Principles of Construction and Operation of Video Codecs on Wavelet Transforms

Mukhammedova Dierabegim Bakhtier kizi
Postgraduate student of the Department of Television and Broadcasting Systems, TUIT

A B S T R A C T
The article discusses the features of the principles of construction and operation of video codecs on wavelet transforms, their advantages, disadvantages and areas of application. It examines in detail the construction of the experimental open-source DIRAC video codec, which has become widespread in scientific research. It also discusses the main design features and technical characteristics of such varieties of the Dirac video codec as Dirac Pro, Schredinger and VC2 (SMPTE).

A R T I C L E I N F O
Article history:
Received 11 Aug 2024
Received in revised form
10 Aug 2024
Accepted 25 Sep 2024

Keywords: TV images, video stream, wavelet transform, wavelet video codec, video codec Dirac, Dirac Pro, Schredinger, VC2 (SMPTE).

Introduction.

Despite the fact that wavelet transforms were developed by Grossman and Morlet back in the 80s of the last century, their application for video coding systems is limited mainly to experimental studies [1, 2]. Basically, wavelet transforms are widely used to compress static images in the JPEG-2000 standard and video sequences in the MJPEG-2000 standard. But in video coding of broadcast and applied television, wavelet transforms are rarely used, and then only for processing reference frames [2]. This is mainly due to the fact that in television, the main compression of video data is provided by minimizing interframe differences based on various block motion compensation methods (Pic. 1), which provide a good compromise between the accuracy of motion compensation and the speed of the algorithms used [2, 3].

Pic. 1. General principle of block estimation and CD of video objects in adjacent frames.

At the same time, for more accurate compensation of video object motion, the block sizes in modern video codecs working in MPEG-4 AVC, H.264 and H.265 formats can vary from 4x4 to 32x32 pixels, which requires the use of short decorrelating transforms. Therefore, such codecs are built mainly on the use of a modified discrete cosine transform (MDCT) [4-6]. However, this transform, at high video data compression ratios, does not eliminate image distortions in the form of a block effect. These distortions are clearly visible after coding a test fine-grained image of a mountain waterfall (Pic. 2) with the XviD video codec, as shown in Pic. 3.

Pic. 2. Original image of a mountain waterfall.

Pic. 3. Image after encoding with XviD codec with a bitrate of 1 Mbit/s.

Pic. 4. Image after encoding with the H.264 codec with a bitrate of 1 Mbps.

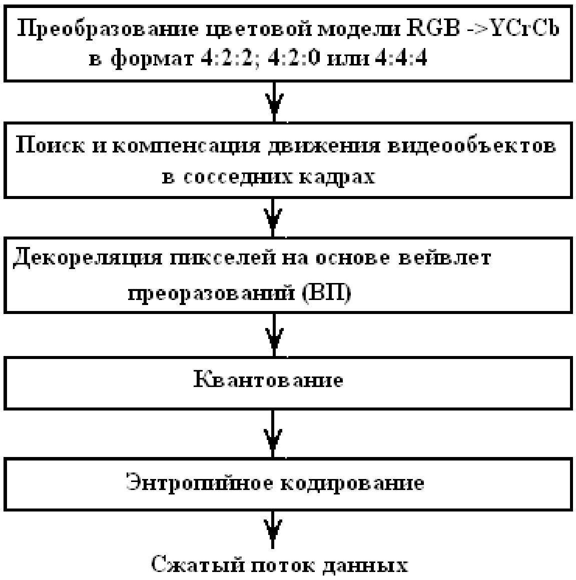
As can be seen from Pic. 3, the decoded image has a very noticeable mosaic structure consisting of DCT blocks. Therefore, to reduce the visibility of block distortions in the H.264 and H.265 codecs, special low-pass deblocking filters are used, which smooth out the brightness differences at the block boundaries. On the one hand, the use of deblocking filtering improves the visual quality of images, but on the other hand, averaging pixel values leads to a decrease in clarity and in some cases the images after deblocking filtering look blurry (Pic. 4). As for codecs on wavelet transforms (WT), they usually do not divide images into blocks, but process them entirely. For this purpose, interpolating non-harmonic wavelet functions are used, implemented in low- and high-pass wavelet filters. Such a mechanism better approximates the image, does not create a block effect and provides image quality 1.5-2 times better than with DCT [2, 7]. However, blockless image processing in wavelet filters does not go well with block methods of motion compensation, which reduces their efficiency by 2-3 times compared to MPEG standard codecs [2]. This is due to the fact that in interpolation prediction, the values of the analyzed pixel take on the values of several pixels to the left and right in horizontal prediction and above and below in vertical prediction. Within the image field, this condition is met, but there are no more boundary pixels at the image boundaries. Therefore, missing pixels at the image boundaries are usually replaced by the values of the boundary pixels. Within the frame, this replacement works fine, since there are no pixels outside the frame. But in block methods of motion compensation, prediction is made inside the pixel block. Accordingly, with small block sizes, the prediction error of pixel values increases greatly, which leads to a decrease in the efficiency of video compression and the visual quality of the displayed images. Nevertheless, research continues around the world on adapting relatively long wavelet functions to various motion compensation methods, which is reflected in the creation of experimental and commercial samples of video codecs. The most popular wavelet video codecs include Dirac, Dirac Pro, Schredinger, VC2 (SMPTE) and some others [1].

Main part.

One of the most interesting video codecs on wavelet transforms is the Dirac codec. The Dirac codec was developed by specialists from the research center of the British Broadcasting Corporation (BBC) with open source and named after the famous British mathematician and physicist Paul Dirac (1902...1984) of Swiss origin [1]. At the same time, the Dirac codec is a hybrid video codec with motion compensation according to the MPEG-2 standard, in which wavelet transforms are used not only for encoding reference frames, but also in interframe encoding. Moreover, this codec corresponds in efficiency to the H.264 standard, but has lower algorithmic complexity and does not require licensing costs [1]. The Dirac codec provides encoding of high-definition and ultra-high-definition television signals and has the following technical characteristics [1]:

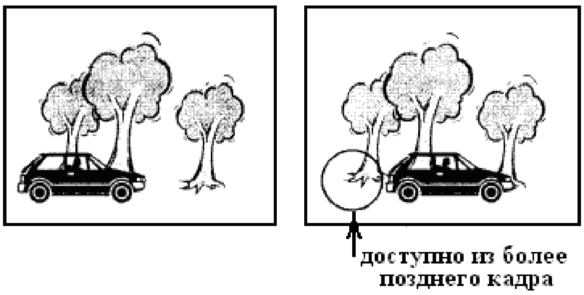
- supports a variety of television formats with resolutions from 176x144 (QCIF) to 4096x3112 (4K);
- supports the following color formats for representing digital video data 4:4:4/4:2:2 and 4:2:0;
- supports quantization bit depth of input video signals from 8 to 16 bits;
- supports processing of video with progressive and interlaced scanning using metadata;
- uses 32-bit frame numbering for both reference I frames and predicted L1(P) and L2 (B) frames;
- uses the following set of wavelet filters for image processing: Deslauriers-Dubuc(9.7), LeGall(5.3), Deslauriers-Dubuc(13.7), Haar, Daubechies(9.7);
- there is a set of 64 tables of values of wavelet coefficient quantizers;
- to ensure the best quality/volume ratio, an adaptive selection of a set of quantizers of the wavelet coefficients is used, which are calculated based on the Lagrange equation;
- the depth of the wavelet transform can vary from 2 to 5;
- coding in constant or variable bitrate modes is supported. In this case, the compression ratio in the variable bitrate mode is determined by the set value of the quality factor, which is set in the range (0.0-10.0).

The Dirac video codec is built on a typical video data processing scheme, the main stages of which are shown in Pic. 5. In this case, at the first stage of processing, the original RGB signals are converted into a luminance signal Y and color difference signals of red Cr and blue Cb, used in analog color television. This is due to the fact that, in accordance with the requirements of the international recommendation ITUR-601, digital television uses the YCrCb color model, which in the digital format 4:2:2 contains 1.5 times less video data than the original RGB signals.



Pic. 5. Main stages of video data processing in the Dirac video encoder.

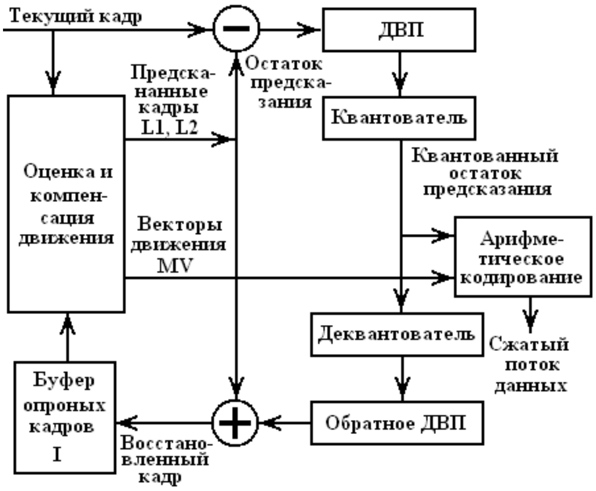
- Next, the motion of video objects in adjacent frames is estimated and compensated, which allows minimizing interframe differences and thereby significantly increasing the compression ratio of the video stream. In this case, the Dirac wavelet codec uses the principle of interframe coding with block motion compensation similar to the MPEG-2 standard, which consists of using three types of frames [2]:
- **I (Intracoded frame or simply Intra)** - "independent", which are coded without any connection with other video frames and act as reference frames, relative to which interframe differences are determined. These frames store the full amount of information about the image structure, so only intraframe coding is performed in them. In this case, the reference frames are the first frame of the video stream and the first frames of the change in the video plot. In addition, for reliable operation of the decoder, the reference frames can be frames following every 10-30 frames of the video sequence.
- **P (Predicted frame)** – direct prediction frames used in the inter-frame coding system with video object motion compensation and carrying information about changes in the image structure compared to the previous I or P frame. In the Dirac codec, P frames correspond to L1 frames (Level of the first level).
- **B (Bidirectional predicted frame)** – bidirectional prediction frames using the previous (I-frame or P-frame) and the next (I-frame or P-frame) to perform bidirectional analysis and motion compensation. The use of B-frames is due to the fact that when objects in the scene move, the background image also changes. Therefore, to transmit its missing sections, it is convenient to use data from a later frame, as shown in Pic. 6. In the Dirac codec, P frames correspond to L2 frames (Level of the second level).



Pic.6. Prediction of an image region from a later frame.

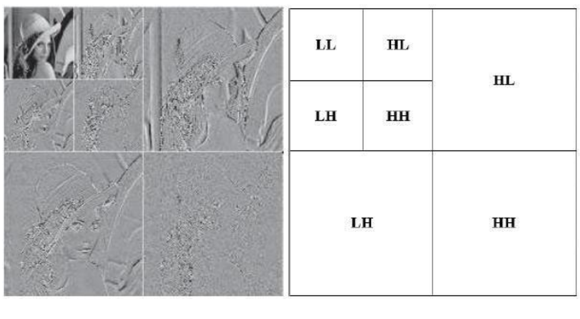
At the next stage of encoding, the video data of the reference and predicted frames undergo spectral processing based on two-dimensional wavelet transforms to eliminate inter-element statistical redundancy of images. Such processing allows eliminating the correlation between pixel values and recording them in a convenient form for more efficient operation of the video compressor.

Since in television the amount of redundant or predictable information strongly depends on the type of video plot, this leads to the fact that frames with large objects will be compressed well, and fine-grained images will be compressed poorly. Accordingly, the volumes of video data of such frames will differ significantly and this causes problems with their transmission over communication channels with constant bandwidth. Therefore, to control the amount of compression of the video stream, a mechanism for quantizing spectral coefficients is used, which is based on dividing the obtained coefficients by certain numbers with subsequent rounding of the remainder to the nearest integer values. This mechanism reduces the bit depth of the encoded numbers and accordingly improves the operating conditions of the video compressor, thereby increasing the amount of compression of video data. Well, at the final stage, a video compressor based on entropy coding comes into operation, which more economically represents quantized samples with a binary code of variable length. The essence of the variable-length code is to encode the most probable samples with short code words, and to encode less probable samples with longer code words, thereby improving the efficiency of video compression. Let us consider in more detail the principle of construction and operation of the Dirac video encoder according to the generalized structural family shown in Pic. 7 [8].



Pic. 7. Generalized structure of the Dirac encoder.

The frame sequence arriving at the encoder input, on the one hand, arrives at the input of the interframe processing block for evaluation and compensation of video object motion, and on the other hand, through the subtractor, passes to the discrete wavelet transform block. In this case, if the first frame of the video sequence arrives at the encoder input, then, with an empty reference frame buffer, this frame is declared reference and it passes to the DWT block without change. In the DWT block, the reference frames are processed entirely using the intraframe coding method in accordance with JPEG-2000. In this case, the Dirac coder has the ability to select one of 7 different wavelet filters for encoding, implemented according to the lifting scheme (Deslauriers-Dubuc (9.7), LeGall (5.3), Deslauriers-Dubuc (13.7), Haar, no shift per level, Haar, one shift per level, Fidelity wavelel, Daubechies (9.7)), which is set in the settings for encoding reference frames [9,10]. Moreover, Dirac uses a multi-level wavelet transform, where the transformation depth can be set from 2 to 5, which affects the amount of video data compression. Thus, an example of a two-level WP is shown in Pic. 8.

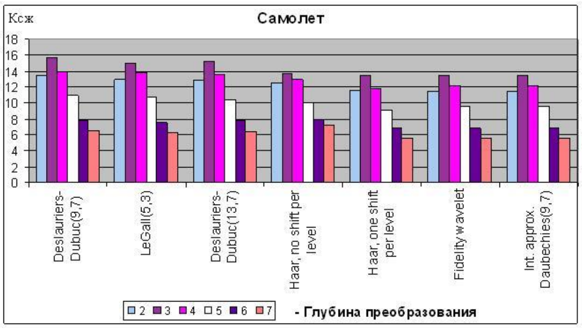


Pic. 8. An example of a two-dimensional two-level wavelet filter.

Thus, as a result of a two-dimensional wavelet filter, 4 sub-arrays of coefficients are formed from the pixel values of the original image:

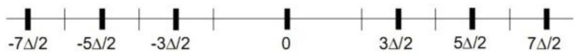
- **LL** - an array of high-frequency coefficients containing prediction errors of pixel values;
- **HH** - an array of low-frequency coefficients containing the sum of the original pixel values and prediction errors;
- **HL and LH** are combined arrays of low-frequency and high-frequency coefficients, respectively.

As the conducted studies have shown, the greatest compression of images is achieved with a three-fold wavelet filter, and the most effective wavelet filters are Deslauriers-Dubuc (9.7), LeGall (5.3), and Deslauriers-Dubuc (13.7), which give approximately the same results, as shown in the histograms (Pic. 8) [7].



Pic. 8. Comparative results of compression of the test image "Airplane" by different wavelet filters at different depths of the WP.

Then the obtained arrays of WP coefficients are sent to the quantizer block, where the compression value of video data is controlled by processing each array from low-frequency to high-frequency. Dirac uses the uniform quantization method with a dead zone, for which the zero zone is twice as large as this zone in the uniform quantization method, as shown in Pic. 9.



Pic.9. View of a uniform quantizer with a dead zone with a step Δ.

Using a dead zone allows converting small values of wavelet coefficients to zero levels, which is equivalent to a simple but effective noise suppression method. Moreover, to improve the efficiency of the compressor, the Dirac codec uses a quantization step close to the optimal value from a set of 96 tabular values that are selected from the results of calculating the Lagrange optimization equation. [10]. For this purpose, Dirac uses the RDO (Rate Distortion Optimization) method for selecting quantizers, balancing between the Lagrange coefficients and the transmission rate, using measurements of quantization errors and direct counts of the symbols of quantized values to calculate the entropy. The higher the value of the Lagrange multiplier λ, the lower the resulting data transfer rate and vice versa. Accordingly, for each sub-range, its own Lagrange parameter is set. In this case, lambda is initially calculated using the quality factor (QF) set in the encoder settings. The calculation of the initial Lagrange coefficients in the codec for reference (I) and predicted frames (L1) and (L2) is performed according to the following formulas:

$$I\lambda = \frac{1}{16}1^{\frac{(10-QF)}{2,5}} \quad (1)$$

$$L1\lambda = 32 * I\lambda \quad (2)$$

$$L2\lambda = 256 * I\lambda \quad (3)$$

These lambda variables are used to select quantizers in the reference and predicted frame images I, L1, and L2. From them, lambdas are derived for motion estimation:

$$L1\_me\lambda = 2.0 * \sqrt{L1\lambda} \quad (4)$$

$$L2\_me\lambda = L1\_me\lambda \quad (5)$$

Moreover, when selecting the "Lossless" encoding mode, all lambda variables are set to 0 [10].

The selection of quantizers in the Dirac video codec consists of three main stages:

1. Calculating prediction errors
2. Calculating the Lagrange coefficients
3. Selecting the best quantizer.

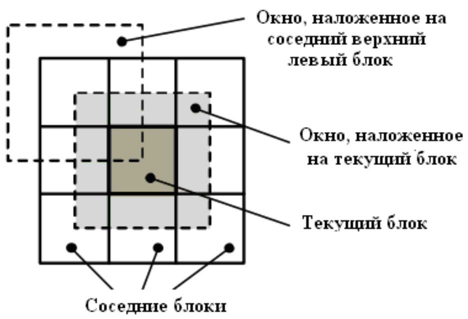
After the quantizer, the quantized array of WP coefficients, on one side, goes to the decoding part, which contains a dequantizer and an inverse wavelet transform block, which are part of the interframe processing system based on5 motion compensation. And on the other side, it goes to the arithmetic coding block, which is the output compressor of video data, after which the compressed digital video stream goes either to the transmission channel or is saved as a file.

The Dirac coder uses a more efficient type of statistical coding than the Huffman coder, called arithmetic coding and based on coding not individual symbols, but their sequences. Arithmetic coding is characterized by greater algorithmic complexity and, accordingly, lower performance, but it provides higher video data compression values.

If the encoder input does not receive reference frames, then the mechanism of interframe coding with video object motion compensation is activated. Motion estimation and compensation units and a reference frame buffer are used for this purpose.

As noted earlier, the main problem with using WP in video coding is that long wavelet functions do not work well with block methods of motion compensation. As a result, due to high-frequency "bursts" of wavelet spectrum coefficients corresponding to block boundaries, the efficiency of video coding decreases. Therefore, the Dirac codec uses overlapped block motion compensation (OBMC), which is a compromise between image quality and coding efficiency. At the same time, it ensures less noticeable block distortions in the image, and also allows obtaining more accurate displacement vectors when estimating the motion (MD) of video objects [10].

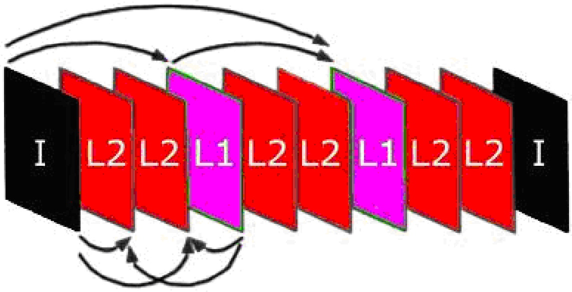
The essence of this method is that block artifacts can be partially smoothed by interpolating predicted brightness values based on the motion vectors of neighboring blocks. In this case, before a block from the reference frame is moved to a new location in the predicted frame, a weight "window" is imposed on it, overlapping eight neighboring blocks (Pic. 10). Thus, the prediction of the pixel value from the current block will depend on the brightness values of four pixels, i.e. from the current and three neighboring blocks [11].



Pic.10. Using a weight window to smooth out block distortions.

Using block overlap when constructing a predicted frame allows eliminating sharp differences in brightness values at block boundaries. However, at low bitrates, some reduction in the clarity of the reconstructed images may be observed.

The encoder works with standard group of pictures (GOP) modes, according to which the number of frames between L1 can vary depending on the encoding mode. The prediction method for encoding a frame using the standard GOP structure is shown in Pic.11. In this Picure, the number of L1 frames between I frames is 2, and the distance between L1 frames is 3 [12]. In this case, motion compensation is based on assessing the motion of video objects in adjacent frames of the video stream.



Pic.11. Prediction of L1 and L2 frames.

Motion estimation of the encoding device is the most complex part of the encoding system and requires huge computing resources. Thus, the Dirac video codec for encoding and decoding uncompressed video uses a 3-stage approach:

- At the first stage, motion vectors are detected for each block and in each block, links are determined with pixel accuracy using a hierarchical (multi-level) motion estimation method.
- At the second stage, these vectors are refined with sub-pixel accuracy.
- At the third stage, a calculation method is selected that combines motion vectors and blocks with similar motion [13].

In this case, the OD turns out to be the most accurate only with the participation of all three components, however, from the point of view of calculations, the use of all three components requires huge costs for computing resources and is not very suitable for working in real time. Therefore, in practice, the hierarchical OD method is often used, which significantly speeds up the work. In this case, the encoded and base frames are first reduced by half in height and width, then the resulting images are reduced again, and so on until the desired result is obtained. The process is repeated L times, where L is the algorithm parameter. As a result, L+1 hierarchy levels are formed. At each stage of the hierarchy, vectors from lower levels are used as directions for searching in higher levels. This dramatically reduces the number of steps in the search process for a large flow. At the same time, the block size remains constant, so at each level there are only a quarter of the blocks, which are distributed so that each block corresponds to four blocks. Thus, each block provides a vector with a guide of four blocks at the next higher level. Moreover, each check for block compliance is performed against the background of a search in a small range around the guide vector. For this purpose, the RDO (Remote Data Objects) metric system is often used.

After executing the hierarchical motion estimation algorithm, a quadratic partition of the image into variable-sized blocks is performed. In this case, the partition is performed independently for each segment. Accordingly, the hierarchical approach significantly reduces the computational costs used in motion estimation for an equivalent search range. However, there is a risk of missing small movements that prevent the correct decision when there are many movements close to each other. Therefore, to reduce the likelihood of this event, the Dirac encoder always uses a zero vector (0,0), which allows tracking both slow and fast video objects. In addition, already found motion vectors in neighboring blocks can also be used as direction vectors if they have not already been involved in the current action. Since each layer has two horizontal and vertical resolutions, the search is performed in the area of +/- 1 pixels of direction vectors. In fact, the search ranges are always larger than the limited area, since this can cause a systematic interruption of motion estimation in a local minimum.

The workflow of the hierarchical motion estimation algorithm based on quadratic partitioning is based on dividing the image into large square blocks (e.g. 128×128 pixels), which are called segments. Then each segment, in turn, can be divided into four square blocks, and so on [4, 13]. In this case, the algorithm can be used both in the motion estimation mode with an already known quantization step, and in the motion estimation mode with a limitation.

First, the mode of using motion estimation with an already known quantization step is described. For this, a “fast” algorithm for hierarchical motion estimation is first described, with the help of which a variety of segment encoding options are formed: an encoding option with one motion vector, with four vectors, and so on.

Then, a recursive algorithm for dividing an L1 frame into variable-size blocks is described, which minimizes the bit costs of the hierarchical tree, provided that the bit costs of the difference blocks are minimal [10, 14].

After this, the procedure for encoding the motion vectors of the P-frame is described and it is shown that with this encoding method the recursive partitioning algorithm also minimizes the bit costs of the motion vectors, provided that the bit costs of the difference blocks are minimal in the set of segment encoding options. Then the features of using this algorithm for L2-frames are described.

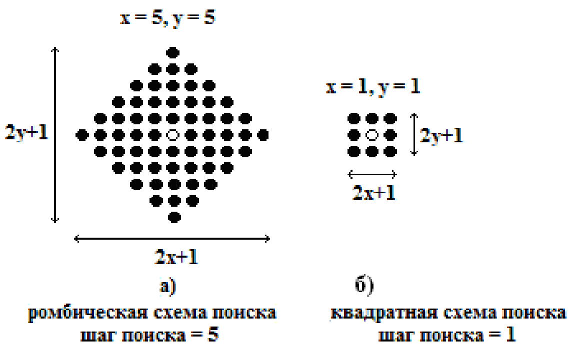
The peculiarity of L2-frame processing is that three variants of partitioning are formed for each segment. The difference between these three variants is in the method of forming difference macroblocks:

- in the first variant, difference blocks are formed using the previous base frame;
- in the second variant, difference blocks are formed using the next base frame;
- in the third variant, difference blocks are formed using the previous and next base frames.

Then, the best of the three partitioning variants is selected, providing the minimum bit costs for difference macroblocks. In this case, the first step in the hierarchical motion estimation in Dirac is a 12-level down conversion (reducing the size by n times) of internal (inter) frames (both P and B). The number of levels depends on the image format (linear dimensions - height and width) and can be calculated using formula (6) [41]. Reception, depending on the video format used - CIF or HD (1920x1080) the number of size conversion levels can be 4 or 6 respectively. In this case, during this conversion, the width and height of the frames are reduced by 2 times at each conversion level.

$$level = \left[\min\left(\log_2\left(\frac{width}{12}\right), \log_2\left(\frac{height}{12}\right)\right)\right] \quad (6)$$

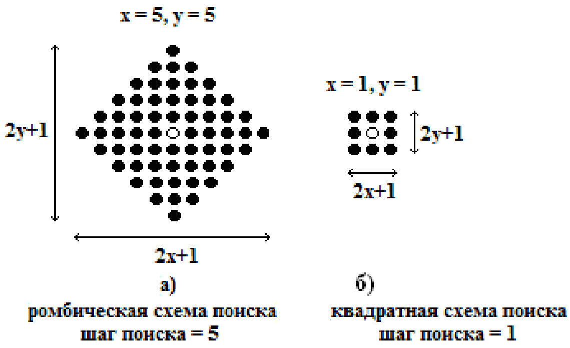
Motion estimation starts from the lowest resolution (smallest frame sizes) and proceeds sequentially until it is performed for frames of the original size. The Dirac codec uses a rhombic and square motion estimation scheme (Pic. 12)



Pic. 12. Search schemes in the Dirac video codec.

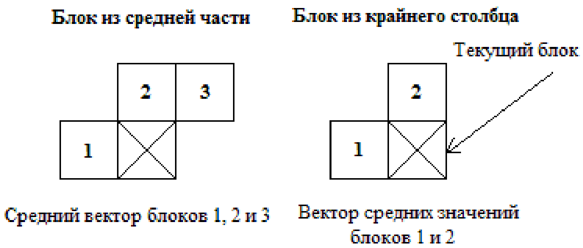
Accordingly, at the lower level of the transformation, a diamond shape is used for the area in which the search is performed with a step of [10]. At all other levels, a square shape of the search area with a step of 1 is used.

Pic. 13 shows both forms of the search area, where a diamond shape with 61 search points is used for the lower level, and a square shape with 9 points is used for the other levels [10, 15]. At the beginning of the search, candidate lists for the search are generated. This list consists of several points that are selected according to a certain scheme (Pic. 13) and are centered relative to the predicted motion vector. The predicted motion vector can be **zero, predicted in space, or a direction motion vector.**



Pic.13. Search schemes in the Dirac video codec.

The predicted motion vector in space is the average vector of blocks 1, 2 and 3 or the vector of average values of blocks 1 and 2, as shown in Pic.14, depending on the position of the current block being processed.



Pic. 14. Spatial prediction of the motion vector in Dirac.

The direction vector is a motion vector that best reflects the position of the corresponding block at the nearest lower level of the hierarchy and is not available for the lowest level.

Pic. 15 shows a 4-level motion estimation used for the CIF format in the experimental video codec Dirac [16, 17].

As can be seen from Pic. 15, two candidate lists are formed for the lower level - for zero and predicted vectors in space using a rhombic search scheme. In this case, the similarity metric between image blocks SAD (Sum of the Absolute Difference) is used for motion estimation, which is used due to the low algorithmic complexity and high speed of the algorithm [14]. In addition, there is a possibility of parallelizing the SAD finding for different blocks, since each block is processed separately. Accordingly, after finding the candidate blocks, the final part of the motion estimation occurs, for which more accurate algorithms are usually used that better take into account the features of human perception.

Pic.15. 4-level motion estimation for the CIF format in the Dirac video codec.

At the initial stage of the search, the SAD is calculated only for the central point of the rhombic scheme in each of the candidate lists and a list is found for which the SAD value is minimal. The candidate lists for the search are selected as follows: the minimal SAD value is multiplied by 1.5 and all candidate lists for which the SAD is less than this value are selected. Thus, there can be a maximum of 2 candidate lists and 122 search points that can be involved in the search at the lower level if there are no intersections between the two lists. Next, the SAD is calculated for all selected candidate lists at the corresponding search points and the coordinates of the search point that gives the smallest SAD value are taken as the best motion vector. In this case, the search procedure for the following levels remains the same, except for adding another candidate list, which is centered relative to the direction vector. Thus, for the square search scheme (Pic. 14), three candidate lists are used, and the maximum number of search points is 27 if the lists do not intersect.

After passing all these levels, the accuracy of motion vectors of one pixel is achieved. Dirac allows achieving an accuracy of 1/8 pixel. For this, motion estimation occurs at the subpixel level, where the current and reference images are doubled, and the motion vector for pixel accuracy is multiplied by 2, then a search is performed to achieve the accuracy of the motion vector of 1/2 pixel. The presented procedure is repeated until the required accuracy is achieved.

After achieving the required accuracy of motion vectors for each of the blocks, the last stage of motion estimation begins - mode decision. OD is carried out using the RDO motion estimation metric [15], which consists of the basic block coincidence metric (SAD) and some smoothness constant of the motion vectors. The smoothness value is based on the difference between the candidate motion vector and the median of neighboring, previously calculated motion vectors. The complete metric is a combination of these two metrics. It specifies a vector V with components Vx and Vy in Cartesian coordinates that links a block P from the current frame to a block R from the reference frame by the given relation R = V(P) [10]:

$SAD(P,R)+\lambda_{ME} \times \max(|V_x - M_x|) + |V_y - M_y|$, (7)

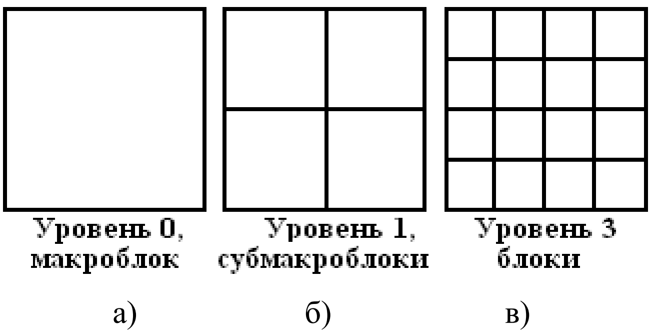
Where, M_x and M_y – the corresponding components along the x and y axes of the average motion vector calculated from the motion vectors of the left, upper and upper-left blocks;

λ_{ME} - weighted Lagrange multiplier (λ). Dirac uses the Quality Factor (QF) parameter to control the quality of encoded frames.

QF plays an important role, as it is used in the RDO process for motion estimation and quantization as a Lagrange multiplier (λ). The relationship between λ and QF is given by the following expression [15]:

$\lambda = (10^{(10-QF)/2,5})/16$ (8)

- The Dirac video codec provides 12 motion compensation modes, which include 3 levels of macroblock division and 4 prediction modes. Moreover, a macroblock is represented by an array of 4x4 blocks. There are 3 ways to divide a macroblock (Pic. 16) [17, 18]:
- Split level 0: no division, in which the macroblock is described by one motion vector from the reference frames (Pic. 16, a).
- Split level 1: uses division of the macroblock into 4 sub-macroblocks, each of which is an array of 2x2 blocks and is described by 1 motion vector from the reference frames per 1 sub-macroblock (Pic. 16, b).
- Split level 2: uses division of the macroblock into 16 blocks (Pic. 16, c).



Pic.16. Macroblock division levels.

At the same time, during coding, the best prediction mode is selected for each block (block, sub-macroblock or macroblock):

➢ INTRA: coded based on the I frame;

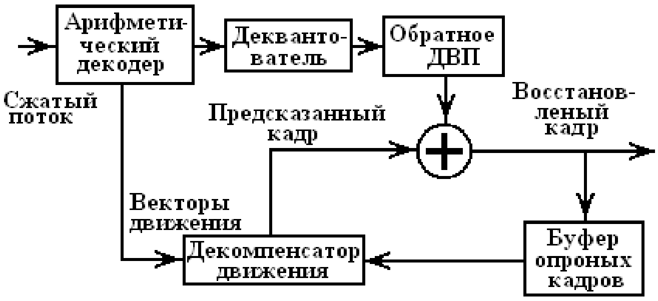
➢ REF1_ONLY: coded based on the first reference frame;

➢ REF2_ONLY: coded based on the second reference frame, if it exists;

➢ REFIAND2: two-way, bi-directional prediction.

Thus, depending on the macroblock division level and the prediction mode, 12 combinations are possible, the choice of which is influenced by equation (6).

The Dirac codec can work with images with a resolution from QCIF (176x144) to HDTV (1920x1080), which allows it to compete with other modern video codecs. At the same time, the implemented algorithms of OD in Dirac, based on the hierarchical method, accelerate the algorithm and do not require significant costs for computational processes. However, this method has the risk of missing small movements, which prevents the correct decision when there are many movements next to each other. Therefore, Dirac always uses the zero vector (0,0) as a similar direction vector, which allows tracking both slow and fast objects. In addition, Dirac uses a multi-level OD system, which, in combination with CD based on block overlapping, provides it with fairly good image quality indicators when compressed 70 times or more, which allows it to compete with professional video codecs. Accordingly, the restoration of encoded images in the decoder is performed in the reverse direction of encoding, as shown in Pic. 17.



Pic.17. Generalized structure of the Dirac decoder.

The second improved implementation of the Dirac codec is the "Schrödinger" project, named after the Austrian physicist Erwin Schrödinger. It was a joint development of the BBC and Fluendo. The goal of this project was significant algorithmic optimization to ensure encoding of high-definition television images in real time on laptops with two core processors. In February 2008, the final version of this codec was released. Since the Schrödinger codec is based on the Dirac codec, its operating principle and main technical characteristics correspond to the Dirac codec [1]. Another type of the Dirac codec is the Dirac Pro codec (Pic. 18), which is a commercial version of the Dirac video codec.

Pic. 18. The appearance of the Dirac Pro video codec.

The professional version of the Dirac Pro codec was released in September 2008, which is designed for use in editing, archival and broadcasting systems, where there are increased requirements for the quality of compressed video. The compression algorithm has an open source code and is distributed under a Royalty Free license. At the same time, the hardware version of the Dirac Mezzanine Compression codec was created by Numedia Technology, which is one of the BBC's partners in the implementation of projects. However, this version of the codec does not use a motion compensation mechanism and all frames of the video stream are processed only by intra-frame coding methods [1].

According to the specification, the codec supports video formats from QSIF525 (176 × 120 pixels) to the latest UHDTV 8K-50 (7680 × 4320 pixels, 50 frames/s) and digital cinema formats 2K and 4K D-

Cinema. The codec has the following characteristics:

- frame rate - 23.97…60 Hz
- sampling - 4:2:0, 4:2:2, 4:4:4 and RGB
- bit depth - 8, 10, 12 (up to 16)
- support for interlaced and progressive scanning.

It should be noted that the Dirac Pro codec was submitted to SMPTE for standardization as VC-2 Codec. In 2010, SMPTE adopted VC-2 as a video compression standard [1].

- SMPTE 2042-1-2009 VC-2 Video Compression;
- SMPTE 2042-2-2009 VC-2 Level Definitions;
- RP (Recommended Practices) 2047-1-2009 — VC-2 Mezzanine Level Compression of 1080P High Definition Video Sources;
- SMPTE 2047-2-2010 Carriage of VC-2 Compressed Video over HD-SDI — Transmission of VC-2 Compressed Video over HD-SDI Interface;

RP 2042-3-2010 VC-2 Conformance Specification.

Conclusion.

As follows from the conducted analytical review of video codecs on wavelet transforms, most of them are based on the experimental video codec Dirac which in its operational and some quality indicators is comparable with the codecs of the H.264 standard. And in some quality indicators it even surpasses it. However, to date, the development of the Dirac, Schrödinger and Dirac Pro projects has been discontinued. This is due to the fact that in the world there is a huge variety of encoding television devices oriented to the use of discrete cosine transform. Accordingly, in broadcast television, it is economically inexpedient to ensure the compatibility of two different technologies on DCT and WP. Nevertheless, codecs on wavelet transforms can find application in volumetric and applied television. Therefore, interest in the development of more advanced video codecs on wavelet transforms actively continues, as evidenced by the results of published works and conducted research, for example [19-21].

List of used literatures:

1. Питер Уилсон, Тим Борер, Томас Дэвид. Семейство систем цифрового сжатия DIRAC расширяется. Журнал 625. №3, 2007.
2. И.А.Гаврилов, Т.Г.Рахимов, А.Н.Пузий, Х.Х.Носиров, Ш.М.Кадиров “Цифровое телевидение” Ташкент, 2016 г., 380 с. с ил.
3. Гришин С.В., Ватолин Д.С., Лукин А.С., Путилин С.Ю., Стрельников К.Н. Обзор блочных методов оценки движения в цифровых видео сигналах.// Программные системы и инструменты. 2008. №9 С.50-62.
4. Ричардсон Ян. Видеокодирование. Н. 264 и MPEG-4 – стандарты нового поколения. – М.: Техносфера, 2005г. 368 с.
5. Дворкович А. В. Эффективное кодирование видеоинформации в новом стандарте H.264/AVC. https://docplayer.ru/43232231-Effektivnoe-kodirovanie-videoinformacii-v-novom-standarte-h-264-avc.html
6. Кодек H.265 и H.265+: особенности работы. https://hikvision.org.ua/ru/articles/kodek-h265-i-h265-osobennosti-raboty
7. Гаврилов И.А, Отто С.Э., Бобобекова Д.Б., Пузий А.Н. Влияние выбора вейвлет фильтра на эффективность сжатия изображений. Статья в журнале Вестник ТУИТ №1-2 2013 г с. 37-44.
8. Дам Чонг Нам. Оптимизация метода квантования для вейвлет-видеокодека.// Цифровая Обработка Сигналов №3, 2018. с.38-43.

9. Open source video codec Dirac // https://sourceforge.net/projects/dirac
10. Dirac Specification Version 2.1.0 Issued: January 24, 2008. PDF. 123 c
11. Коплович Е. А. Алгоритм оценки смещения видеокадра на основе выбора характерных блоков с использованием коэффициентов ДКП// Интеллектуальные системы в производстве. — 2006. — №2(8).—С. 25–30.
12. Myo Tun. Research and Developments of Dirac Video Codec // A thesis submitted for the degree of Doctor of Philosophy. September 2008 - page 28
13. BBC Research and Development., "Dirac Fundamentals." http://dirac.sourceforge.net/documentation/algorithm/algorithm/toc.htm.
14. SAD, https://ru.wikipedia.org/wiki/Sum_of_the_Absolute_ Difference
15. RDO, https://en.wikipedia.org/wiki/Rate_distortion_ optimization
16. Рыбаков Г., Суслов А. JPEG-2000. http://rain.ifmo.ru/cat/view. php/theory/data- compression/ jpeg-2006.
17. Kamilov Mirzayart, Nosirov Khabibullo. Brightness transformation of image with adaptive blocks, research of its efficiency in compression and estimation of reconstructed image quality. Статья в научном журнале «European Sciences review» Scientific journal № 1-2 2017 (January-February). c.189-192.
18. Гаврилов И.А., Пузий А.Н. Вейвлет видеокодек с большой величиной сжатия видеопотока. Статья в сборнике докладов Республиканской научно-технической конференции «проблемы информационных и Телекоммуникационных технологий» проходившей 12-13 марта 2015 года в г. Ташкенте ТОМ III с. с.389-392.
19. Шаронов Игорь Олегович Видеокомпрессия на основе дискретного вейвлет-преобразования и блочной компенсации движения. Автореферат диссертации по специальности 05.12.13.2018 г.
20. Дам Чонг Нам. Исследование и разработка методов сжатия подвижных изображений с использованием расширенных вейвлет-разложений. Автореферат диссертации по специальности 05.12.13.2019 г.
21. Быстров Кирилл Сергеевич Исследование и разработка видео кодека на основе многоканального вейвлет преобразования. Автореферат диссертации по специальности 05.12.13. 2020 г.