Regression Analysis of Correlation between Video Coding parameters and Sequence Modification Analyzers

Ilya Brailovskiy, Natalya Solomeshch
Intel Corporation,
{ilya.v.brailovskiy, natalya.solomeshch}@intel.com

Abstract

In this paper we analyze correlation between video sequence modification scoring and video coding algorithms such as Motion Estimation and Rate Control. Using two open source codecs X.264 and IPP H.264 we conducted number of tests. By means of linear regression analysis we processed obtained test data. Based on this analysis we conclude that Decimation, Motion and Noisy Frame techniques have high correlation with different codec algorithms. Moreover, we refined our experiments further and found out that higher rank in Noisy Frames insertion analysis correlates with better Scene Change Detection algorithm and with higher visual quality. However, the abovementioned methods have tangible correlation with such encoding characteristics as number of B pictures between reference frames and thus can't be used in competitive ranking without certain assertions. We also point out that Overall analyzer has high correlation with all key encoding decision algorithms and potentially very good for "black box" codec competitive analysis.

Keywords: *Video codec analysis, visual quality metrics, linear regression, scene change detection*

1. INTRODUCTION

Automated analysis of visual quality for video codecs such as MPEG codecs or AVC [1,10,11] is one of the most crucial and important aspects which arises in design, development and implementation of video codecs. Full reference image quality metrics like PSNR or SSIM [2-3] provide relatively good correlation with subjective human perception. Such Visual Quality metrics however do not provide a way to characterize codecs in a holistic way. Codecs quality depends on a great number of input parameters such as bitrate, input video source and different algorithmic options. Visual Quality metrics could characterize only one point in the codec parameter space since a full reference metric just compares one input video source with one output video sequence. Such comparison could provide certain indication on how good a codec performs on specific input and with fixed parameter settings but gives no information on how the codec operates for different bitrate or another source clip.

To overcome this limitation number of methodologies had been proposed [5-7]. In this paper we will focus on the methodology proposed in papers [6,7] which allows systematically analyze and compare video codecs by iterating though different encoding parameters following certain rules. Compared to some other methodologies (for example [5]) this approach presumably allows some insight into codec internals as well as connecting these internals with the output visual quality.

Goal of this paper is by means of computational experiments for existing codec implementations analyze and prove or disprove correlation between codecs scoring provided by the methodology [6,7] and specific encoding options. Another goal we would like to achieve is to find analyzer's and codec's parameters limitations (if any) for "black box" testing based on empirical study for the selected set of analyzers proposed in [6,7]. Please note that initially we do not assume that a more advanced algorithm provides higher scoring, or higher scoring provides better visual quality. Our goal is just to establish a correlation between codec scoring and encoding parameters. However during the course of our research we found certain empirical connection between higher scoring and better visual quality when more advanced encoding configurations are applied.

2. CODEC ANALYSIS METHODOLOGY

We briefly recall codec assessment and analysis methodology proposed in [6,7]. We will strictly follow these papers in our overview.

2.1 Video Codec Scoring with Specially Prepared Video Sequence

The overall high-level scheme of the examined method is provided at Fig. 1.

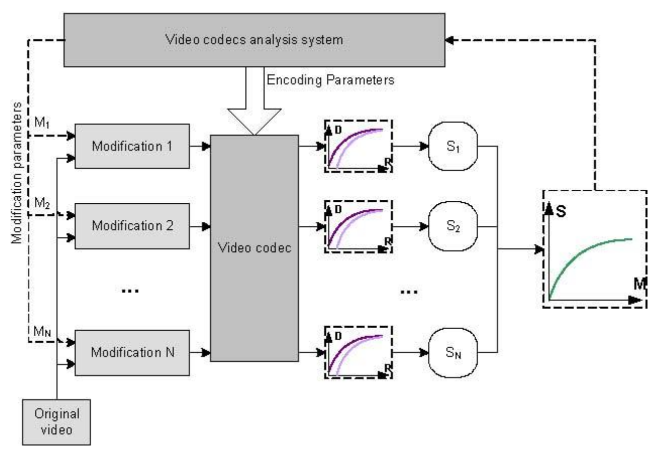

Fig. 1. General scheme of video codec analysis.

The first stage involves generation of a series of video feeds according to known modification parameters M_i. These video feeds could be created by modifying natural video or could be generated fully artificially. Regardless the way these feeds are created we will call them "modified sequences". Modified sequences are submitted to the codec under the test, which encodes and decodes each of the modified streams. Resulted reconstructed data is compared with initial modified sequences and Rate-Distortion (RD) curves are calculated. The same procedure is applied to the reference video codec and RD curves for the reference codec are constructed. Based on calculated RD curves, baseline scoring S^r is computed for each of the modifications by the following formulae:

$$S_{1,2}^{[a,b]} = \exp\left\{\frac{1}{b-a}\int_a^b \ln\left(\frac{R_1(D)}{R_2(D)}\right)dD\right\},$$

where $R_1(D)$ and $R_2(D)$ are the RD curves for the tested and reference codecs correspondingly and *[a,b]* is the range of the quality metric that we use to conduct our comparison. Value of $S_{1,2}^{[a,b]}$ characterizes the average ratio of bitrate, for a same quality for a set amount of introduced distortion that can be achieved by a video codec with corresponding RD characteristics.

At the next phase the video codec analysis system takes the arrays of scoring results S^r, S^t, $S_{ref,tested}^{[a,b]}$ for the original sequence (if any) and the array of modification parameters M as the input and calculates the scoring value estimate for a sequence modification by following formula:

$$Q(M, S^r, S^t) = (1-\gamma)\left(\alpha(M, S^t) - \alpha(M, S^r)\right) + \gamma\, S_{ref,tested}^{[a,b]},$$

$$\alpha(M, S) = \frac{n \sum_i M_i S_i - \sum_i M_i \sum_i S_i}{n \sum_i {M_i}^2 - (\sum_i M_i)^2},$$

where γ is a constant in the range *[0,1]*, i is modified sequence number, n - total number of modifications for given analyzer and $\alpha(M,S)$ is the slope of the approximating line (the average rate of change of the coding efficiency).

The final step combines estimates for several different codec analyzers and produces the final codec score. For more details on scoring value calculation please refer to [6,7].

Let us briefly recall what the Noisy Frames, Decimation, Motion and Overall analyzers are.

2.2 Modification analyzers overview

Noisy frames analyzer. The analysis lies in modification of input video sequence in order to complicate temporal prediction algorithms. In modified sequence some frames are replaced by "noisy frames", the whole consisting of random noise. One can vary the number of such frames from test to test, usually from one to ten. This method is assumed to reveal a quality of work of the scene detector and of the rate control method in the first place.

Decimation analyzer compares RD-curves obtained during encoding of original and modified video sequences. The modified sequence is created by systematic removal of frames from the original one. One can vary the number of the removed frames – one out of each two frames, two out of three frames, etc. This method is expected to have the highest correlation with scene analyzer and rate control method.

Motion analyzer makes the comparison of RD-curves, obtained during encoding of synthetic video sequences. Synthetic sequences are composed of blocks of randomly generated sizes and colors. Some of the blocks are moving with random velocity. Complexity of this analyzer could be varied by varying number of such blocks, their speed and color. Motion analyzer is as it can be by judged by the name supposed to allow evaluating methods of Motion Estimation implemented in the encoders.

The main idea behind the **Overall analyzer** is to use a set of objective metrics over number of source streams simultaneously in order to increase the approximation with subjective measurements. It is expected that the Overall analyzer would correlate with all the encoding parameters with some emphasis on commonly accepted encoding algorithms responsible for the quality of codecs: Rate Control, Scene Change Detection, Rate-Distortion optimizations, MacroBlock Decisions as well as some others algorithms.

2.3 Regression Analysis and Test Setup

For studying dependency of modification analyzers scores on codec encoding options we are using quite straightforward and well-proven linear regression analysis [12]. Against each possible combination of encoding option calculate analyzer score and linear regression coefficients for encoding options. We consider correlation between chosen analyzer and encoding option as a high if corresponding regression coefficient is "high". We would divide regression coefficients on "highs" and "lows" empirically by examining the difference between the maximal and minimal coefficient values.

Unfortunately we can't execute such experiments for all possible codecs and for all possible encoding options. We need to limit our analysis to certain codecs, specific streams and specific encoding options bound by many practical limitations. As we will show a little later even limited range of options leads to quite significant computational time.

We used 4 streams of High Definition (HD), Standard Definition and Common Intermediate Format (CIF) resolutions in our experiments. For each resolution we selected 5 bitrates for RD curves calculation, more details on selected test streams can be found in the Table 1. We run our tests iterating though all encoding parameters simultaneously but for each resolution (HD, SD and CIF) separately.

We conducted our experiments with several codecs: X.264 Open Source H.264 encoder [8], IPP Media Sample H.264 encoder and IPP Media Sample MPEG2 encoder [9]. All these codecs are Open Source thought distributed under difference licenses. Compared to some other Open Source codecs such as JM reference model these codecs are significantly faster which makes execution of bigger number of test runs possible. In the Table 2 you could see which systems we used for testing and in the Table 3 you could see how long it took us to run those tests even with selected relatively fast codecs.

Although we didn't run some other codecs such as MPEG4 part 2 or VC1 [1] because of computational resource limitations we think that the results obtained for H.264 codecs wouldn't be significantly different for other codecs. Some indications that that might be true are provided by the fact that for several runs of MPEG2 encoder we observed quite similar to H.264 codecs results. Extending results of our research for the whole family of AVC encoders looks quite reasonable for us since we took different AVC encoders implementations and observe certain consistency in the results. However, strictly speaking, our study applies for those codecs we tested only.

Type stream	Name	Resolution	Bitrate (Kbit/s)
CIF	Foreman	352x288	200, 400, 800, 1600, 2200
SD	Iceage	720x576	750, 1200, 2000, 4000, 7000
	Lotr	720x416	
HD	Troy	1920x1072	2000,5000,8000,11000,15000

Table 1. Parameters of the used sequences.

Computer numbers	OS	CPU	Cores Num	CPU (GHz)	MEM (Gb)
1	Windows Server 2003 R2 Ent	Xeon	16	2.4	16
2	Windows Vista Enterprise	Xeon 3230	8	2.66	2
3	Windows Server 2003	Core 2 Quad	4	2.4	4
4	Windows Server 2003 R2 x64	P4	2	2.8	1

Table 2. System Configuration used for testing.

	Noise frames	Overall	Decimation	Motion
X264 Cif	0.83 day (1)*	2 days(4)	2 days(4)	2 days (4)
X264 SD	25 days (1)	17 days(4)	17.5 days(4)	
X264 HD	25 days(1)	14.5 days (3)	30 days(3)	
IPP264 Cif	2.5 days (1)	2 days (3)	2 days(3)	3.33 days (4)
IPP264 SD	22 days(2)	16.33 days(3)	13 days(3)	
IPP264 HD	26 days(1)	27 days (3)	19 days(3)	

Table 3. Average time of executing various measurements. In the brackets one can see the number of used computer.

* The numbers in the brackets correspond to used computer numbers.

For the iterations we choose following encoding parameters mostly available with command line options of the selected encoders:

1. The number of frames between the Intra-frames (keyint: 100, 200, 300);

2. The minimal number of Bi-predictive frames between the Intra and Predicted frames (Bframes: 0, 2);

3. Bitrate control method (rc method): for X.264 the options are (variable bitrate and constant bitrate bufsize 1100 kbit), and for IPP.264 (variable bitrate and constant bitrate);

4. Motion Estimation method: for X.264 (me method: diamond search, hexagonal search, uneven multi-hexagon search, exhaustive search, hadamard exhaustive search); for IPP.264 (MV_SEARCH_TYPE: LOG, EPZS, FULL_ORTHOGONAL, LOG_ORTHOGONAL, UMH, SQUARE, FTS, SMALL_DIAMOND);

5. Macroblocks size: (subblock split:

1. I16x16 P16x16 B16x16;
2. I16x16 P16x16 B16x16 I8x8 P8x8 B8x8;
3. I16x16 P16x16 B16x16 I8x8 P8x8 B8x8 I4x4 P4x4 B4x4);

6. Combination of various optimizations: for X.264 (subme + trellis: 1&0, 3&0, 4&1, 6&1, 7&2, 9&2) and for IPP.264 (combination of various optimizations – specific value in par-file which turns on/off similar options: 0,1,2,3,4,5);

7. Flag which indicates presence of Scene Change Detector Analyzer: for X.264 (no-scenecut), for IPP.264 (internal building preprocessing directive option) IPP MPEG2.

We wouldn't go into the details of the meaning of these options, please refer to [8,9,11].

3. RESULTS

We summarized results of our regression analysis coefficients calculation in the Tables 4 and 5. We present only SD resolution results since HD and CIF results have essentially the same pattern for both AVC codecs. Let us highlight some of the observations one can make by looking into the tables.

At first we would like to go though the results which are common for both codecs. Scene Change detection algorithm, Rate Control method and Subblock split option all have very high impact for both codecs for almost all analyzers. KeyFrame interval has small impact on both codecs. ME algorithm choice have small impact on the quality for both codecs mostly because it has high impact on performance which is out of the scope of this study. Noisy Frame and Decimation analyzers both are mostly influenced by RC method and Scene Change Detection algorithms. Overall analyzer demonstrates similar correlation pattern for both codecs though Scene Change Detection algorithm has much smaller impact on this Analyzer for IPP H.264 compared to X.264. Regression coefficients for Motion analyzer are mostly influenced by Subblock Split and Scene Change Detection options because these algorithms contribute most to Mode Decision process.

Second, let us emphasize some differences between codecs. B-frames number has significant impact on IPP H.264 while it has almost negligible impact on X.264 encoder. IPP H.264 is more influenced by Rate Control for all analyzers while X.264 is not.

Varying parameters	Noise frames	Decimation	Motion	Overall
Keyint	0.1506	1.3935	0.0156	0.3789
Bframes	0.0110	0.06545	0.3854	0.0301
Rc method	24.2025	9.0777	0.6607	7.4427
Me method	0.1876	0.0732	0.8339	0.4331
Subblock split	7.5508	2.3768	9.5350	7.1469
Regimes of optimization	4.0886	2.0157	1.3081	6.6421
Scene Change Detection	19.7372	22.2783	14.9058	26.2165

Table 4. Linear Regression Analysis Coefficients for X.264 Open Source H.264 encoder on SD sequences.

Varying parameters	Noise frames	Decimation	Motion	Overall
Keyint	1.1311	2.020	0.0000	1.9778
Bframes	6.4867	8.018	9.7880	0.4410
Rc method	25.3586	31.0780	14.0873	23.555
Me method	0.0529	0.0103	1.1588	0.0828
Subblock split	4.6134	1.9246	10.6397	2.2763
Regimes of optimization	4.9054	0.3718	0.0357	8.8911
Scene Change Detection	10.5375	4.6043	8.8423	2.0616

Table 5. Linear Regression Analysis Coefficients for IPP Media Sample H.264 encoder on SD sequences.

Most of the differences between two codecs are observed for Motion analyzer, and they fall into the general codecs' differences most likely caused by the difference in the implementation of these codecs. However Scene Change Detection and Subblock Split regression coefficients for Motion Analyzer are on the higher side for both encoders.

In order to examine influence of Scene Change Detection and Subblock Split options on Motion Analyzer more precisely we performed additional test on X.264 Open Source H.264 encoder using extended set of parameters. We have conducted additional linear regression analysis for Motion Analyzer (Table 6) and included chromaticity motion estimation option (no-chroma-me on and off) into consideration. The names of the synthetic sequences 1, 2, …, 15 represent the complexity of the Motion

Analyzer. Smaller numbers represent simpler sequences with slower motion; bigger numbers represent higher complexity and higher motion. From obtained data we can observe that influence of Scene Change Detection goes up with higher speed and more complex sequences. This effect could be explained that when complexity goes up Scene Change Detection algorithm can't find good correlation between two consecutive frames and often assumes a new scene.

Varying parameters	Names of the comparing sequences			
	1,2,3,4,5,6, 7,8,9,10,11 , 12,13,14,1 5	1,15	1,2	14,15
Me method	1.2517	0.8376	1.9692	0.8367
Subblock split	7.8453	9.9343	12.9871	8.9431
Regimes of optimization	1.3446	1.0707	5.9788	0.7874
Scene Change Detection	18.9256	18.4604	4.4785	19.4741
Chroma-ME	1.8508	1.1622	0.1311	1.4784

Table 6. Results of the regression analysis for the Motion analyzer for X.264 Open Source H.264.

From the results of this experiment we can suggest that if Motion Analyzer is planned to be used for evaluating quality of Motion Estimation and Macroblock Decisions algorithms it should set smaller speed for moving blocks. Complexity classes from 1 to 3 are probably enough to evaluate the complexity of core Motion encoding algorithms.

4. INFLUENCE OF THE SCENE CHANGE DETECTION ON THE QUALITY OF ENCODED VIDEO SEQUENCE.

As we mentioned earlier the influence of Scene Change Detection (SCD) algorithm is quite significant for both codecs especially in case of Noisy Frames. (Please note that we will be using acronym SCD for Scene Change Detection in this section). It's naturally to look further and understand if a better SCD algorithm provides better quality in case of inserted Noisy Frames. For that we will investigate cases when SCD algorithm is on or off. We assume that when SCD is on the quality should be better, and our goal is to check this assumption.

For verifying our assertions in regard with SCD quality impact we need to make sure that 1) encoder with SCD produces different encoded streams compared to an encoder without SCD for the streams with noisy frames. 2) quality of the sequences encoded with SCD has better than quality 3) it's good to see if the encoder uses more Intra frames and places them where the Scene Changes occurs (it's not necessary – sometimes encoders can mitigate a scene change by different means).

To see that SCD indeed improves the visual quality we've built RD curve to compare quality of the encoder when SCD on and off. As you can see from Fig.2 curve with SCD on is higher which means better visual quality.

Table 7 summarize results of analyzing Frame Type statistics for the streams with and without SCD and answers the question 1). Table 7 partially answers question 3) since we see increase of I frames number in case of SCD. We also did more thorough analysis where SCD algorithm places I frames, in most cases excessive I frames are placed where Noisy Frames inserted – which means they are placed where Scene Change occurs.

SCD on	Percentage of frame types	Ordinal numbers of the frames coming after an Intra-type frame				
		№ 1	№ 2	№ 3	№ 4	№ 5
	i, %	15.4	9.2	12.3	1	<1
	p, %	47.4	64.9	53.6	55.6	64.9
	b, %	37.1	25.7	34	43.2	35

SCD off	Percentage of frame types	Ordinal numbers of the frames begin with the frame on which the encoder with analyzer placed an Intra-type frame					
		№ 0	№ 1	№ 2	№ 3	№ 4	№ 5
	i, %	5.1	<1	<1	<1	<1	<1
	p, %	40.2	40.2	19.5	40.2	40.2	19.5
	b, %	54.6	59.7	80.4	59.7	59.7	80.4

Table 7. Frame type summary for SD-type Iceage video sequence with eight noisy frames inserted.

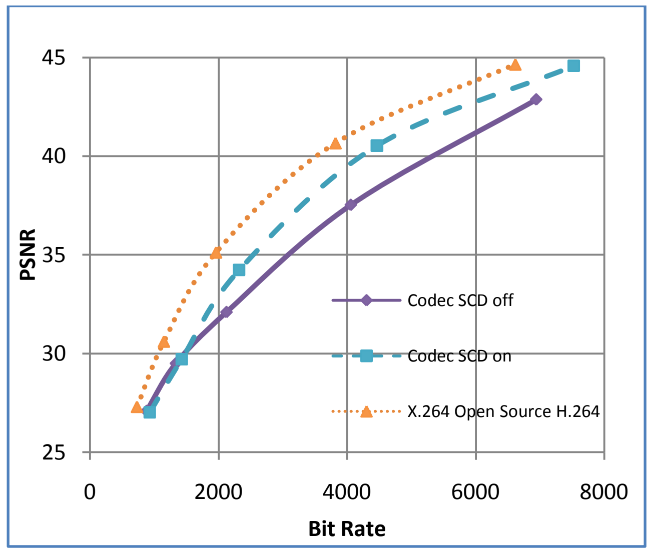

Figure 2. R-D curves for IPP H.264 on a Noisy Stream (8 noise frames inserted) on Iceage stream.

To understand Noisy Frame correlation with SCD better for other class of codecs we extended our experiments further and did similar analysis for IPP Media Sample MPEG2. We obtained quite consistent results such as on Figure 2 for MPEG2 as well which we don't copy to this paper to avoid redundancy with AVC results.

Additionally for both codecs we did visual inspection with enabled SCD and disabled SCD. For both AVC and MPEG2 codecs we noted better or much better visual quality depending on bitrate or on the number of inserted frames. Video clips with higher bitrate or with lesser number of inserted frames looked less different for codecs with SCD on and off. However when there are more inserted frames we observe much better quality for the codecs with SCD on. We put example of observed visual differences on Figure 3. There are significant blocking artifacts on Figure 3 (a) compared to Figure 3 (b). As the only difference during the coding was the presence of the SCD, it can be concluded, that the analysis of a scene change makes a tangible contribution to the visual quality which can be practically demonstrated with Noisy Frame Insertion technique.

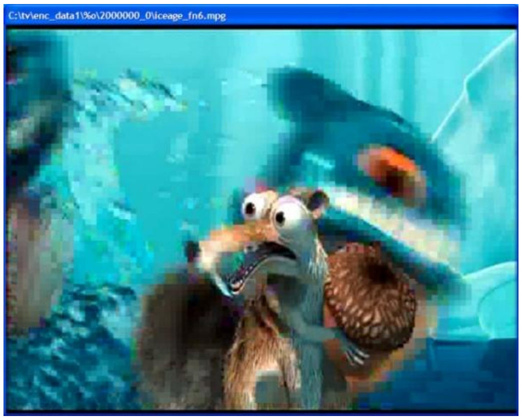

Figure 3.A Scene Change Detection is absent.

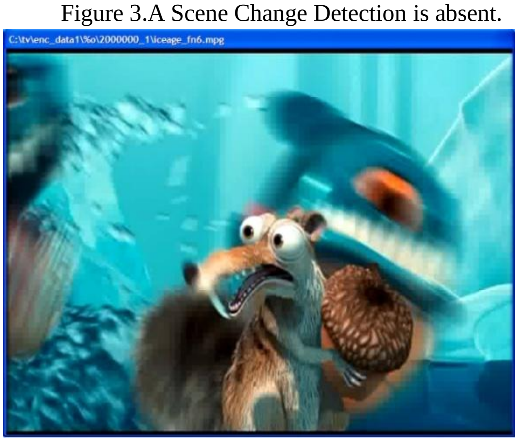

Figure 3.B Scene Change Detection is present.

Figure 3. The visual difference of the quality on 2000 Kbits, 6 noisy frames, for IPP MPEG2 codec with Scene Change Detection (SCD) on and off. (A) SCD is absent. (B) SCD is present.

5. CONCLUSION

Based on performed analysis we conclude that for evaluated X.264 and IPP H.264 codecs the analyzer methodology proposed in [6,7] provides high correlation with key internal encoding algorithms and can be effectively used for testing and comparing video encoders.

Some of the methods described above are not suitable however for "black box" codec testing. For example, Decimation Analysis has high correlation with the number of B Frames for IPP H.264 encoder and not neglible correlation with keyframe interval for X.264 encoder. In order to effectively use Decimation Analysis we recommend aligning GOP structures of the codecs under consideration.

If Motion analyzer is intended to evaluate codecs difference related to Motion Estimation and Macroblock Decisions then less complicated modifications should be used, preferable in the complexity range from 1 to 3.

Nosiy Frame analysis can be effectively used as an indicator of quality of Scene Change Detection algorithm.

Overall analyzer has expectedly high correlation with all major codec quality algorithms and potentially could serve as a measure of overall codec quality. Based on our study applicability of overall analyzer to "black box" codec competitive analysis is highly probable. At the same time, overall analyzer should be examined further since in this paper we haven't covered correlation of overall analysis ranking with visual quality.

6. ACKNOWLEDGMENTS

We would like to thank Alexey Leonenko for very useful discussions and feedback during whole course of this research.

We also would like to thank Alexander Parshin and Dmitry Vatolin for their kind permission of publishing Figure 1 and some other materials obtained during joint research works.

7. LITERATURE

[1] E. G. Richardson H.264 and MPEG-4 video compression: Wiley and Sons, 2003

[2] Digital video quality: vision models and metrics. Stefan Winkler. Published by John Wiley and Sons, 2005

[3] Digital video image quality and perceptual coding Hong Ren Wu, Kamisetty Ramamohan Rao. Published by CRC Press, 2006

[4] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600-612, Apr. 2004

[5] S. Wolf and M. H. Pinson, "Low bandwidth reduced reference video quality monitoring system", First Int'l Workshop on Video Proc. And Quality Metrics, Jan 2005

[6] A. Parshin, D. Vatolin, I. Brailovskiy and P. Corriveau, Novel Approach for Video Codecs Scoring using Specially Prepared Video Sequences. In Proceedings of the Fourth International Workshop on Video Processing and Quality Metrics for Consumer Electronics conference, Jan 2009.

[7] A. Parshin and D. Vatolin Methods of analysis of videocodecs based on modifications of natural of natural videosequencies. Graphicon 2008 conference, June 2008

[8] X.264 encoder: http://x264.nl/

[9] IPP H.264 and MPEG2 encoders: http://software.intel.com/en-us/intel-ipp/

[10] ITU-T and ISO/IEC JTC 1, "Generic coding of moving pictures and associated audio information – Part 2: Video," ITU-T Rec. H.262 and ISO/IEC 13818-2 (MPEG-2), Nov. 1994 (with several subsequent amendments and corrigenda).

[11] Draft ITU-T recommendation and final draft international standard of joint video specification (ITU-T Rec. H.264/ISO/IEC 14 496-10 AVC). Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, JVTG050, 2003.

[12] Introduction to Linear Regression Analysis (3rd ed.). Douglas C. Montgomery, Elizabeth A. Peck, and G. Geoffrey Vining . New York: Wiley, 2001.

About the authors

Natalya Solomeshch is an Engineering Intern at Intel and student of MIPT. Her contact email is natalya.solomeshch@intel.com.

Ilya Brailovskiy, PhD, is a Video and Platform Architect with Intel Corporation. His contact email is ilya.v.brailovskiy@intel.com.