Method for Encoding Video Frame Fragments Based on Non-Equilibrium Codes with Minimization of Service Data

Yuriy Ryabukha [1[0000-0001-9821-598x]] and Dmitry Barannik [2[0000-0003-4235-300X]],
Vitaly Tverdokhlib [2[0000-0002-2996-9523]] and Valeriy Barannik [2[0000-0003-3516-5553]],
Oleg Kulitsa [3[0000-0003-2589-6520]] and Emil Faure [4[0000-0002-2046-481X]],
Roman Odarchenko[5,6,7 [0000-0002-7130-1375]]

[1] Kharkiv National Air Force University, Kharkiv, Ukraine
vvbar.off@gmail.com
[2] Kharkiv National University of Radio Electronics, Kharkiv, Ukraine
valera462000@gmail.com
[3]National University of Civil Defence of Ukraine, Cherkasy, Ukraine
okulitsa@gmail.com
[4]Cherkasy State Technological University, Cherkasy, Ukraine
faureemil@gmail.com
[5] National Aviation University, Kyiv, Ukraine
[6] State Scientific and Research Institute of Cybersecurity Technologies and Information Protection, Kyiv, Ukraine
[7] Yessenov University, Aktau, Kazakhstan
odarchenko.r.s@ukr.net

Abstract The principle of constructing a method for encoding a video stream, aimed at reducing the structural redundancy of fragments of video frames, is considered. Moreover, the developed method allows us to consider fragments of a video frame in the form of separate code structures

Keywords: structural redundancy, video processing, telecommunication systems, data protection.

1 Introduction

Modern information and communication networks are characterized by a constant increase in the intensity of the transmitted traffic as well as its constant complication.

The largest part of this traffic is video service data, regarding which there is a constant increase in quality requirements.

Simultaneously with the growth of traffic volumes, its complication is observed, caused by the development of existing and the emergence of new types of network services.

These include the development of video communications services, virtual reality. New types of services include concepts such as Smart City, Safe City, holographic video communication systems, and UHD television.

It is clear that in addition to increasing the volume of traffic both video and other types, its complication arises due to the next reasons:

1. High dynamic range of packet sizes to be processed on network nodes;
2. A large number of packages of small length generated within the Smart City or the Safe City by sources such as:
   - sensors (data collection from various systems);
   - robotic systems (remote control commands).
3. Using firewalls and cyber protection modules at the level of network nodes, which creates an additional delay.
4. Using hidden information transmission systems (in particular, based on steganographic algorithms), which increases the load on the nodes and creates an additional transmission delay due to the need for additional traffic analysis.
5. Using intelligent video traffic processing and analysis systems.

At the same time, the increase in network bandwidth lags behind the growth in the volume of transmitted traffic. Consequently, there is a contradiction, as shown in Fig. 1.

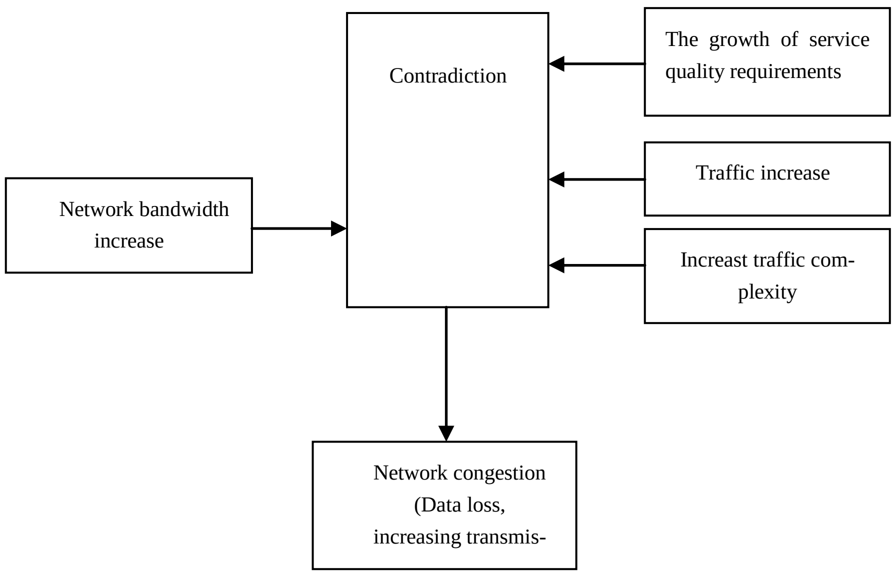

Fig. 1. Contradictions in the development of modern networks.

Under these conditions, traditional encoding methods partially provide efficient transmission of video data in terms of intensity/quality.

At the same time, their applicability is limited, on the one hand, by the fact that a further increase in the compression ratio leads to a decrease in the quality of the reconstructed video data [1-3].
On the other hand, on the basis of traditional methods, it is not possible to construct effective mechanisms for matching video intensity with network bandwidth. In this regard, an approach focused on eliminating the structural redundancy of bit description of fragments of a frame is of interest.

In this case, it is necessary to take into account that video traffic has a pulsating nature, as a result of the influence of a number of factors. This dependence is expressed in the fact that the intensity $R(t)$ of the video stream over time will be variable, as shown by the next expression:

$$R(t) = \varphi(\theta_1, \theta_2, \theta_3, \theta_4), \tag{1}$$

where φ is the functional, describing the relationship between the frame bit intensity and the factors affecting its value;
θ_1 are sets of patterns, respectively, defining statistical, semantic and psycho-visual features of the frame;
θ_2 is a parameter, that determines whether the frame belongs to P, I, or B type, i.e. technological aspect, followed by processing in the MPEG scheme;
θ_3 is a set of technological concepts and architectures embedded in the codec. For example, the presence of motion compensation technology, the presence of technology to mask the contours of objects, the presence of technologies of M-dimensional orthogonal transformation;
θ_4 is a plurality of controlled codec parameters.

In turn, the dynamics of network bandwidth $B(t)$ formed as a result of the influence of a number of other external factors, which is given by the next expression:

$$B(t) = \varphi(\vartheta_1, \vartheta_2, \vartheta_3, \vartheta_4, \vartheta_5, \vartheta_6), \tag{2}$$

where ϑ_1 is a nominal network bandwidth;
ϑ_2 is a network load factor;
ϑ_3 is an interference exposure factor;
ϑ_4 are the topological features of the network;
ϑ_5 an indicator of traffic intensity depending on the time of day;
ϑ_6 an indicator of traffic complexity.

Thus, the formation of values $B(t)$ and $R(t)$ occurs independently of each other. In this regard, it is possible to consider effective the transmission of video traffic at which the video stream intensity is consistent with the current bandwidth, i.e.:

$$R(t) \to B(t), \tag{3}$$

at the same time, the level of introduced error should be minimal.

From this, in the general case, we can conclude that the most effective methods are those that allow, at the level of the video source, to carry out a flexible change in the intensity of the video stream depending on the current level of bandwidht.

Therefore, the method should simultaneously provide both a decrease in information intensity and the ability to control its level in accordance with expression (3).

2 Rationale for controlling the intensity of the video at the source

In general, the quality Q of the video data received at the receiving side depends on such indicators as:

- an indicator of packet loss ζ;
- transmission delays Δt.

In turn, one of the reasons for the occurrence of data loss is due to the presence of a delay exceeding the permissible value, i.e.:

$$Q = f(\zeta, \Delta t), \tag{4}$$

where f is a functional determining the method of processing video at the source level.

In turn, the delay in transmitting packets on a network section from a source to a data recipient is generally described by the next expression:

$$\Delta t = t_{pf} + t_{nt} + t_{rs} + t_{np}, \tag{5}$$

where t_{pf} is the time of packet formation on the transmitting side, which includes all stages of processing at the source level, including buffering and channel coding;
t_{nt} is the packet transit time on the network;
t_{rs} is the packet delay at the receiving side, taking into account the time of buffering and decoding data;
t_{np} - processing time on network nodes.

In turn, time is the sum of the delay values introduced at each of the network nodes in the area from the source to the recipient, i.e.:

$$t_{np} = \sum_{i=1}^{I} t_{np}^{(i)}, \tag{6}$$

where $t_{np}^{(i)}$ is the processing time on the $i-$th network node;

I is the number of network nodes in the path of traffic from the source to the recipient.

At the same time, the delay introduced at the network node can be described by the next expression:

$$t_{np}^{(i)} = \varphi(\Phi_{perf}; R_{\Sigma}; R_{pack}; \Phi_{und}), \tag{7}$$

where Φ_{perf} is an indicator of the performance of node switching equipment;
R_{Σ} is a total intensity of the load arriving at the i – th network node;
R_{pack} is an average packet size arriving at the i – th network node;
Φ_{und} is an indicator of traffic heterogeneity.

Performance indicator Φ_{perf} value is limited. Therefore, you can reduce the delay by:

- reduce the intensity of the incoming load on the node;
- traffic management at the level of network nodes.

At the same time, traffic control at the node level can be ineffective as a result of such reasons:

- when redistributing traffic using prioritization and processing of smart queues, data reduction is not ensured;
- in the conditions of the presence of many packets of minimum length, uncertainty arises when one or another part of high priority packets may not be serviced during processing.

In such conditions, the most appropriate is to reduce the total load R_{Σ} entering the node.

In other words, it is necessary to reduce the intensity of video data entering the network from each of the sources

3 General principles of construction and method requirements

Based on the fact that most of the existing video coding technologies (in particular, the MPEG family) use DCT transformation, the DCT transformant will be selected as the structural unit of video data with respect to which the elementary code structure will be built.

In order for the developed method to allow further manipulation of the number of bits to describe the frame of the video stream, it is necessary that there is the possibility of changing the bit volume to represent the DCT transform.

Then it is advisable to carry out the coding of the transform in the bit, and not the component region.

In other words, the code description of the transformant will be considered as a union of the code constructs of transformants of individual bit planes.

At the same time, in order not to have a significant additional computational load, the method should have the following characteristics:

- be simple and trivial to implement [4];
- not require large amounts of service data [5-9];
- use existing informative features to identify structural features of the processed fragments [10-14].

4 Method realization

Each coefficient $y(p)_{k\ell}$ of the k,ℓ-th component description of the p-th transformant is a set of bits $\alpha_{k\ell}^{(n)} 2^{n-1}$, i.e. $y(p)_{k\ell} = \sum_{n=1}^{\mu} \alpha_{k\ell}^{(n)} 2^{n-1}$, where μ is a number of a bit-planes and $\alpha_{k\ell}^{(n)}$ is a weight coefficient, determines the bit belonging to a specific bit plane [14,18-20]. Such a way of describing transformants allows identifying chains of binary elements $\{\ell_{k\ell}^{(1)},...,\ell_{k\ell}^{(\theta)},...,\ell_{k\ell}^{(\Theta)}\}$ within each bit plane.

The detected lengths of binary elements can be used as a simple informative feature that describes the structural features of the bit plane. This meets the requirements of the method.

In terms of the possibilities of increasing the lengths of binary series, it seems more appropriate to process the bit representation of transformants in the direction of horizontal planes.

This approach is explained by the fact that the unit elements of the higher order bit planes corresponding to the high-frequency components will mainly be absent.

Also, processing the bit transformant representations in the direction of horizontal planes will be appropriate if necessary to ensure the restoration of images according to the hierarchical principle.

As a rule, the regions of high-frequency components of the transform correspond to the region of zero elements within the bit plane. In this regard, in order to increase the lengths of detected binary series, it is proposed to bypass the bit plane in the following manner:

- in the direction of the diagonal bypass (Fig. 2);

- every time, starting again from an element of the bit plane (BP) at the position with coordinates.

In this case, the lengths of binary series form an array M by size of $\upsilon \times \varepsilon$, each line of which contains the corresponding length of the binary series as they are identified.

In this case, a code description of the column of the array M as a nonequilibrium positional number can be obtained, namely:

$$E(q)^{(\mu)}_{\alpha} = \ell^{(\mu)}_{\alpha,1} \prod_{\phi=2}^{\Theta_\alpha} (b_\phi + 1) + \ldots + \ell^{(\mu)}_{\alpha,\theta} \prod_{\phi=\theta+1}^{\Theta_\alpha} (b_\phi + 1) + \ldots + \ell^{(\mu)}_{\alpha,\Theta_\alpha} = \tag{8}$$

$$= \sum_{\theta=1}^{\Theta_\alpha} \ell^{(\mu)}_{\alpha,\theta} \prod_{\phi=\theta+1}^{\Theta_\alpha} (b_\phi + 1) ,$$

here $E(q)^{(\mu)}_{\alpha}$ is a α - th sequence of lengths of binary series for μ - th bit-plane q - th transformant;

$\ell^{(\mu)}_{\alpha,\theta}$ is a lengths of θ - th binary series related to α - th sequence identified within μ - th bit-plane;

$(b_\theta + 1)$ is an base element $\ell^{(\mu)}_{\alpha,\theta}$, as an element of the non-equilibrium positional number (NEPN);

$\prod_{\phi=\theta+1}^{\Theta_\alpha} (b_\phi + 1)$ - weight coefficient for the length of θ - th series;

Θ_α - the number of lengths of binary series (BS) in α - th sequence.

Here, the base element $(b_\theta + 1)$ of a non-equilibrium positional number is defined as follows:

$$b_\theta = \psi_{b\alpha}(\ell^{(\mu)}_{\alpha,\theta,1}, \ldots, \ell^{(\mu)}_{\alpha,\theta,\varphi}) , \tag{9}$$

where ψ_{bm} is a functional determining the size of the base b_θ, depending on the lengths of the binary series;

φ is the number of BS lengths for which a common base b_θ is formed.

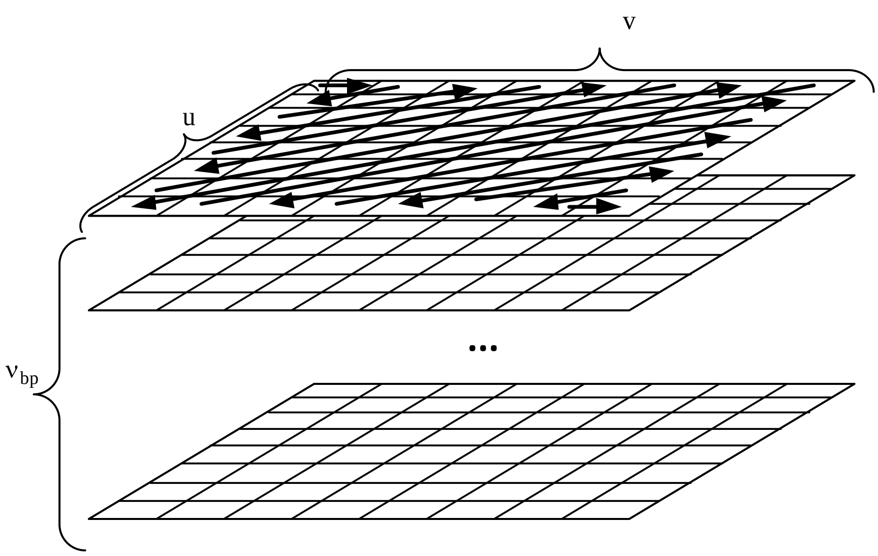

Fig.2. Bit plane traversal order.

The basements $\{b_\theta\}$ system of non-equilibrium positional numbers is here the main service information, using which decoding can be performed.

At the same time, in accordance with the advanced requirements for the construction of the method, the amount of overhead data should be minimal.

With this in mind, the base of a nonequilibrium positional number will be calculated for several lengths of binary series, namely, by the value of the maximum series length in an array row M, namely:

$$b(q)_\alpha^{(\mu)} = \max\{\ell_{\alpha,1}\ \ell_{\alpha,2} \ldots \ell_{\alpha,\beta} \ldots \ell_{\alpha,\varepsilon}\}+1, \quad \alpha=\overline{1,\varepsilon}. \tag{10}$$

The procedure for creating an array M taking into account that the number of columns Φ is not known in advance (since the array is filled as binary series are detected), and one column of the M array can include both a sequence of binary series lengths generated for one bit-plane and a sequence of binary series lengths obtained for total bit transfornant representation, is to implement the following steps:

- for $\alpha=1$ and $\beta=1$, forms element $\ell_{1,1}$ equal to $\ell_{1,1}=\ell_1$, where ℓ_1 is the length of the first binary series obtained for the bit transformant representation;
- if $(\alpha;\beta)$-th element value step $\ell_{\alpha,\beta}$ equal to the value $\ell_{\alpha,\beta}=\ell_\varphi$ where ℓ_φ - is the length of ϕ-th binary series, then $\upsilon(\beta-1)+\alpha<\Phi$, i.e. not all formed binary series lengths distributed over the array M, the relation $\ell_{\alpha,\beta}=\ell_{\varphi+1}, \to \alpha\le\upsilon$ or $\ell_{1,j+1}=\ell_{\varphi+1}, \to i>s$.

In the opposite case, for the $\upsilon(\beta-1)+\alpha=\Phi$ calculation process is considered complete and $\upsilon_\varepsilon=\alpha$, and $\varepsilon=\beta$.

With this in mind, the number of bits $W(E(q)_\alpha^{(\mu)})$ to represent the code value of the nonequilibrium positional number will be determined as follows:

$$W(E(q)_\alpha^{(\mu)}) \le \begin{cases} [\ell og_2 \prod_{\alpha=1}^{\upsilon} h_\alpha - 1]+1, \to \beta<\varepsilon; \\ [\ell og_2 \prod_{\alpha=1}^{\upsilon_\varepsilon} h_\alpha - 1]+1, \to \beta=\varepsilon. \end{cases} \tag{11}$$

Thus, the code description $E(q)^{(\mu)}$ of the bit plane of the DCT transform will be a combination of individual code values of non-equilibrium positional numbers, as shown by the following expression:

$$E(q)^{(\mu)} = \bigcup_{\alpha=1}^{\varepsilon} E(q,\alpha)_m^{(\mu)} \tag{4}$$

Similarly, the code description of the transform will be formed by a set of separate independent code constructions of bit planes, i.e.:

$$E(q) = \bigcup_{\mu} E(q)^{(\mu)} \tag{5}$$

In the general case, based on expression (5), an algorithm for controlling the intensity of a video stream can be implemented.

The main control parameter is the number of bit planes that will be used to describe the transform.

It should be borne in mind that in order to decide which bit planes will be excluded from consideration, it is necessary to first evaluate the contribution of each of the bit planes to the description error of the reconstructed transform. This need is due to the fact that the distribution of individual elements to describe the component will be random [15-17, 21].

Therefore, for a real transform, the dependence of the introduced error on the position of the excluded bit plane will not be described by the classical curve bit-rate/distortion [22-26].

Conclusions

A method for the code description of fragments of a video frame, which are used as DCT transformants, is developed.

In this case, the code description of the transform is formed by separate and independent code structures of bit planes, each of which is obtained by combining the code values of nonequilibrium positional numbers.

This approach makes it possible, in addition to reducing the information intensity of the video stream through the use of nonequilibrium position coding, to control the intensity of the video data on the source side.

References

1. Recommendations of the International Telecommunication Union ITU-T G.1010 "End-User multimedia QoS categories".
2. Barannik V.V., Podlesny S.A., Yalivets K., Bekirov A. E. The analysis of the use of technologies of error resilient coding at influence of an error in the codeword. Modern Problems of Radio Engineering. Telecommunications and Computer Science (TCSET), 2016 13th International Conference, 2016. pp. 52-54. DOI: 10.1109/TCSET.2016.7451965.
3. Barannik V.V., Ryabukha Yu.N., Tverdokhleb V.V., Barannik D.V. "Methodological basis for constructing a method for compressing of transformants bit representation, based on non-equilibrium positional encoding". 2nd IEEE International Conference on Advanced Information and Communication Technologies, AICT 2017, Proceedings, Lviv, 2017, pp. 188. DOI: 10.1109/AIACT.2017.8020096.

4. Barannik V.V., Podlesny S.A., Krasnorutskyi A.O., Musienko A.P., Himenko V.V. The ensuring the integrity of information streams under the cyberattacks action. East-West Design & Test Symposium (EWDTS), 2016 IEEE, 2016. pp. 1-5. DOI: 10.1109/EWDTS.2016.7807752.
5. Christophe E., Lager D., Mailhes C. Quality criteria benchmark for hiperspectral imagery. IEEE Transactions on Geoscience and Remote Sensing. Sept 2005. Vol. 43. No 9. P. 2103–2114.
6. R.C. Gonzales and R.E. Woods, "Digital image processing," in Prentice Hall, New Jersey, edition. II, 2002. – 1072 p.
7. Ian Richardson, H.264 and MPEG-4 Video Compression: Video Coding for Next-Generation Multimedia / Ian Richardson, pp. 368, 2005.
8. Y. Zhang, S. Negahdaripour and Q. Li, "Error-resilient coding for underwater video transmission," OCEANS 2016 MTS/IEEE Monterey, CA, 2016, pp. 1-7.
9. Tanenbaum A., Van Steen M. Distributed systems. Pearson Prentice Hall, 2007.
10. H. Baccouch, P. L. Ageneau, N. Tizon, N. Boukhatem, "Prioritized network coding scheme for multilayer video streaming", 2017 14th IEEE Annual Consumer Communications & Networking Conference (CCNC), pp. 802-809, 2017.
11. W. J. Tsai, Y. C. Sun, "Error-resilient video coding using multiple reference frames", 2013 IEEE International Conference on Image Processing, pp. 1875-1879, 2013.o transmission", 2016 8th IEEE Intern. Conf. on Commun. Software and Networks (ICCSN), pp. 561-564.
12. A. Chigorin, G. Krivovyaz, A. Velizhev, A. Konushin "A method for traffic sign detection in an image with learning from synthetic dat" 14th International Conference Digital Signal Processing and its Applications, 2012, pp. 316-335.
13. D. Salomon. "Data Compression: The Complete Reference". Fourth Edition. Springer-Verlag London Limited, 2007. - 899 p.
14. Vatolin D., Ratushnyak A., Smirnov M. and Yukin V., Methods of data compression. The device archiver, compression of images and videos. M.: DIALOG MIFI, 2013, 384 p.
15. Lee S.. Yoon J. Temporally coherent video matting. Graphical Models 72. 2010. P. 25-33.
16. Barannik V.V., Alimpiev A., Bekirov A.E., Barannik D.V., Barannik N. Detections of sustainable areas for steganographic embedding. East-West Design & Test Symposium (EWDTS), 2017 IEEE, 2017. P. 1-4. DOI: 10.1109/EWDTS.2017.8110028.
17. Ding Z., Chen H., Gua Y., Peng Q. GPU accelerated interactive space-time video matting. In Computer Graphics International P 163-168. 2010.
18. Ablamejko S.V., Lagunovskij D.M. Obrabotka izobrazhenij: teh-nologija, metody, primenenie. - Minsk: Amalfeja, 2000. – 303 p.
19. K. R. Rao and J. J. Hwang, Techniques and Standards for Image, Video and Audio Coding. EnglewoodCliffs, NJ: Prentice-Hall, 1996.
20. Shi, Yun Q. Image and video compression for multimedia engineering: fundamentals, algorithms, and standards / Yun Q Shi, Huifang Sun, NY, CRC Press, 2008, 576 p.
21. Bai X., Wang J. Towards temporally-coherent video matting. Proceedings of the 5th international conference on Computer vision/computer graphics collaboration techniques. MIRAGE'11, Springer-Verlag. 2011. P. 63-74.
22. M. Al Hadidi, J. S. Al-Azzeh, R. Odarchenko et al, Adaptive Regulation of Radiated Power Radio Transmitting Devices in Modern Cellular Network Depending on Climatic Conditions, Contemporary Engineering Sciences, Vol. 9, № 10, pp. 473-485, 2016.
23. JPEG 2000 image coding system: Core coding system, itu-t recommendationt. t. 800 and iso/iec 15444-1. ITU-T and ISO/IEC JTC 1, 2000.

24. Al-Azzeh J.S., Al Hadidi M., Odarchenko R., Gnatyuk S., Shevchuk Z., Hu Z. Analysis of self-similar traffic models in computer networks, International Review on Modelling and Simulations, № 10(5), pp. 328-336, 2017.
25. In J., Shirani S., Kossentini F. Jpeg compliant efficient progressive image coding.Department of Electrical and Computer Engineering,University of British Columbia, Vancouver BC V6T 1Z4, Canada.
26. R. Odarchenko, V. Gnatyuk, S. Gnatyuk, A. Abakumova, Security Key Indicators Assessment for Modern Cellular Networks, Proc. of the 2018 IEEE First Intern. Conf. on System Analysis & Intelligent Computing (SAIC), Kyiv, Ukraine, 2018, pp. 1-7.