iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** IV **Month of publication:** April 2021

DOI: https://doi.org/10.22214/ijraset.2021.33676

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue IV Apr 2021- Available at www.ijraset.com

An Overview of the Legacy of Video Compression Technologies and New Advancement in Video Conference Technology

Sameer Mahajan[1], Gayatri Bangar[2], Vedang Naik[3]
[1]*Student, Department of Computer Engineering, TEC, University of Mumbai, Mumbai, India.*
[2]*Student, Department of Information Technology, TEC, University of Mumbai, Mumbai, India.*

Abstract: Video compression is piece of communication as the demand for higher quality videos and lower data consumption increases so are newer standards being adopted to keep up with those demands. We look at various video formats from the past as well as a new way, developed by Nvidia, for talking head video-conferencing that is far-more efficient than the previous standards.
Keywords: Video compression, H.261, H.263, H.264, HEVC, Video Conferencing, Neural Talking-head Synthesis.

I. INTRODUCTION

The primary goal of most digital video coding standards has been to optimize coding efficiency. Coding efficiency is the ability to minimize the bit rate necessary for the representation of video content to reach a given level of video quality—or, as alternatively formulated, to maximize the video quality achievable within a given available bit rate.[1]. Various standards have been adopted over the years, we will look at some of the standouts as well as future proposals.

II. H.261

This is an older video format where two types of image frames are defined: intra-frames (l-fames) and inter-frames (P-frames). I-frames are treated as independent images. P-frames are not independent. They are coded by a forward predictive coding method in which current macroblocks are predicted from similar macroblocks in the preceding I-frame or P-frame, and differences between the macroblocks are coded. Temporal redundancy removal is performed via the P-frame coding, whereas spatial redundancy removal is done by the I-frame coding.[2]. Discrete cosine transform (DCT) is used for spatial redundancy reduction while, temporal prediction works using a block-based motion estimation (ME) and compensation (MC).[3]

III. H.263

H.263 is a standard that, despite sharing many of its core features with H.261 exists to cover many of the latter shortcomings and thus offers significant advantages over it. Few of those are

1) *Bit rate:* For H.261 the target bit-rate is p*64 bits where p is number between 1 and 30 whereas H.264 supports bit rate less than 64 bits. [3,4]
2) *Picture Formats:* H.263 supports 3 picture formats Common Intermediate Format (CIF), Quarter-CIF (QCIF), Sub-CIF (SCIF). H.261 supports only CIF and QCIF. [3,4]
3) *Block Structure:* Both formats use the same macro block (MB) and group of blocks (GOB) structure but in H.263, for error resilience each GOB contains only one macroblock row. Thus, for QCIF, each GOB has 11 MBs, compared to the 11 x 3 = 33 MBs used in H.261. Furthermore, in H.263, optional header information can be inserted in the GOB layer. This allows the coder to insert extra synchronization codewords for improved error resilience. [3,4]
4) *Motion Compensation Accuracy:* Motion vectors for H.263 have half-pixel accuracy compared to the integer pixel accuracy for motion vectors used in H.261. Half pixel values are found using bilinear interpolation. [3,4]
5) *Loop Filter:* H.261 employs a spatial-domain loop filter in the coding loop to reduce the block effects due to block-based motion estimation. H.263 does not employ such a filter since the bilinear interpolation used in H.263 for half-pixel motion compensation introduces some low-pass filtering as a side-effect. [3,4]

There are also other changes introduced to error correction, variable-length code tables (VLC), quantization parameters, and macro-block addressing. [3,4]

IV. H.264

In the H.264 format the videos are transformed into a bitstream by the encoder and the decoder then recreates the video from this bitstream. Both the encoder and decoder follow three processes each. [5]

A. *Encoder Processes*

1) *Prediction:* Based on previously-coded data encoder forms a prediction of the current macroblock, using either intra-prediction from the current frame or from other frames that have already been coded and transmitted using inter prediction. The encoder subtracts the prediction from the current macroblock to form a sample called a residual.
2) *Transformation and Quantization:* A block of residual samples is transformed using a 4 × 4 or 8 × 8 integer transform, an approximate form of the Discrete Cosine Transform (DCT). The transform outputs a set of weighting values for a standard basis pattern. When combined, the block of residual samples is recreated. The output is quantized according to a quantization parameter (QP). QP determines the level of compression: higher the QP higher the compression and lower the video quality whereas a lower QP means lower compression but higher video quality.
3) *Bitstream Encoding:* To form a compressed bitstream the values of quantized transform coefficients, information for the decoder to recreate the prediction, information about the structure of the compressed data and the compression tools used during encoding and information about the complete video sequence must be encoded.

B. *Decoder Processes*

1) *Bitstream Decoding:* The encoded bitstream that is received is decoded to get the information required to recreate the video images.
2) *Rescaling and Inverse Transform:* The quantized transform coefficients are re-scaled. Each coefficient is multiplied by the QP value to restore its original scale. An inverse transform combines the standard basis patterns, weighted by the re-scaled coefficients, to re-create each block of residual data.
3) *Reconstruction:* For each macroblock, the decoder creates a prediction similar to the one created by the encoder and adds it to the decoded residual and that reconstructs the decoded macroblock which becomes part of the displayed video.

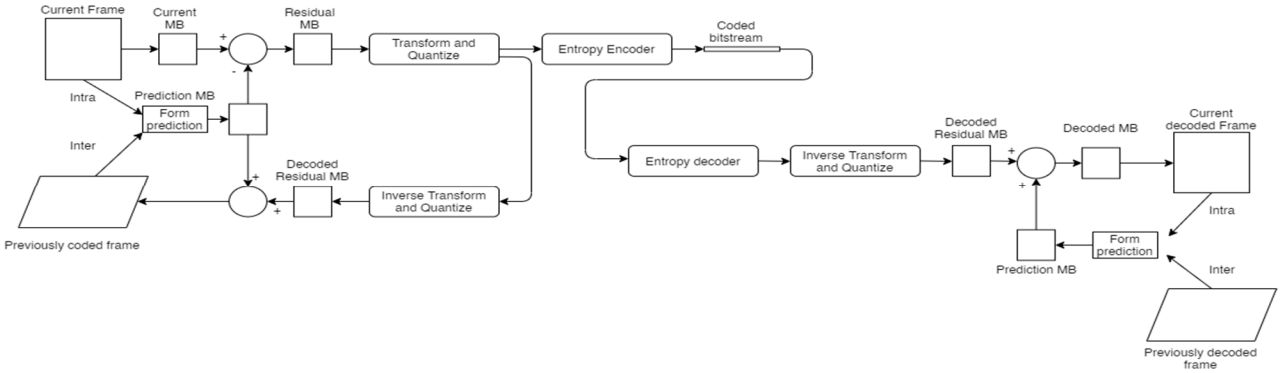


Fig 1. H.264 Encoder and Decoder

The H.264 format introduced many additional features to enhance the ability to predict the values of the content of the picture to be encoded, increase coding efficiency and make the operations more robust and flexible [6].

a) *Better Predictions:* Features introduced to improve the predictions made for the image content are, Variable block-size motion compensation with small block sizes, Quarter-sample-accurate motion compensation, Motion vectors over picture boundaries, Multiple reference picture motion compensation, Decoupling of referencing order from display order, Decoupling of picture representation methods from picture referencing capability, Weighted prediction, Improved “skipped” and “direct” motion inference, Directional spatial prediction for Intra coding, In-the-loop deblocking filtering

b) *Increased Efficiency:* Small block-size transform, Hierarchical block transform, Short word-length transform, Exact-match inverse transform, Arithmetic entropy coding, Context-adaptive entropy coding

c) *Robustness and Flexibility:* Parameter set structure, NAL unit syntax structure, Flexible slice size, Flexible macroblock ordering (FMO), Arbitrary slice ordering (ASO), Redundant pictures, Data Partitioning, SP/SI synchronization/switching pictures

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue IV Apr 2021- Available at www.ijraset.com

V. HEVC

The High Efficiency Video Coding (HEVC) doubles the coding efficiency compared to the H.264/AVC.[7] with a larger focus on increased video resolution and more use of parallel processing architectures.[8] The fundamental difference between the two is the use of an adaptive quadtree structure is the use of a coding tree unit (CTU). The quadtree coding structure is described by means of blocks and units. A block is an array of samples and sizes, while a unit encapsulates one luma and corresponding chroma blocks together with syntax needed to code these. A CTU includes coding tree blocks (CTB) and syntax specifying coding data and further subdivision which results in coding unit (CU) leaves with coding blocks (CB). Each CU incorporates more entities for the purpose of prediction, called prediction units (PU), and of transform, so-called transform units (TU). Similarly, each CB is split into prediction blocks (PB) and transform blocks (TB). This approach is useful in larger resolution videos.[7]

The features added in the HEVC can be divided into three categories:

A. Video Coding Layer

Features added to improve video coding efficiency Coding tree units and coding tree block (CTB) structure, Coding units (CUs) and coding blocks (CBs), Prediction units and prediction blocks (PBs), TUs and transform blocks, Motion vector signaling, Motion compensation, Intrapicture prediction, Quantization control, Entropy coding, In-loop deblocking filtering, Sample adaptive offset (SAO)

B. High-Level Syntax Architecture

Features added to improve robustness and flexibility are Parameter set structure, NAL unit syntax structure, Slices, Supplemental enhancement information (SEI) and video usability information (VUI) metadata

C. Parallel Decoding Syntax and Modified Slice Structuring

Features introduced to enhance the parallel processing capability or modify the structuring of slice data for packetization purposes tiles, Wavefront parallel processing, Dependent slice segments.[8]

VI. NEURAL TALKING-HEAD SYNTHESIS

The methods discussed above transmitted the compressed video using various compression techniques. Ting-Chun Wang et.al [8] proposed a neural model and demonstrates a major application in video conferencing. This method not only compresses the video efficiently but generates the video using only a source image and some information. The model takes a reference image containing the intended person's appearance and a driving video illustrating the motion required in the output. This technique performs better than H.264 which is an efficient and commonly used standard for video compression.

Wang's method performs better as it achieves better visual quality as compared to the state-of-the-art methods with a better data rate as the model is able to reduce 10 times more bandwidth than the commercial H.264 standard. H.264 transmits more than 6 to 12 times the data transmitted by Wang's method and thus the quality significantly decreases during video conferencing using H.264 if we limit the optimum data transmission which doesn't happen in neural talking-head synthesis.

Moreover, this model doesn't require 3D graphics and allows for changing the viewpoint of the talking-head. Keypoint representation is used to encode the motion which unsupervisedly breaks down identity-specific and motion-related information. This approach renders a one-shot talking head using a deep network with 2D graphics. The 2D graphics overpowers the 3D graphics approach as 2D graphics can better render hair, beard, etc. 2D graphics can directly combine accessories like hats, eyeglasses and unlike 3D graphics, they are not expensive and strenuous.

Adding to the excellent video compression technique this method can also perform face frontalization, face rotation, and motion transfer.

Generative Adversarial Network plays a key role in talking head synthesis. In 2014, Goodfellow *et al.* [9] and others from the University of Montreal first proposed the idea of Generative Adversarial Networks (GAN). In any domain, GANs are able to imitate any data distribution: images, audio, voice, and prose. A GAN network is made up of a parallel working discriminator (D), and a generator (G) [10]. Wang's method uses GANs to synthesize talking-head videos.

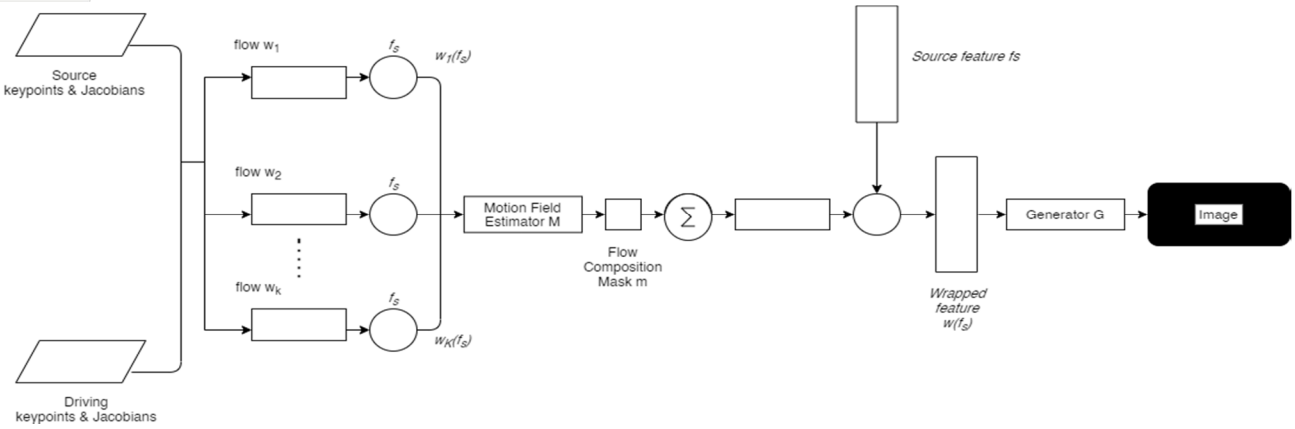


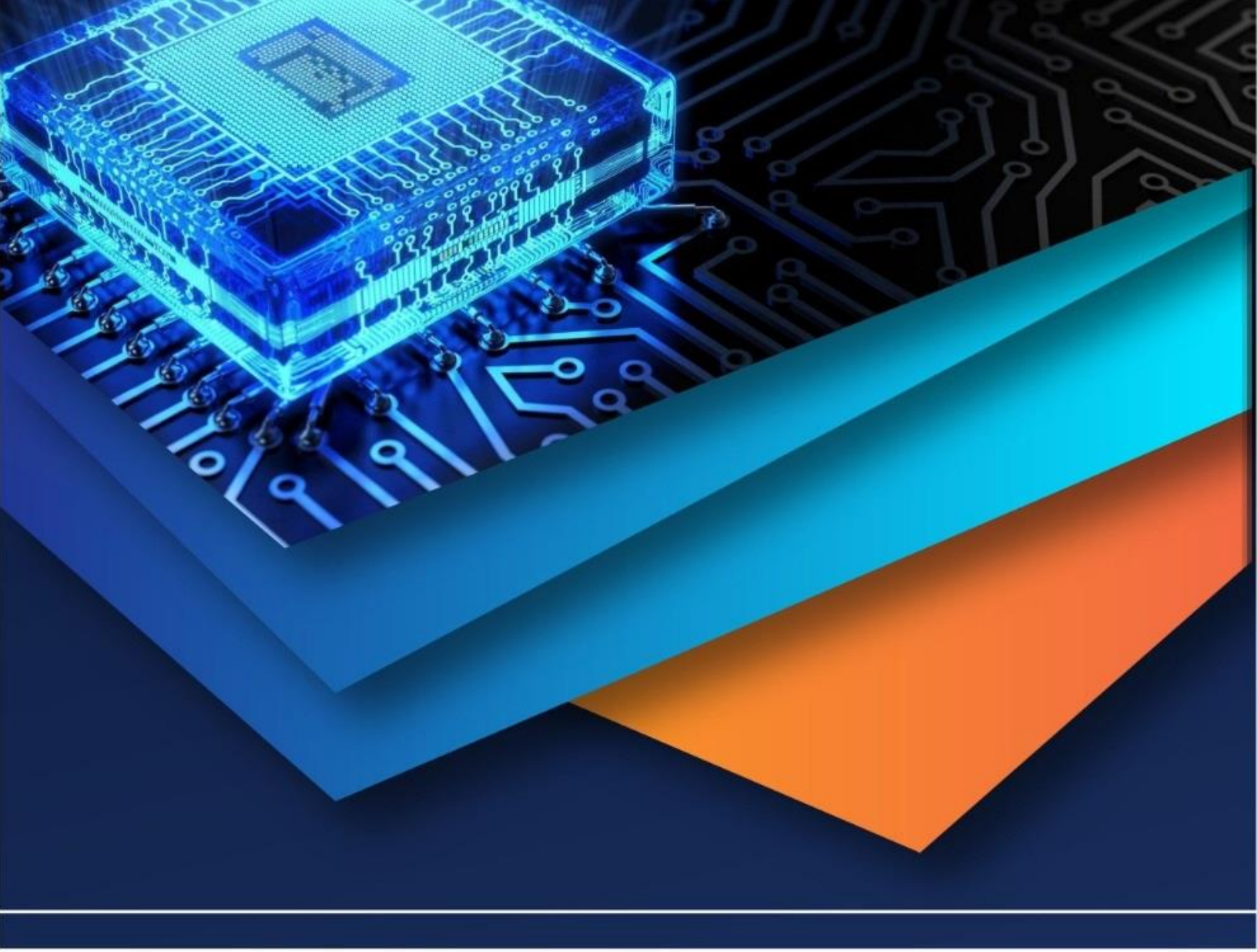
Fig. 2. Neural Talking-head synthesis flowchart

VII. CONCLUSION

Thus we have looked at various video-coding formats that are presently in use and also seen how neural networks and artificial intelligence can be used to make the process of video compression far more efficient.

REFERENCES

[1] Ohm, J., Sullivan, G., Schwarz, H., Tan, T. and Wiegand, T., 2012. Comparison of the Coding Efficiency of Video Coding Standards—Including High Efficiency Video Coding (HEVC). IEEE Transactions on Circuits and Systems for Video Technology, 22(12), pp.1669-1684.
[2] Ze-Nian Li, Mark S. Drew, and Jiangchuan Liu. 2014. Fundamentals of Multimedia (2nd. ed.). Springer Publishing Company, Incorporated.
[3] Bernd Girod, Eckehard G. Steinbach, and Niko Faerber "Comparison of the H.263 and H.261 video compression standards", Proc. SPIE 10282, Standards and Common Interfaces for Video Information Systems: A Critical Review, 102820D (25 October 1995); https://doi.org/10.1117/12.227952
[4] Vasudev Bhaskaran and Konstantinos Konstantinides. 1997. Image and Video Compression Standards: Algorithms and Architectures (2nd. ed.). Kluwer Academic Publishers, USA.
[5] Iain E. Richardson. 2010. The H.264 Advanced Video Compression Standard (2nd. ed.). Wiley Publishing.
[6] T. Wiegand, G. Sullivan, G. Bjontegaard and A. Luthra, "Overview of the H.264/AVC video coding standard", IEEE Transactions on Circuits and Systems for Video Technology, vol. 13, no. 7, pp. 560-576, 2003. Available: 10.1109/tcsvt.2003.815165.
[7] G. Sullivan, J. Ohm, W. Han and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard", IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, no. 12, pp. 1649-1668, 2012. Available: 10.1109/tcsvt.2012.2221191.
[8] T. Wang, A. Mallya and M. Liu, "One-Shot Free-View Neural Talking-Head Synthesis for Video Conferencing", NVIDIA Corporation, 2021.
[9] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville,and Yoshua Bengio. Generative adversarial networks. In *NeurIPS*, 2014
[10] "GAN — What is Generative Adversarial Networks GAN", Medium, 2021. [Online]. Available: https://jonathan-hui.medium.com/gan-whats-generative-adversarial-networks-and-its-application-f39ed278ef09.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY