UDC 681.327

Victor Porev

FEATURES OF RUN-LENGTH ENCODING FOR TRUE COLOR RASTER IMAGE FORMAT

The paper considers the approach to improving the compression method based on run-length encoding. Describes a modification of the compression scheme RLE-БП to provide the ability to compress lossless bitmaps of True Color format. The main goal is to achieve a high decompression rate with a competitive degree of compression due to the variability of the construction of code sequences according to the RLE-БП.

Key words: compression, lossless, run length encoding, RLE-БП.

Fig.: 1. Bibl.: 8.

Relevance of the research topic. An important role in the digital era is played by the search and implementation of effective information coding methods, in particular data compression methods.

Formulation of the problem. For many applications, there are requirements to provide the highest possible degree of compression for data storage and to obtain the fastest possible decompression when reading such data. To a large extent, such requirements are contradictory, so it is necessary to find some compromise. The balance between the degree of compression and the speed of decompression can change as a result of improving the encoding of information.

Actual scientific researches and issues analysis. Research on information compression has been going on for many decades. One of the first known compression methods is the Huffman method, according to which the most popular symbols are encoded with a shorter prefix code, and the less popular ones with a long one [1]. Also, the run length encoding (RLE) method was proposed quite a long time ago, according to which chains of the same symbols are coded by the number of repetitions-character code pair [2]. For text compression, this encoding is ineffective, but the RLE method proved to be useful for encoding bitmap images with a limited number of colors - no more than 256. Later, dictionary LZ encoding methods [3, 4] were used to compress bitmap images. These methods have a significantly higher degree of compression compared to RLE and have been used for such graphic formats as GIF, PNG. The above methods belong to the category of lossless compression - encoding does not change the original data in any way. Such methods are mainly

focused on the 256 color format, although PNG allows you to store True Color images as well.

In general, the True Color format, such as 24 bits per pixel, is very commonly used for photo-type images. But compression methods, even as effective as dictionary LZ, are not used for digital photos, so in many cases they do not provide any compression. Digital photographs use lossy compression techniques such as JPEG [5] or variations of wavelet encoding such as JPEG2000 [6]. Lossy compression methods make it possible to compress the image of digital photos significantly - tens of times - without a noticeable deterioration in the perception of the image by a person. However, such methods cannot be used for some applications where complete preservation of the original data is required.

Among the lossless compression methods, dictionary LZ-like methods seemed to have the absolute advantage due to the highest degree of compression of repetitive data. In any case, compared to the Huffman and RLE methods. But it is not quite so. In particular, the RLE method provides a much higher operating speed and does not require additional memory for the dictionary. In addition, RLE is generally convenient to use for encoding individual rows (or columns) of a raster, and dictionary LZ-like methods lose their effectiveness here, since the contents of the dictionary are created based on all the data. Another advantage of RLE is the possibility of parallel (or multithreaded) encoding and decoding of different raster fragments.

Therefore, the RLE method was chosen as the basis for solving tasks related to providing fast direct access to individual fragments of large rasters. Around 2004, to support the geoinformation system, methods were invented to modify the RLE method to significantly increase the degree of compression while maintaining the speed of decompression and providing direct access to raster fragments. These methods were named RLE-БП according the first letters of the authors' Ukrainian surnames - Блінова, Порєв. Later, in 2008, basic information about RLE-БП was published in particular in [7] and some other publications.

Tatyana Blinova

Victor Porev

Fig. 1. The authors of RLE-БП

The main idea of RLE-БП consists, firstly, in improving and significantly complicating code sequences, and secondly, that for each line of the raster, the software coder looks for such values of the code parameters that provide the minimum total amount of code for each line. Thanks to this, it was possible to significantly increase the degree of compression compared to simple RLE and in many cases for images of business graphics to approach the level of compression, for example, the LZW method [6, 8].

Uninvestigated parts of general matters defining. The question of encoding chains of repeating pixels in the True Color color format is insufficiently researched.

Setting objectives. The main tasks are to invent such methods of encoding chains of repeating pixels that are capable of providing high decompression speed and organizing direct access to compressed data in the form of True Color format images.

The statement of basic materials. To support True Color bitmap compression capabilities, some encoding variants have been added to the RLE-BP set. These variations for the 24-bit format are named methods 2-4 and are described as follows.

Encoding method 2. Three types of code sequences are used:

0 c..c - single pixel of any color (c..c) - 1+24 bits
10 n..n m..m - main color pixel chain (m..m) - 2+N1+M bits
n..n - code of N1 bits to represent (n-1), where n is the length of the string
m..m - M bits code of the main color
11 n..n c..c - chain of pixels of any color (c..c) - 2+N2+24 bits
n..n - code of N2 bits to represent (n-2), where n is the length of the string
where
M = 1...8 – number of bits to represent the main color index
N1 = 0...10 - the number of bits to represent the length of the main color chain
N2 = 0...5 - the number of bits to represent the length of a string of any color

Encoding method 3. Three types of code sequences are used:

0 m..m - single pixel of the main color (m..m) - 1+M bits
10 n..n m..m - main color pixel chain (m..m) - 2+N1+M bits
n..n - code of N1 bits to represent (n-2), where n is the length of the string
m..m - M bits code of the main color
11 n..n c..c - a single pixel or a chain of pixels of any color (c..c) - 2+N2+24 bits
n..n - code of N2 bits to represent (n-1), where n is the length of the string

where

M = 1...8 – number of bits to represent the main color index

N1 = 0...10 - the number of bits to represent the length of the main color chain

N2 = 0...5 - the number of bits to represent the length of a string of any color

Encoding method 4. Two types of code sequences are used:

0 n..n m..m - single pixel or chain of pixels of the main color (m..m) - 1+N1+M bits

n..n - code of N1 bits for the number of (n+1) pixels

m..m - M bits code of the main color

1 n..n c..c - a single pixel or a chain of pixels of any color (c..c) - 1+N2+24 bits

n..n - code of N2 bits for the number of (n+1) pixels

where

M = 1...8 – number of bits to represent the main color index

N1 = 0...10 - the number of bits to represent the length of the main color chain

N2 = 0...5 - the number of bits to represent the length of a string of any color

These methods are used in the software of the geographic information system and in applications for distance learning.

Conclusions. The modified RLE-БП run length encoding method, which allows to achieve a competitive compression ratio at a high decompression speed in the mode of direct access to raster fragments, is highlighted. A modification of the RLE method for compression of images in True Color format is proposed.

References

1. D.A. Huffman. A Method for the Construction of Minimum-Redundancy Codes // Proceedings of the IRE. 40 (1952): 1098–1101. doi:10.1109/JRPROC.1952. 273898.
2. S.W. Golomb. Run-Length Encodings // IEEE Trans. Information Theory, 12:3 (1966) pp. 399-401.
3. Ziv J., Lempel A. Compression of Individual Sequences via Variable-Rate Coding // IEEE Trans. Inform. Theory, 1978, V. 24 (5), pp. 530–536
4. Welch T. A Technique for High-Performance Data Compression // Computer, 1984, V. 17 (6), pp. 8–-19.
5. Gregory K. Wallace. The JPEG Still Picture Compression Standard // IEEE Transactions on Consumer Electronics. Vol. 38, No 1, Feb 1992, pp. xviii-xxxiv

6. High Throughput JPEG 2000 (HTJ2K) and the JPH file format: a primer // ISO/IEC JTC 1/SC 29/WG1 | Document N87018. URL: http://ds.jpeg.org/whitepapers/jpeg-htj2k-whitepaper.pdf
7. Blinova T., Porev V. Some Methods Of The Raster Encoding In Geographic Information Systems // Proc. int. conf. CODATA`21, Kyiv, 2008. – p.153.
8. Victor Poriev. Improving the method of run length encoding // Proceeding of International Conference on Security, Fault Tolerance, Intelligence" (ICSFTI2019), 14-15 may 2019, Kyiv .- pp.165-169.

AUTHORS

Victor Porev – associate professor, Department of Computer Engineering, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute".
E-mail: v_porev@ukr.net

EXTENDED SUMMARY

Victor Porev

FEATURES OF RUN-LENGTH ENCODING FOR TRUE COLOR RASTER IMAGE FORMAT

Relevance of research topic. An important role in the digital era is played by the search and implementation of effective information coding methods, in particular data compression methods.

Formulation of the problem. For many applications, there are requirements to provide the highest possible degree of compression for data storage and to obtain the fastest possible decompression when reading such data. To a large extent, such requirements are contradictory, so it is necessary to find some compromise. The balance between the degree of compression and the speed of decompression can change as a result of improving the encoding of information.

Analysis of recent research and publications. Research on information compression has been going on for many decades. The RLE method was chosen as the basis for solving tasks related to providing fast direct access to individual fragments of large rasters. At one time, methods of modifying the RLE method were invented to significantly increase the degree of compression while maintaining the speed of decompression and providing direct access to raster fragments. These methods were called RLE-БП.

Selection of unexplored parts of the general problem. The question of encoding chains of repeating pixels in the True Color color format is insufficiently researched.

Setting objectives. The main tasks are to invent such methods of encoding chains of repeating pixels that are capable of providing high decompression speed and organizing direct access to compressed data in the form of True Color format images.

Presentation of the main material. To support True Color bitmap compression capabilities, some encoding variants have been added to the RLE-БП set, which are designated as methods 2-4 for the 24-bit format.

These methods are used in the software of the geographic information system and in applications for distance learning.

Conclusions. The modified RLE-БП run length encoding method, which allows to achieve a competitive compression ratio at a high decompression speed in the mode of direct access to raster fragments, is highlighted. A modification of the RLE method for compression of images in True Color format is proposed.

Key words: compression, lossless, run length encoding, RLE-БП.