ANALYSIS OF VIDEO DATA COMPRESSION ALGORITHMS AND OPTIMIZED FOR USE IN REMOTELY CONTROLLED DRONES

Kirill Smirnov [1], **Anastasia Mozhaeva** [2]

[1] Institute of Radio and Information Systems (IRIS), Vienna, Austria; smirnov@media-publisher.eu

[2] The University of Waikato Hamilton, New Zealand

ABSTRACT

Due to the widespread use of unmanned aerial vehicles (UAVs) in the civil sphere, there is a need to improve the video image generation quality and transmission technologies, changing the video streams coding to reduce their size and improve quality. Video streaming technologies are moving further towards complicating the video transmission systems used, which helps improve the received video data quality. Higher of transmitted video signal quality – more requirements for frequency band increase used. This fact leads us to use effective video compression algorithms that allow us to combine high image quality with a narrow bandwidth of frequencies used. In this paper, currently existing coding algorithms, their efficiency and computational complexity are studied. Among the algorithms under consideration there will only be those whose effectiveness has been proven by finding their application in modern realities. Based on the research, conclusions will be drawn regarding the feasibility of using each encoding algorithm for transmitting video data in real time.

KEYWORDS: *Encoder, Video Coding Standards, Video Compression, Drone, Real-Time Video Transmission.*

DOI: 10.36724/2664-066X-2023-9-2-9-16

Received: 30.01.2023
Accepted: 10.03.2023

Citation: Kirill Smirnov, Anastasia Mozhaeva, "Analysis of video data compression algorithms and optimized for use in remotely controlled drones," *Synchroinfo Journal* **2023**, vol. 9, no. 2, pp. 9-16.

1 Introduction

Due to the widespread use of unmanned aerial vehicles (UAVs) in the civil sphere [1], there is a need to improve the quality of video image generation and transmission technologies, in particular, changing the coding of video streams to reduce their size and improve quality. The systems disadvantage for generating and transmitting service video data currently used for UAVs is the use of outdated analog standards for encoding and image transmission, which have image resolution, poor noise immunity and, as a result, tire the UAV operator.

Video streaming technologies are moving further towards complicating the video transmission systems used, which helps improve the quality of received video data. The higher the quality of the transmitted video signal, the more the requirements for the used frequency band increase. This fact leads us to the need to use effective video compression algorithms that allow us to combine high image quality with a narrow bandwidth of frequencies used. There are many different video encoding techniques available today, and choosing the most effective one requires extensive research. In addition, an algorithm that is actively used to solve one specific problem may not be well applicable to another.

The currently existing coding algorithms, their efficiency and computational complexity are studied. Among the algorithms under consideration there will only be those whose effectiveness has been proven by finding their application in modern realities. Based on the research, conclusions will be drawn regarding the feasibility of using each encoding algorithm for transmitting video data in real time.

2 Video coding principles

Encoders work based on the principles of human image perception, which reduces the amount of data required to transmit images, and therefore increases the efficiency of using communication channels. To reduce the amount of data, you can remove redundant or irrelevant information from the data stream. Redundant information is information that has no information content, or simply information that can be easily and losslessly recovered using mathematical algorithms on the receiving side.

Using only one method to eliminate one or another type of redundancy in encoders would be an ineffective solution [2], so most of them use several of them at once. Modern encoders work with the following types of redundancy:

- temp rary redundancy;
- s atial redundancy;
- psych visual redundancy;
- entr pic redundancy.

Various video encoding algorithms analysis

There are many video codec standards developed for different purposes. After analyzing their characteristics and operating principles, it will be possible to draw immediate conclusions about the possibility of using them in remotely controlled drones.

3 Analysis of video coding standards

Motion JPEG video coding standard

MJPEG [3] encodes a video sequence as a series of JPEG images, each corresponding to one video frame, and uses only i-coding. The JPEG standard was not originally intended for this use, but MJPEG has become popular and is used in a number of video communications and data storage applications. This standard makes no attempt to eliminate temporal redundancy in a moving video sequence, and therefore compression performance is poor compared to codecs that use inter-frame coding.

Spatial redundancy is reduced by using discrete cosine transform coding on so-called 8x8 pixel macroblocks. After this, the resulting decorrelated signal is quantized.

Since temporal redundancy is not eliminated in the encoder, it is less sensitive to sudden movements in the frame at the cost of reduced encoding efficiency.

Statistical redundancy is reduced by using entropy coding. To do this, the encoder uses the Huffman algorithm.

MPEG-4 Part 10 (AVC, H.264) video coding standard

The most popular codec at the moment, first released in 2003 [4]. Improvements to this standard are still being made to this day in the form of updates that add support for new functions.

The reduction in spatial redundancy in the algorithm occurs due to the algorithm predicting the calculated blocks based on the previous calculated macroblock. To reduce the number of transmitted bits, a difference signal obtained from the calculated and predicted blocks is used. The intra-frame prediction block can have dimensions of 4x4, 8x8 and 16x16 pixels. Smaller blocks (4x4 and 8x8) are transformed first using a discrete cosine transform, and then they are combined into new 4x4 blocks, which are additionally decorrelated by the Hadamard transform. The algorithm has 52 levels available for quantization.

To reduce temporal redundancy, a macroblock-based motion estimation and compensation method is used. The format uses both p-frames and b-frames. With downsampling, the accuracy of the motion vector is half a pixel with single-step filtering, and a quarter of a pixel with double filtering. When obtaining greater compression, double filtering is more complex, since the secondary calculation requires the result of the primary one. Prediction of motion vectors occurs on the basis of previous frames, and the more frames, the more memory is required for processing, and at the same time, the greater the accuracy of the estimation, which will be reflected in the form of greater compression efficiency.

Statistical redundancy is handled using entropy coding. Context Adaptive Variable Length Coding (CAVLC) or Context Adaptive Binary Arithmetic Coding (CABAC) can be used to increase compression efficiency by approximately 9-14%. At the same time, however, the complexity of calculations also increases, and as a consequence - an increase in processing time.

To reduce the number of image artifacts resulting from block operations in cycles, a deblocking filter is used.

Some AVC profiles also provide error tolerance. One such method is to reduce the sensitivity of macroblocks to data loss in a packet. By using the technique of separating syntactic elements, protection against inequality errors is achieved. To compensate for a lost or corrupted slice or frame, some low-precision data can be resent by increasing redundancy.

The encoder uses so-called parallel computing, the operating principle of which is as follows: the frame is divided into slices, which can be encoded and decoded independently of each other. When combined with delay processing applications, each slice can be encoded and decoded as soon as it reaches the encoder, without having to wait for an entire frame to be added.

DIRAC video coding standard

A free and opensource standard developed by the BBC in 2008 [5]. Designed as a simple and flexible alternative to AVC, it uses the less commonly used discrete wavelet transform for signal decorrelation and motion compensation. The codec supports 4:4:4, 4:2:2 and 4:2:0 color subsampling.

Spatial redundancy is reduced by using the discrete wavelet transform on an entire frame at a time. This makes it easy to extract low-resolution data in the decoder. Small details are preserved better than in block-based algorithms. The vertical and horizontal components are divided into low and high frequency components by sequential filtering. There are several types of supported wavelet filters, differing in their complexity/quality ratio.

Temporal redundancy is reduced by motion estimation and compensation. During motion estimation, p-frames and b-frames are used, each operating on a maximum of two frames. DIRAC uses a hierarchical approach to create motion vectors, with the current and reference frames being processed step by step by a decimator filter.

The picture is divided into superblocks, and calculations can be made for each. To reduce the number of artifacts that arise during motion compensation, an overlapping block motion compensation (OBMC) algorithm is used, with the number of horizontal macroblocks exactly equal to the number of vertical macroblocks. The accuracy of motion prediction directly depends on the bitrate selected for operation.

Statistical redundancy is dealt with by entropy coding, which is used in three steps: binarization, context modeling and arithmetic coding. Binarization is carried out to create a stream of bits that will be used in further steps. Contextual modeling calculates and predicts whether an observed coefficient will decrease by basing calculations on its neighbors. Arithmetic coding is then applied.

VP9 video coding standard

VP9 is an open and free-to-consumer video compression standard developed by the open community AOM (Alliance for Open Media) [6]. Previously developed under the name Next Generation Open Video (NGOV) and VP-Next. It is an evolutionary development and successor to the VP8 standard.

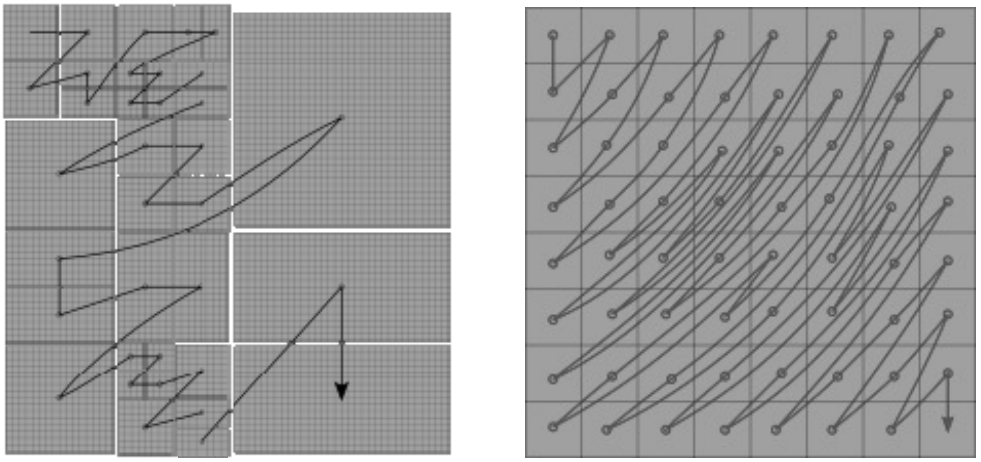
VP9 is a traditional block encoding format. The bitstream structure is much simpler compared to formats that offer similar efficiency at the same bitrate. This makes this format more profitable for streaming.

The main improvement is support for using 64 by 64 blocks. This feature will be useful in Full HD and Ultra HD video. The performance of motion prediction vectors has been improved. In addition to VP8's four modes: fine motion, vertical, horizontal, average, VP9 has support for 6 and oblique directions necessary for linear extrapolation of pixels in intra-frame prediction.

Modern coding tools have:

- eighth p xel precision of motion vectors,
- three diff rent switchable interpolation filters,
- improved a gorithm for selecting support motion vectors,
- improved coding of displacements of motion vectors relative to their reference point,
- improved e tropy coding,
- improved a d adapted (to new block sizes) loop filtering,
- asym etric discrete sinusoidal transformation (ADST),
- larger discrete cosine transform blocks (DCT, 16x16 and 32x32) and improved segmentation of frames into areas of particular similarity.

To ensure parallel processing of frames, video fragments can be divided along the boundaries of encoding blocks up to four rows with equal fragments with a width of 256 to 4096 pixels, each of which will be encoded independently of each other. This is necessary for video resolutions greater than 4096 pixels. The block size in bytes is contained in the header, so decoders can decode each fragment in a separate stream. The image is then divided into encoding blocks called 64x64 pixel superblocks, which are adaptively partitioned into a quadrant encoding structure. They can be divided horizontally, vertically, or in both dimensions; square subblocks can be subdivided recursively into 4x4 pixel blocks. Subblocks are encoded in raster scan order: left to right, top to bottom (Figure 1).



a) b)

Fig. 1: Scanning order: a) superblocks; b) subblocks

Starting with each key frame, decoders store 8 frames in a buffer, for use as reference frames or for display later. Frames that are transmitted have notification markers in which of the buffers will be recorded and additional decoding is possible into another buffer, which is not displayed.

The encoder has the ability to send a minimal frame that simply triggers the display of one of the buffers. Each subframe can reference three buffered frames for temporal prediction. Up to two of these reference frames can be used in each encoding block to compute a spatial offset prediction (motion compensation) from a particular reference frame, or from averaging the 2 reference frames ("prediction coupling mode"). Ideally, the small remaining difference (delta encoding) from the computed prediction to the actual image content is converted using DCT or ADST (for edge blocks) and quantized.

AV1 video coding standard

AV1 is an open video encoding format intended for video transmission over the Internet, conceived as a continuation of VP9 [7]. The codec is designed to allow scaling across different resolutions and bitrates. This allows you to introduce support for different devices with different processing power. For transformation, superblocks of 128x128 or 64x64 pixels are used, which are then divided into smaller blocks. The algorithm uses both i-frames and p-frames.

Spatial redundancy is handled by combining prediction with traditional transformation techniques. Predictions are made based on already processed neighboring elements, using the available 65 angular modes. To eliminate prediction errors, AV1 uses the discrete cosine transform and the asymmetric analogue of the discrete Fourier transform. Combining two 1D transforms is used to apply different transforms for the horizontal and vertical dimensions.

Temporal redundancy is eliminated by motion estimation and compensation. Motion vectors are calculated based on reference frames, their values are remembered and can be used in various motion prediction modes. The motion vector prediction accuracy is 1/8 pixel.

The filtering system consists of several steps consisting of processing vertically, horizontally and along the edges of the selected block. Depending on the size of the filtered block, an adaptive filter is applied to it with varying strength, the size and limits of which are adapted to each block after it is analyzed.

Entropy coding occurs using a non-binary arithmetic coding algorithm. Its peculiarity is that each symbol can take on eight possible meanings. This adds complexity to the system, but allows it to process multiple characters in a loop, thereby improving performance.

DCP video coding standard

DCP – Digital Cinema Package or Digital Cinema Package, developed by a consortium of the world's leading film studios DCI (Digital Cinema Initiatives) [8]. According to this standard, the source images of each frame when saved are compressed according to the JPEG 2000 standard, using the CIE XYZ color space with a color depth of 12 bits per channel. All this, along with the audio, is packaged in a DCP package, which involves the use of an MXF container with a maximum flow limit of 250 megabits per second and is encrypted if necessary. To play encrypted content, a so-called “key” is sent to the cinema by email, which contains information: the number of the cinema server and the period of time in which this cinema will be able to launch this package.

The DCP package consists of 6 files:

• mxf contai er with image (video)

• mxf contai er with sound (audio)

• CPL (Composition Play List) – playlist of the entire package, some type of content (advertising, trailer, movie), timing, etc.

• PKL (Paking List) – description of HASH checksums; when uploading content to movie servers, the checksums are checked to prevent errors during copying.

• ASSETMAP – provides UUID binding to the name and location of files on disk

• VOLINDEX – to indicate on which volume a specific file with such and such a UUID is located, volume number identifier.

NDI video coding standard

NDI is a freely available standard developed by NewTek [9] designed for low latency transmission and reception of video information. The NDI codec is capable of delivering 1080 HD video with a bitrate of about 100 Mbps. Its structure is very similar to MPEG-2, while it supports forms of color subsampling. Each frame is encoded as two fields, with the data packed into 32-bit little-endian words. Each field is divided into four slices, which can be independently decoded; each slice encodes 16 lines, then skips 48 lines (encoded by the other three slices), encodes another 16 lines, and so on until it reaches the bottom of the field, which implicitly ends the slice. The first three bytes of each slice form the length of the slice in bytes (including the three length bytes). The next slice starts in the byte immediately after the end of the previous one, according to their length (this could mean skipping 1-7 bits).

NDI uses a block structure and discrete cosine transform (DCT), which converts video signals into elementary frequency components. In this case, macroblocks have a size of 16x16 pixels (regular blocks are 8x8). Discrete cosine transform coefficients are encoded exactly as in MPEG-2 (same luma/chrominance tables, same scheme for storing coefficient and sign). The coefficient prediction is also taken from the previous block and restarted at 1024 for each new line of the macroblock (not 128, as in MPEG-2). The algorithm can work with 100 quantization levels.

The NDI codec is designed to operate with extremely low latency and is largely implemented in assembly language to ensure the fastest possible video frame compression process. Latency is a factor in both the network connection and the end products. NDI has a technical latency of 16 slices, although in practice most implementations will have a latency of approximately one field. Hardware implementations can provide full end-to-end latency within 8 slices.

4 Conclusions from the comparative analysis

To select a video compression algorithm suitable for use in a remotely controlled drone, it is necessary to compare the standards discussed above. For a clearer comparison, the criteria for each algorithm are presented in Table 1. It is important to note that the table does not include the DIRAC and DCP algorithms. They are both constructed using the wavelet transform, which is not suitable for the intended purpose due to the high complexity of encoding and decoding [10], and therefore the high resource intensity of the algorithms, leading to a long delay time.

Table 1. Comparison of algorithms by spatial compression methods

Codec	Block sizes	Transformation method	Spatial forecasting	In-loop filtration
MJPEG	8x8	2-D DCP	No	No
AVC	16x16, 8x8, 4x4	DCP and Hadamard transformation	Intra-frame macroblock prediction	Horizontal and vertical edge filtering
VP9	Superblocks 64x64, divided into subblocks 4x4	DCP or asymmetric DFT	I-frame macroblock prediction	4-step filtering
AV1	Superblocks 64x64 or 128x128	DCP and asymmetric DFT	I-frame and P-frame prediction	Adaptive Intra-Boundary Filter
NDI	Blocks 8x8, macroblocks 16x16	DCP and Hadamard transformation	I-frame macroblock prediction	Horizontal and vertical edge filtering

When choosing modern algorithms, preference is given to those with larger division block sizes - the effectiveness of such solutions is more noticeable when used with images with resolutions greater than Full HD. The use of blocks divided into subblocks for transformation is also considered more effective in terms of information compression, however, the complexity of the system in this case increases significantly, which leads to an increase in latency [11-13].

Using forecasting based only i-frames, you can increase the encoding speed. At the same time, prediction on p-frames gives even greater results in both coding efficiency and the quality of the resulting image, but this will introduce additional latency into the system.

Table 2. Comparison of algorithms by methods of processing temporal redundancy, entropy coding and characteristic features

Codec	Temporary redundancy	Entropy coding	Peculiarities
MJPEG	No	Variable length Huffman algorithm	No
AVC	Prediction on the previous and next frames, motion vector with ¼ pixel accuracy	CAVLC or CABAC	Parallel computing, error tolerance
VP9	Prediction on the previous frame, motion vector with 1/8 pixel accuracy	Boolean entropy coding	Parallel computing, customizable quality
AV1	Prediction on the previous frame, motion vector with 1/8 pixel accuracy	Non-binary arithmetic encoding	Scaling for different devices
NDI	Prediction on the previous frame	CABAC	Parallel computing, generational stability of processing

Algorithms that use the CABAC algorithm for entropy coding provide a significant gain in coding efficiency compared to others.

Filtering is an important step to maintain the visual quality of an image. Although the procedure itself does not provide any efficiency gains, processing it allows for greater compression to be achieved in other stages of the algorithm. The complexity of the filter directly correlates with the size of the blocks - the larger the size, the more levels of filtering will be needed to eliminate artifacts appearing in the image, and the longer the processing will take.

The possibility of parallel computing, although it somewhat reduces the efficiency of coding, greatly increases the performance of the encoder when using a multi-core processor or a special hardware chip in the system.

Since we are talking about low-latency video encoding, when choosing the right encoder for the task at hand, special attention must be paid not only to encoding efficiency. Modern drone control systems, although significantly superior to their predecessors, still cannot boast of sufficient performance compared to stationary systems. Therefore, it is necessary to choose an algorithm that does not require high performance or has ways to increase response time. This method is the parallel computing capability present in the AVC, VP9 and NDI algorithms. However, the VP9 encoder has a complex system of four filters applied in sequence, which significantly increases the latency of the system. The

remaining standards, AVC and NDI, have a similar structure, since the latter was developed as an alternative to h.264. Due to the great similarity, the difference between the delay time of one and the other will be extremely small, and can be detected exclusively experimentally. Using these encoders to work in drone remote control systems is the optimal solution, combining the required level of efficiency and relative ease of implementation.

5 Conclusion

As a result of the work carried out, the basic principles of video image coding were analyzed, individually or together implemented by algorithms of varying complexity in video encoders. The most popular video compression standards today and the algorithms used in them were reviewed. Based on the analysis, some of the coders considered were considered unsuitable for the intended purpose, while the rest were compared according to a number of criteria derived on the basis of the requirements of the problem being solved. As a result, the two most suitable coders were selected. Since one of them, NDI, is very similar in operating principle to the second, AVC, their difference in performance is extremely difficult to identify analytically in the future. This requires a laboratory experiment, which will be carried out in the course of further work. Therefore, at this stage, the conclusion to the work carried out can be the following statement: both encoders can be effectively used in remote control systems for drones.

REFERENCES

[1] A. Alsoliman, G. Rigoni, D. Callegaro, M. Levorato, C.M. Pinotti, and M. Conti "Intrusion Detection Framework for Invasive FPV Drones Using Video Streaming Characteristics," *ACM Trans. Cyber-Phys. Syst*. 7, 2, Article 12, 2023, doi: 10.1145/3579999.

[2] I.V. Vlasyuk, A.I. Sidorova, E.P. Romanova, “Features of interframe coding of video information according to the MPEG-4 standard,” *T-Comm*. 2010. Vol. 4. No. 9, pp. 50-52.

[3] RTP Payload Format for JPEG-compressed Video [Electronic resource]. Network Working Group – Electronic. Access mode: https://tools.ietf.org/html/rfc2435.

[4] Ya. Richardson, “Video coding. H.264 and MPEG-4 – new generation standards,” Moscow: Tekhnosphere, 2005. 366 p.

[5] Dirac specifications [Electronic resource]. BBC – Electronic. Access mode: https://web.archive.org/web/20150503015104/http://diracvideo.org/download/specification/dirac-spec-latest.pdf.

[6] A. Grange, “VP9 Bitstream and Decoding process specification,” Electron. Text. Access mode: https://storage.googleapis.com/downloads.webmproject.org/docs/vp9/vp9-bitstream-specification-v0.6-20160331-draft.pdf.

[7] P. de Rivaz, D. Houghton, “AV1 Bitstream and Decoding process specification,” Electron. Text. Access mode: https://aomediacodec.github.io/av1-spec/av1-spec.pdf.

[8] Digital Cinema System Specification [Electronic resource]. Digital Cinema Initiatives – Electronic. Access mode: http://www.dcimovies.com/archives/spec_v1/DCI_Digital_ Cinema_System_Spec_v1.pdf.

[9] NDI Technical Brief [Electronic resource]. NewTek – Electronic. Access mode: https://233b1d13b450eb6b33b4-ac2a33202ef9b63045cbb3afca 178df8.ssl.cf1.rackcdn.com//pdf/newtek-ndi-technical-brief.pdf.

[10] I.V. Vlasyuk, V.Yu. Lyubetskaya, “Analysis of methods for suppressing ringing artifacts that appear on images during encoding with wavelet transform,” *T-Comm*. 2017. Vol. 11. No. 4, pp. 53-58.

[11] V. Balobanov, A. Balobanov, A. Potashnikov, I. Vlasuyk, “Low latency ONM video compression method for UAV control and communication,” *2018 Systems of Signals Generating and Processing in the Field of on Board Communications*. 2018, pp. 1-5.

[12] E. Belyaev and S. Forchhammer, "An Efficient Storage of Infrared Video of Drone Inspections via Iterative Aerial Map Construction," *IEEE Signal Processing Letters*, vol. 26, no. 8, pp. 1157-1161, Aug. 2019, doi: 10.1109/LSP.2019.2921250.

[13] A. Chowdhery and M. Chiang, "Model Predictive Compression for Drone Video Analytics," *2018 IEEE International Conference on Sensing, Communication and Networking (SECON Workshops)*, Hong Kong, China, 2018, pp. 1-5, doi: 10.1109/SECONW.2018.8396351.