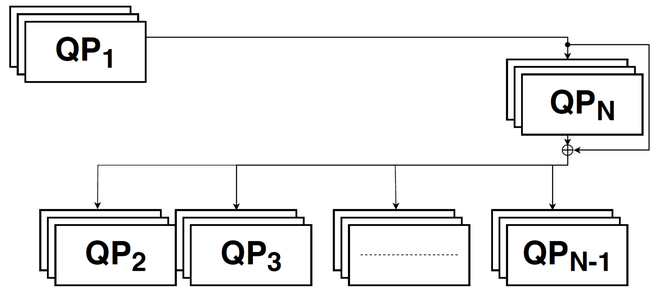
Redundancy between co-located CTUs in different representations can be exploited to reduce this complexity.



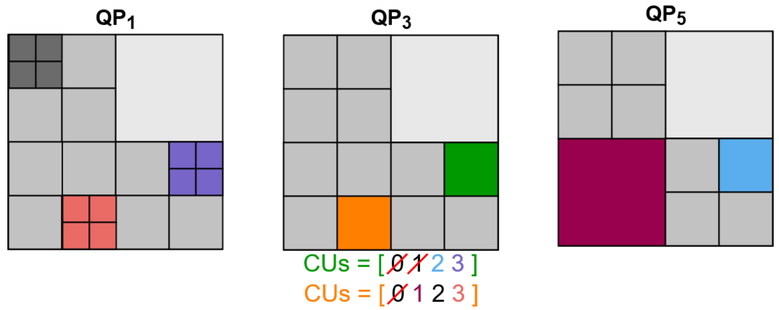
Proposed Method

Existing methods [1] use depth of CTUs in the highest quality representation as a bound to eliminate searching larger depths than the co-located CTU in the highest quality representation.

In this paper, information of co-located encoded CTUs in both the **highest and the lowest** quality representations are used to limit the search process of each CTU.



CTU Depth Search: **Highest depth level in the lowest quality representation** and the **lowest depth level in the highest quality representation** forms the CU depth level bounds for searching.



Reference Frame Selection: If both the highest and the lowest quality chose the same reference frame, that reference frame is used in intermediate representations and further searches are skipped.

Parallel Encoding: Can be started with two CTU delay (QP_1 and QP_N)

Encode the highest quality using unmodified HEVC

Use CTU information to encode the lowest quality

Use information from both the highest and the lowest quality to encode intermediate representations

Results

Seven standard test video sequences are used at different resolutions (1080p and 360p):

- CrowdRun (50 fps), ParkJoy (50 fps), Kimono (24 fps), BlueSky (25 fps), RiverBed (25 fps), RushHour (25 fps), and Sunflower (25 fps)

Results are compared against the reference HEVC software (HM 16.20) and the state-of-the-art method [1].

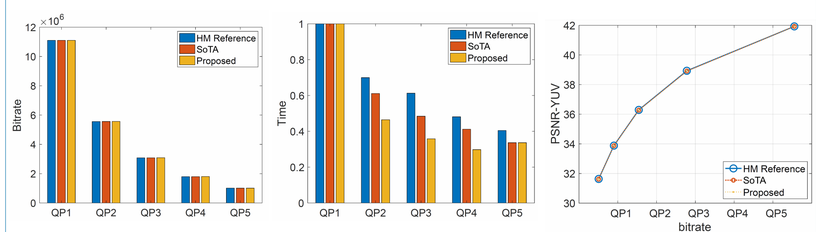
Bjontegaard delta PSNR (BD-PSNR) and Bjontegaard delta rate (BD-rate) [2] are measured as metrics using PSNR-YUV as the objective metric.

Five QP values are used:

- QP1=22, QP2=26, QP3=30, QP4=34, QP5=38

ΔT indicates the average time differences at five QPs. **BD-Rate / ΔT** is used as the final metric to measure the effect in both time and size complexity.

Sequence		State-of-the-art				Proposed Method			
		BD-Rate	BD-PSNR	ΔT	BD-Rate/ΔT	BD-Rate	BD-PSNR	ΔT	BD-Rate/ΔT
1920x1080	CrowdRun	0.27	-0.011	12.01 %	2.26	0.34	-0.015	28.31 %	**1.23**
	ParkJoy	0.49	-0.019	19.88 %	2.48	0.34	-0.013	30.95 %	**1.11**
	Kimono	0.73	-0.018	40.83 %	1.81	0.67	-0.016	47.77 %	**1.41**
	BlueSky	1.02	-0.042	37.29 %	2.73	-0.50	0.0021	46.51 %	**-1.09**
	RiverBed	0.33	-0.011	47.29 %	0.71	0.35	-0.012	50.26 %	**0.69**
	RushHour	1.03	-0.020	39.99 %	2.58	1.05	-0.021	45.42 %	**2.32**
	Sunflower	0.92	-0.032	49.95 %	1.86	0.47	-0.018	54.81 %	**0.86**
Average		**0.68**	**-0.021**	**35.32 %**	**2.06**	**0.53**	**-0.010**	**43.43 %**	**0.99**
480x360	CrowdRun	0.26	-0.013	9.36 %	2.85	0.51	-0.026	29.18 %	**1.76**
	ParkJoy	0.17	-0.008	15.06 %	**1.13**	0.46	-0.023	30.78 %	1.51
	Kimono	0.77	-0.031	15.65 %	4.97	0.78	-0.031	30.86 %	**2.53**
	BlueSky	0.13	-0.006	23.79 %	**0.55**	1.15	-0.058	35.51 %	3.24
	RiverBed	0.10	-0.003	19.66 %	**0.52**	0.37	-0.015	25.95 %	1.46
	RushHour	0.57	-0.024	20.88 %	2.75	0.77	-0.032	33.54 %	**2.32**
	Sunflower	0.78	-0.003	35.28 %	2.21	0.78	-0.033	42.91 %	**1.82**
Average		**0.39**	**-0.012**	**19.95 %**	**2.14**	**0.68**	**-0.031**	**32.67 %**	**2.09**
Total Average		**0.48**	**-0.017**	**27.63 %**	**2.10**	**0.53**	**-0.020**	**38.05 %**	**1.44**



Acknowledgement

This research has been supported in part by the **Christian Doppler Laboratory ATHENA** (https://athena.itec.aau.at/)

References

[1] D. Schroeder, P. Rehm, and E. Steinbach, ''Block structure reuse for multi-rate high efficiency video coding'' in *2015 IEEE International Conference on Image Processing (ICIP),* Sep. 2015, pp. 3972-3976.
[2] Gisle Bjontegaard, ''Calculation of average PSNR differences between RD-curves'' , VCEG-M33, 2001.