Implementation of Video Compression Based on Spatial-Temporal Resolution Adaptation

Dr. Y.L. AJAY KUMAR[1], Dr. M. VENKATA SREERAJ[2], Mr. N. NAVEEN KUMAR[3]

[1] Associate Professor, Department of ECE, Anantha Lakshmi Institute of Technology and Sciences, Anantapur, AP, India.

[2] Professor, Department of ECE , Anantha Lakshmi Institute of Technology and Sciences, Anantapur, AP, India.

[3] Assistant Professor, Department of ECE, Anantha Lakshmi Institute of Technology and Sciences, Anantapur, AP, India.

Abstract The ViSTRA machine dynamically resamples the enter video spatially and temporally at some stage in encoding, relying on a quantization-decision choice, and reconstructs the entire decision video on the decoder. Frame repetition is used for temporal up sampling, whilst a Convolutional Neural Network (CNN) super-decision version is used for spatial decision up sampling. The High Efficiency Video Coding (HEVC) fashionable software program now consists of ViSTRA (HM 16.14). With BD-charge will increase of 15 based on PSNR and a mean MOS distinction of 0.five primarily based totally on subjective visible best testing, experimental findings proven through an worldwide project monitor extensive improvements.

Keywords—Video compression, spatial resolution adaptation, temporal resolution adaptation, perceptual video compression, CNN-based super-resolution.

1. INTRODUCTION

Video content producers have been expanding the video parameter space by adopting greater spatial resolutions, frame rates, and dynamic ranges to meet the growing demand for more immersive visual experiences. This greatly increases the bitrate necessary to store and distribute video material, putting present bandwidth constraints to the test and necessitating more compression efficiency than current video codecs can provide. Previous research has revealed that the best video representation parameters in terms of perceived quality are significantly content dependent [1, 2].

Bitrates might be greatly decreased while keeping equal perceived visual quality by dynamically forecasting these characteristics. Several publications have advocated decreasing spatial resolution for low bitrate encoding in this context [3, 4], but no viable adaption mechanism exists. Others have created prediction models [5, 6] or implemented resolution adaptation as one of the rate-distortion optimised modes at the block level (CTU) [7], but only for H.264 or intra coding. A few frame rate selection approaches have been presented in [8, 9] for temporal adaptation.

However, video compression methods have not yet been fully incorporated. Furthermore, the video resampling technique used has a significant impact on the reconstructed video quality. For the reconstruction of full resolution video frames, previous spatial resolution adaption algorithms generally used linear filters, such as bicubic. However, due to better reconstruction quality, CNN-based superresolution approaches [10, 11] have been prominent in the field of computer vision in recent years. However, machine learning-based techniques for video compression have yet to be substantially investigated.

We propose ViSTRA, a spatio-temporal resolution adaptation framework for video compression that dynamically predicts the optimal spatial and temporal resolutions for the input video during encoding and attempts to reconstruct the full resolution video at the decoder, based on our previous work on quality assessment [1, 2, 12, 13] and spatial resolution adaptation for intra coding [14]. The integration of both spatial and temporal adaptation into a single framework is one of our paper's primary achievements.

• A Quantization-Resolution Optimization (QRO) module that uses perceptual quality measurements and machine learning approaches to make accurate resolution adaption decisions;

• The utilization of a CNN-based super goal model prepared especially for compacted material to remake full spatial goal content;

• The ViSTRA structure's association with HEVC reference programming (HM 16.14).

The discoveries gave here depend on test successions used in the IEEE ICIP 2017 Video Compression Grand Challenge [15]. When contrasted with the first HEVC anchor codec, they exhibit critical coding benefits of 14.5 percent BD-rate (PSNR) and 0.52 normal MOS distinction (from autonomous abstract test) (HM 16.14). The remainder of the paper is organized as follows: Section II presents the proposed structure; Section III dives further into the QRO module's plan; Section IV digs into the techniques utilized for spatial and transient goal resampling; Section V presents and talks about the exploratory plan and results; lastly, Section VI gives ends and ideas for future examination.

2. LITERATURE STUDIES

A. Mackin, F. Zhang, and D. R. Bull This study introduces the BVI-HFR video database, which comprises content with frame rates ranging from 15 to 120 frames per second and may be used to highlight the benefits and drawbacks of higher frame rates, as well as investigate the function of frame rates from capture through distribution. The need to expand the present video parameter space of spatial resolutions and display sizes, to include, among other things, a larger colour gamut, richer dynamic range, and higher frame rates, grows as the demand for higher quality and more immersive video content grows. Increased frame rate can reduce motion blur while simultaneously reducing temporal aliasing and the accompanying visual artefacts, resulting in a more realistic depiction of a scene.

M. Shen, P. Xue, and C. Wang Oversampling a still picture before pressure doesn't guarantee satisfactory picture quality, as per reports. In low piece rate video coding, downsampling before video pressure might decrease the impeding impact and increment the pinnacle signal-to-commotion proportion of the decoded outlines. Oversampling a still picture before pressure doesn't guarantee satisfactory picture quality, as per reports. In low piece rate video coding, downsampling before video pressure might decrease the impeding impact and increment the pinnacle signal-to-commotion proportion of the decoded outlines.

3. PROPOSED METHOD

In order to maximise rate-quality performance, the proposed architecture, depicted in Fig. 1, blends spatiotemporal adaptability with video encoding. The QRO module, which is responsible for estimating the appropriateness for both audio and video, processes video frames from the full quality video as the first step.

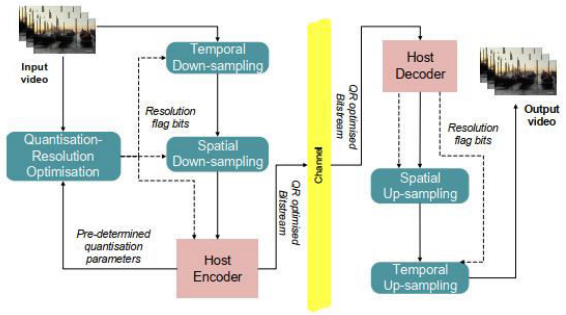


Diagram of the proposed resolution adaptation framework for video compression

Given the video's content and the input quantization value, spatial and temporal adaptation is required (QP). There are two choices: one for geographical adaptation and one for temporal adaptation. The modules that perform spatial and temporal downsampling are then controlled by these decisions. Flag bits are used to signify the adaption in the bitstream. The resolution optimised video is therefore the host encoder's input.

The flag bits are removed from the bitstream at the decoder, and the host decoder decodes resolution resampled video frames. Finally, based on the encoder's resampling selections, video frames are spatially and momentarily upsampled to the original resolution for display. Temporal adaptation choices are performed every two frames, and each frame requires the addition of one flag bit to the bitstream.

Spatial adaptation decisions, on the other hand, are decided for each Group-of-Pictures (GOP), and each GOP requires one flag bit. If two following GOPs have distinct spatial choices, such as the second GOP's resolution being half that of the first, a split is introduced at that point, and two independent bitstreams are encoded. Because HEVC does not enable the encoding of multiple spatial resolutions by default, this process is necessary.

4. SOFTWARE USED

MATLAB

MATLAB® is a high-overall performance technical pc language. It combines computation, visualization, and programming in a user-pleasant interface with troubles and answers written in not unusual place mathematical notation.

MATLAB is an interactive gadget with an array as its primary records detail that doesn't want dimensioning. This lets in you to clear up many technical computing problems in a fragment of the time it takes to construct a programme in a scalar non interactive language like C or FORTRAN, mainly ones the use of matrix and vector formulations. Matlab is an acronym for matrix laboratory. MATLAB became created to make matrix software program evolved via way of means of the LINPACK and EISPACK programmes greater accessible.

Today, MATLAB engines encompass the LAPACK and BLAS libraries, making software program for matrix computing state-of-the-art. MATLAB has developed over a length of years with enter from many customers. In college environments, it's far the usual educational device for introductory and superior guides in mathematics, engineering, and science. In industry, MATLAB is the device of preference for high-productiveness research, development, and analysis. Toolboxes are a form of add-on application-unique answer to be had in MATLAB.

Toolboxes are important for maximum MATLAB customers on account that they can help you recognize and use expert technologies. Toolboxes are units of MATLAB functions (M-files) that amplify the MATLAB surroundings to address positive difficulty types. Signal processing, manage systems, neural networks, fuzzy logic, wavelets, simulation, and plenty of different regions have toolboxes to be had.

5. RESULTS AND ANALYSIS

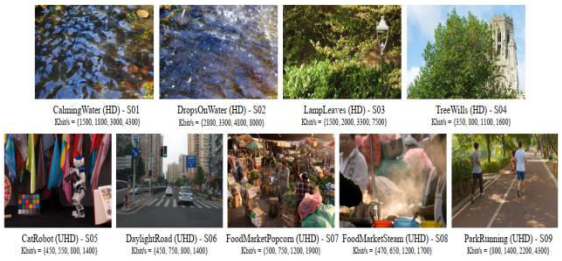
The suggested framework was integrated into the HEVC test model HM 16.14 and submitted to the International Conference on Image Processing (ICIP) 2017 Grand Challenge on Video Compression Technology [15]. The goal of this challenge was to find methods that increase video compression far beyond what is already available. It's worth noting that the subjective findings reported were achieved separately by the challenge's organisers. The JVET (Joint Video Exploration Team) UHD test set [21] and the BVI Texture database [22] were used to create the test dataset, which consists of 9 sequences, 4 HD (1920 1080) and 5 cropped UHD (2560 1600).

The organisers also provided 4 target rate points per sequence, as well as the corresponding HEVC (HM 16.14) anchors for each rate point. These were chosen primarily to give low-quality anchors that may be enhanced perceptually by the submissions. The QP values were systematically changed until the output bitrates reached sufficiently near to the target bitrates in order to satisfy the target bit rates for the test sequences. Figure 1 shows a sample frame and target rate points for each sequence.

PSNR, Video Multimethod Assessment Fusion (VMAF) [23], and subjective assessments were used to generate the test findings. Using the submissions and anchors, a single-stimulus technique was used for the subjective testing, with 28 subjects adhering to the home environment criteria stated in BT.500-13 [24]. TABLE I compares the proposed technique to the HEVC anchor using Bjntegaard measures [25] on PSNR, VMAF, and subjective MOS.

The rate-quality curves for two test sequences, LampLeaves (S03) and Cat Robot, are also shown in Fig (S05). The BD-MOS values were calculated using the technique described in [26]. It should be emphasised that the suggested technique has made

considerable improvements over the anchor codec, with an increase



Test sequences and target bitrates used for experimental results: proposed for the Grand Challenge on Video Compression Technology at ICIP 2017. All sequences are 60 fps except for Park Running which is 50fps.

Average BD-rate increases of 14.5 percent (using PSNR) and 0.55 BD-PSNR. The results are more obvious when utilising VMAF, which correlates better with subjective quality [27], with an average of 21.2 percent BD-rate and 6.1 BD-VMAF. Finally, subjective tests show that the proposed framework improves perceptual quality, with an average BD-MOS of 0.52. The results reveal that ViSTRA produces greater increases at higher spatial resolutions, with 17.7% BD-rate (PSNR) for 25601600 test sequences against 10.4% for 1920 1080. This is because higher resolutions have more spatial redundancy, which implies that the downsampling process will lose less information.

For the two sample sequences, the figure also contrasts ViSTRA's rate-quality performance without the usage of the CNN at the decoder. The CNN improves the quality of reconstructed frames by 0.19 dB and 3.5 VMAF values on average, resulting in a BD-rate improvement of 6.0 percent based on PSNR and 14.1 percent based on VMAF. Except for TreeWills, only spatial resampling is used for the majority of sequences and rate points investigated.

This is because the greatest frame rate employed in the test sequences is 60 frames per second, and temporal resampling is most advantageous at higher frame rates or for slow motion sequences, which TreeWills include. According to a different research, spatial adaptability is responsible for 12.9 percent of the overall BD-rate increases for this sequence based on PSNR. Furthermore, at only one rate point, LampLeaves at 7500 kbps, no adaptation was used, and the original resolution was encoded (see Fig. 4 (a)).

The average encoding time is 0.58 times faster than the HM 16.14 encoder, despite the suggested approach's complexity. This is because ViSTRA enables for the encoding of films with lower spatial and temporal resolution, which reduces the encoding time dramatically. However, because to the use of the CNN for spatial resolution upscaling, the average decoding time of ViSTRA is 61 times that of HM. These results were acquired using a shared cluster from the University of Bristol, which features SandyBridge CPUs with 16 cores, a clock speed of 2.6 GHz, and 4GB of memory. NVIDIA K20 GPU nodes were used to execute the decoding work.

Experimental results of the proposed method compared to HEVC HM 16.14 Anchor.

Sequence	BD-rate (PSNR) [%]	BD-PSNR [dB]	BD-rate (VMAF) [%]	BD-VMAF	BD-rate (MOS)	BD-MOS
S01	-12.7	0.33	-19.2	4.3	-18.1	0.26
S02	-1.4	0.13	-11.2	2.8	-41.4	0.49
S03	-9.9	0.28	-11.9	2.8	-26.4	0.32
S04	-17.5	0.52	-23.5	5.8	-29.6	0.50
Avg. HD	**-10.4**	**0.32**	**-16.5**	**3.9**	**-28.9**	**0.40**
S05	-19.6	0.87	-26.3	8.9	-44.0	0.81
S06	-14.9	0.51	-21.5	6.8	-34.8	0.66
S07	-16.9	0.80	-24.4	8.2	-25.4	0.42
S08	-20.9	1.04	-29.7	9.4	-46.5	0.77
S09	-16.4	0.49	-23.1	5.7	-36.0	0.44
Avg. UHD	**-17.7**	**0.74**	**-25.0**	**7.8**	**-37.3**	**0.62**
Avg. total	**-14.5**	**0.55**	**-21.2**	**6.1**	**-33.6**	**0.52**

6. CONCLUSIONS

This paper proposes a spatio-temporal resolution adaptation framework for video coding, ViSTRA, which optimally resamples input video frames during encoding and reconstructs the full resolution video frames at the decoder. We propose a quantization-resolution module which computes features from the original uncompressed input video frames and determines the optimal spatial and temporal

resolution at which to encode them. At the decoder, we apply frame repetition and a Convolutional Neural Network (CNN) for temporal and spatial resolution upscaling, respectively. This framework has been integrated into HEVC test model HM 16.14 and extensive experimental results were conducted using objective quality metrics and subjective tests. These show that significant coding gains can be achieve by applying the proposed framework for video coding. Future work will focus on improving the performance and reducing the complexity of the CNN for spatial-temporal resampling and on testing more immersive video formats including 4K resolution and 120 fps sequences.

REFERENCES

[1] A. Mackin, F. Zhang, and D. R. Bull, "A study of subjective video quality at various frame rates," in IEEE International Conference on Image Processing, Sept 2021, pp. 3407–3411.

[2] A. Mackin, M. Afonso, F. Zhang, and D. R. Bull, "A study of subjective video quality at various spatial resolutions," in IEEE International Conference on Image Processing, 2020.

[3] M. Shen, P. Xue, and C. Wang, "Down-sampling based video coding using super-resolution technique," IEEE Transactions on Circuits and Systems for Video Technology, vol. 21, no. 6, pp. 755–765, 2018.

[4] G. Georgis, G. Lentaris, and D. Reisis, "Reduced complexity superresolution for low-bitrate video compression," IEEE Transactions on Circuits and Systems for Video Technology, vol. 26, no. 2, pp. 332–345, 2017.

[5] R. Wang, C. Huang, and P. Chang, "Adaptive downsampling video coding with spatially scalable rate-distortion modeling," IEEE Transactions on Circuits and Systems for Video Technology, vol. 24, no. 11, pp. 1957–1968, 2017.

[6] J. Dong and Y. Ye, "Adaptive downsampling for high-definition video coding," IEEE Transactions on Circuits and Systems for Video Technology, vol. 24, no. 3, pp. 480–488, 2014.

[7] Y. Li, D. Liu, H. Li, L. Li, F. Wu, H. Zhang, and H. Yang, "Convolutional neural network-based block up-sampling for intra frame coding," IEEE Transactions on Circuits and Systems for Video Technology, pp. 1–1, 2017.

[8] Z. Ma, M. Xu, Y.-F. Ou, and Y. Wang, "Modeling of rate and perceptual quality of compressed video as functions of frame rate and quantization stepsize and its applications," IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, no. 5, pp. 671–682, 2016.

[9] Q. Huang, S. Y. Jeong, S. Yang, D. Zhang, S. Hu, H. Y. Kim, J. S. Choi, and C.-C. J. Kuo, "Perceptual quality driven frame rate selection (PQD-FRS) for high-frame-rate video," IEEE Transactions on Broadcasting, vol. 62, no. 3, pp. 640–653, 2016.

[10] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super resolution using deep convolutional networks," IEEE Transactions on pattern analysis and machine intelligence, vol. 38, no. 2, pp. 295–307, 2016.

[11] J. Kim, J. Kwon Lee, and K. Mu Lee, "Accurate image super resolution using very deep convolutional networks," in IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1646–1654.

[12] F. Zhang, A. Mackin, and D. R. Bull, "A frame rate dependent video quality metric based on temporal wavelet decomposition and spatiotemporal pooling," in IEEE International Conference on Image Processing, Sept 2015, pp. 300–304.

[13] A. Mackin, M. Afonso, F. Zhang, and D. R. Bull, "SRQM: a video quality metric for spatial resolution adaptation," in Picture Coding Symposium (PCS), 2015.

[14] M. Afonso, F. Zhang, A. Katsenou, D. Agrafiotis, and D. Bull, "Low complexity video coding based on spatial resolution adaptation," in IEEE International Conference on Image Processing. IEEE, 2015, pp. 3011–3015.

Dr.Y. L. AJAY KUMAR had graduated B.Tech from G.Pulla Reddy Engineering College, Kurnool. M.Tech from JNTUA, Anantapuramu and Ph.D from JNTUA, Anantapuramu. Currently he is working as Associate Professor and Research and Development Director in Anantha lakshmi Institute of Technology &Sciences, Ananthapuramu, AP, India. His areas of interest are VLSI and Embedded. He has 12 years of experience in teaching. He published 42 international journals and attended 5 Conferences. He has one Patent Journal.

Dr.M.VENKATA SREERAJ. Has been Completed his PhD from Jawaharlal Nehru Technical University, Anantapur. In the Data Communication and Networking Discipline . Having 17 Years of Experience. His areas of interest are Wireless Sensor Networks and Data Communication. He has published 6 International Journals and attended 5 International Conferences.

Mr. N.NAVEEN KUMAR had graduated B.Tech from JNTUA, Ananthapuramu, M.Tech from JNTUA, Ananthapuramu. Currently Pursuing Ph.D from JNTUA, Ananthapuramu and he is working as an Assistant Professor and NSS Coordinator in Anantha Lakshmi Institute of Technology & Sciences, Ananthapuramu, AP, India. He has Published 6 International Journals. His areas of interest are Microwave Antennas, Communication Systems, VLSI and Image Processing.