Comparison of Objective Quality Assessment Methods for Scalable Video Coding

Denis Vranješ [1], Drago Žagar [1], Ognjen Nemčić [2]
[1] University of Osijek, Faculty of Electrical Engineering, Kneza Trpimira 2B, Osijek, HR-31000, Croatia
[2] Supra Net Projekt d.o.o., Erlichova 2, Zagrab, HR-10000, Croatia
denis.vranjes@etfos.hr

Abstract - Appropriate delivery of a video material over networks under various conditions represents a certain challenge. It is necessary to adapt video content in order to ensure the best possible quality of transmitted material in every moment, regarding to variable network conditions. This problem can be solved with scalable video coding – SVC. Since subjective video quality tests are complex and comprehensive, in this paper the scalable coded video materials are evaluated with several objective video quality metrics. The general idea was to find out which of them gives the best match with the results of subjective tests for scalable coded video materials. There is shown the correlation between subjective and objective tests and general conclusions are drown.

Keywords - SVC, video quality evaluation, objective quality metric

I. Introduction

Multimedia applications, such as videoconferencing, IPTV, e-learning etc., are broadly used nowadays. Therefore, efficient video transmission over networks with various conditions is very important. Quality of transmitted video material affects user's experience, so the quality evaluation is essential to achieve satisfactory Quality of Service (QoS). Subjective tests are good indicators of quality, but they are complex, expensive and time consuming. Besides subjective methods, there are available objective algorithms for video quality evaluation which are less complex.

Several papers compare subjective and objective methods of non scalable video quality evaluation and study which objective methods have the best correlation with subjective scores. In [1] Moorthy et el. used 160 sequences coded by H.264 codec with different bit rates and simulated transmission over wireless channel. Transmitted materials were evaluated with subjective and several objective methods such as Peak signal-to-noise ratio (PSNR), Visual signal-to-noise ratio (VSNR) [2], Video quality metric (VQM) [3], Visual information fidelity (VIF) [4] and Multi-scale structural similarity (MS-SSIM) index [5]. The best correlation is showed for MS-SSIM and VQM algorithms. In [6] Seshadrinathan et el. presented LIVE Video Quality Database [7]. It contains 150 distorted video sequences from 10 different source video content coded by MPEG-2 and H.264 codec which were evaluated by 38 human observers. Besides that, they show performances of several freely available full reference (FR) algorithms. In their study the best results showed MOtion-Based Video Integrity Evaluation (MOVIE) [8] index, while still noteworthy were VQM and MS-SSIM index. In [9] Chikkerur et el. made objective tests on existing LIVE Video Quality Database. In their tests also the best performances were shown by MOVIE, MS-SSIM and VQM.

Since network conditions are time variant, additional efficiency can be achieved with scalable video coding (SVC). There are three types of scalability: quality, spatial and temporal scalability. In [10] Lee et el. presented database with scalable coded video materials and their subjective evaluation [11]. Since the comparison of subjective and objective methods wasn't done for SVC, it is the topic of this paper. The paper is organized as follows. Section II describes codecs, materials and objective algorithms that were used for quality evaluation, while Section III gives resuslts review and analysis. In section IV conclusions of this paper are presented.

II. Scalable Codecs and Video Materials

As it is presented in [10], three different raw sequences, DucksTakeOff, IntoTree and ParkJoy, with different spatial and temporal complexity (Table I) with spatial resolution (WxH) of 1280 x 720 and temporal frequency (F) of 50 fps are coded with two different codecs: scalable video coding (SVC) and wavelet-based scalable video coding (WSVC). One frame form each of those three sequences is presented in Fig. 1. Sequences differ by spatio-temporal activity which is measured by Spatial perceptual Information (SI) value, Temporal perceptual Information (TI) value and the product of SI and TI (SITI) (Table I).

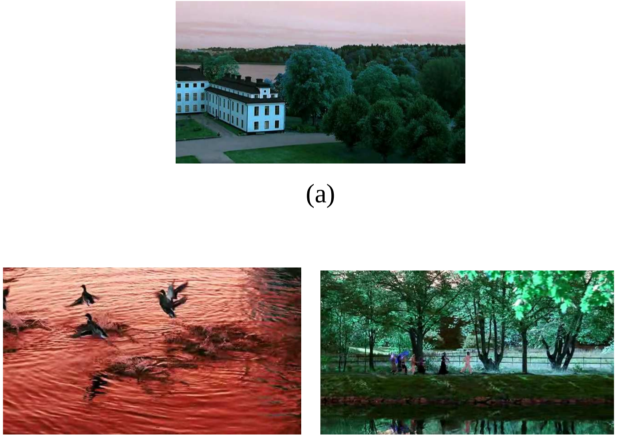
A. *Scalable Codecs*

SVC is scalable extension of H.264/AVC codec where coded bit stream contains several different layers. This extension enables spatial, temporal and quality scalability with the slight bit rate increasing in comparison to H.264/AVC codec. Scalable coded bit stream consists of one base layer and several enhancement layers. Each of them increases quality, but also a bit rate of coded material. Scalable video coding for experiments made in [10] is done with JSVM 9.18 [12] reference software.

By WSVC codec a spatio-temporal decomposition using wavelet transform is done thus ensuring possibility of spatial and temporal scalability. Using the motion estimation, motion information used for computing wavelet coefficients, is given. Compressed bit stream consists of several layers. In experiments made in [10] the method from [13] is used. Bit stream consists of 5 temporal layers, 3 spatial layers and several quality layers.

TABLE I. Spatial and Temporal Activity Values: SI, TI and SITI=SI*TI

Sequence	**SI**	**TI**	**SITI**
IntoTree	7,44	18,64	138,68
DucksTakeOff	13,28	23,18	307,83
ParkJoy	16,32	42,27	689,85



(a)

(b) (c)

Figure 1. Sample frames from the test sequences: (a) IntoTree (b) DucksTakeOff (c) ParkJoy

B. *Video Materials*

As it is already mentioned, scalable coded bit stream consists of several layers in which different combinations of three types of scalability, i.e. temporal, spatial and quality scalability, are possible. In [10] Lee, De Simone and Ebrahimi coded three mentioned sequences with the same or similar bit rates using different types of scalability. Sequences are coded on 4 to 6 bit rates with both SVC and WSVC codec, as it can be seen in Tables II and III. Materials are coded on three different resolutions (W x H), i.e. 320 x 180, 640 x 360 and 1280 x 720, and 4 different frame rates (F), i.e. 6.25, 12.5 25 and 50 fps. As a measure of quality pixel bit rate (B_p) is used and it is mathematically defined in (1) [10]

$$B_p = \frac{B}{H*W*F} \quad , \qquad (1)$$

where B is a bit rate.

It should be noted that IntoTree sequences coded with WSVC codec were not available, so we analyzed only results for ParkJoy and DucksTakeOff sequences. For subjective tests in [10] the sequences coded with lower resolution are upsampled on 1280 x 720 resolution using bilinear filter.

For objective evaluation resolution and frame rate of reference and test material have to be the same. Because of that test materials are upsampled on 1280 x 720 resolution, and their frame rate is set to 50 fps.

TABLE II. Selected Comparison Tests Composed of Multiple Layers Having (Nearly) the Same Bit Rates From the Bit Streams Encoded by SVC. Each Layer is Shown as (B, WxH, F, B_P) Where B, WxH, F and Bp are the Bit Rates in kbps, Spatial Resolution, Temporal Resolution and Pixel Bit Rate, Respectively [10]

IntoTree				DucksTakeOff			
B	W x H	F	Bp	B	W x H	F	Bp
508	320x180	12,5	0,71	358	320x180	6,25	0,99
528	640x360	6,25	0,37	365	320x180	12,5	0,51
1527	1280x720	12,5	0,13	533	320x180	12,5	0,74
1550	640x360	25	0,27	536	640x360	6,25	0,37
1932	1280x720	6,25	0,34	638	1280x720	6,25	0,11
1960	1280x720	25	0,09	642	640x360	6,25	0,45
2350	1280x720	12,5	0,2	753	1280x720	6,25	0,13
2447	1280x720	50	0,05	790	640x360	12,5	0,27
ParkJoy				926	1280x720	12,5	0,08
344	320x180	12,5	0,48	971	640x360	12,5	0,34
365	320x180	6,25	0,51	1542	1280x720	25	0,07
509	320x180	12,5	0,71	1552	640x360	25	0,27
531	640x360	6,25	0,37				
1542	1280x720	6,25	0,27				
1556	640x360	25	0,27				
4062	1280x720	50	0,09				

TABLE III. Selected Comparison Tests Composed of Multiple Layers Having (Nearly) the Same Bit Rates From the Bit Streams Encoded by WSVC. Each Layer is Shown as (B, WxH, F, B_P) Where B, WxH, F and Bp are the Bit Rates in kbps, Spatial Resolution, Temporal Resolution and Pixel Bit Rate, Respectively [10]

ParkJoy				DucksTakeOff			
B	W x H	F	Bp	B	W x H	F	Bp
520	320x180	6,25	1,44	520	640x360	6,25	0,36
520	640x360	6,25	0,36	544	320x180	6,25	1,51
768	320x180	12,5	1,07	768	320x180	12,5	1,07
768	640x360	12,5	0,27	768	640x360	12,5	0,27
1024	320x180	12,5	1,42	1024	320x180	12,5	1,42
1024	640x360	6,25	0,71	1024	640x360	6,25	0,71
1024	640x360	12,5	0,36	1024	640x360	12,5	0,36
1024	640x360	25	0,18	1024	640x360	25	0,18
1024	1280x720	6,25	0,18	1024	1280x720	6,25	0,18
1024	1280x720	12,5	0,09	1024	1280x720	12,5	0,09
3048	1280x720	6,25	0,53	3048	1280x720	6,25	0,53
3048	1280x720	12,5	0,26	3048	1280x720	12,5	0,26
3048	1280x720	25	0,13				
3048	1280x720	50	0,07				

III. Results and Discussion

Objective quality evaluation of the sequences from database [11] is made using 5 algorithms: VQM, MS-SSIM, PSNR, VSNR and SSIM. The set of the test materials evaluated with several objective algorithms consists of 53 sequences, 27 sequences coded with SVC codec and 26 sequences coded with WSVC codec. Since in database [11] the subjective results of the video quality evaluation are given, after linearization, Pearson's correlation coefficient between objective and subjective measurements is computed and results are presented in Table IV. For all of 53 sequences coded with different codecs, resolutions, frame rates and bit rates, highest correlation (0,83) is shown by SSIM objective algorithm. Scatter diagram for all sequences evaluated with SSIM metric is presented in Fig. 2.

In the other part of experiment coded sequences are firstly divided by codecs and then by contents, resolutions and frame rates. Correlation coefficients for each group are presented in Table V. Objective video quality evaluation for sequences coded by SVC and WSVC codecs separately and together is done. It is obvious that correlation between subjective and objective evaluation is significantly lower for both codecs together than for each of codecs separately. So, it can be concluded that using objective video quality metrics for combination of different codecs isn't suitable.

When sequences were divided by content there are different results for different codecs. Although the number of tested sequnces is relatively small, it can be concluded that sytematically neither of objective metrics gives good enough correlation. As it is presented in Table V (a), it can be seen that for DucksTakeOff sequence coded by SVC codec all metrics, except VQM, show correlation higher than 0.9 while the best result is presented for SSIM metric with correlation of 0.9351. For ParkJoy sequence there is only VQM metric with correlation higher than 0.9, while all other metrics have correlation lower than 0.8. Unexpectedly, for IntoTree sequence, sequence with the lowest spatial and temporal activity, all metrics achieve correlation lower than 0.8 while the best is MS-SSIM with correlation of 0,7945. For WSVC codec and DucksTakeOff sequence all metrics except PSNR have correlation higher than 0.8 while the best is SSIM metric with correlation of 0.8652. For ParkJoy sequence the best is VQM metric with correlation of 0,9417, while all other metrics have correlation lower than 0.7.

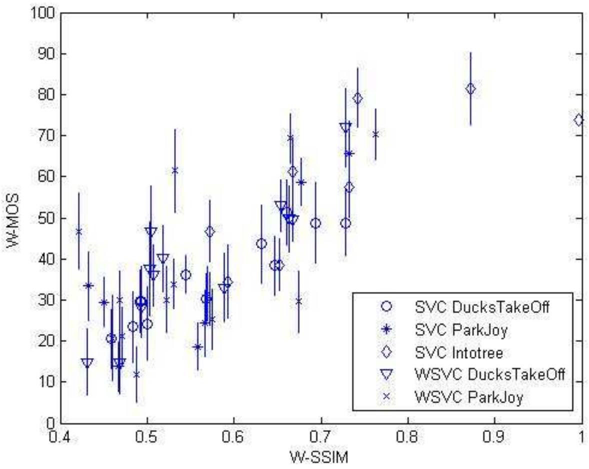


Figure 2. Scatter diagram for all sequences evaluated with SSIM algorithm

TABLE IV. PEARSON'S CORRELATION COEFFICIENT FOR ALL 53 SEQUENCES

Objective algirithm	Pearson's correlation coefficient
SSIM	0,8258
MS-SSIM	0,7995
VQM	0,7221
VSNR	0,7506
PSNR	0,6845

TABLE V. PEARSON'S CORRELATION COEFFICIENT BETWEEN SUBJECTIVE AND OBJECTIVE EVALUATIONS FOR SEQUENCES DIVIDED BY A) CONTENT B) RESOLUTION C) FRAME RATE

SVC	VQM	MS-SSIM	PSNR	VSNR	SSIM
Intotree	0,7347	0,7945	0,6449	0,643	0,7747
DucksTakeOff	0,7302	0,9032	0,9082	0,9179	0,9351
ParkJoy	0,9337	0,7421	0,6875	0,7645	0,7851
WSVC					
DucksTakeOff	0,8527	0,818	0,7291	0,8337	0,8652
ParkJoy	0,9417	0,6817	0,422	0,6278	0,5613
SVC and WSVC					
DucksTakeOff	0,8183	0,7938	0,7576	0,8246	0,8509
ParkJoy	0,9313	0,7034	0,474	0,6763	0,6431

(a)

SVC	VQM	MS-SSIM	PSNR	VSNR	SSIM
320x180	0,2723	0,918	0,8723	0,8396	0,7414
640x360	0,6749	0,7259	0,7186	0,7716	0,8458
1280x720	0,8963	0,8554	0,7644	0,7864	0,9139
WSVC					
320x180	0,4241	0,7902	0.7711	0,6279	0,9143
640x360	0,017	0,8154	0,7814	0,8265	0,7572
1280x720	0,7557	0,7765	0,6699	0,4925	0,7295
SVC and WSVC					
320x180	0,0219	0,7809	0,7346	0,6422	0,6369
640x360	0,4842	0,7585	0,7345	0,8059	0,8154
1280x720	0,8384	0,8028	0,6051	0,531	0,8087

(b)

SVC	VQM	MS-SSIM	PSNR	VSNR	SSIM
6,25	0,5138	0,4113	0,545	0,6196	0,2344
12,5	0,3746	0,7957	0,7872	0,7566	0.8370
25	0,6529	0,2782	0,7755	0,7906	0.7220
WSVC					
6,25	0,7705	0,3847	0,0971	0,2317	0,1861
12,5	0,441	0,7272	0,4601	0,6405	0,656
25	0,5401	0,787	0,6799	0,0806	0,9927
SVC and WSVC					
6,25	0,666	0,3715	0,2597	0,3851	0,134
12,5	0,3438	0,776	0,619	0,6337	0,7673
25	0,6332	0,4225	0,6127	0,6598	0,6144

(c)

Sequences are also coded on three different resolutions: 320x180, 640x360 and 1280x720. When sequences are divided by resolutions (Table V (b)) it is obvious that the best results are presented for SSIM metric with correlation from 0.7295 (WSVC 1280x720) to 0.9143 (WSVC 320x180) and MS-SSIM with correlation from 0.7259 (SVC 640x360) to

0.918 (SVC 320x180). It can be noticed that for this division PSNR metric also showed pretty good correlation with subjective results.

Since sequences are coded on three different frame rates, comparison by that criteria is also done. It can be seen that this division gives significantly lower correlation coefficients between subjective and objective measurements than any other. This is expected because for evaluation all sequences must have the original frame rate of 50 fps. To achieve this, some frames have to be repeated several times, depending on coded frame rate. It reduces the quality of evaluated materials and also correlation between objective and subjective measurements. For division by frame rate, as it is expected, the lowest correlation is presented for frame rate of 6.25 fps and only VQM metric for WSVC codec showed correlation higher than 0.7. For 12.5 fps frame rate and SVC codec the highest correlation (0.837) is achieved for SSIM metric, while for WSVC the best result is presented for MS-SSIM metric (0.7272). For 25 fps frame rate and SVC codec the highest correlation is shown for VSNR metric while for WSVC the best result is presented for SSIM metric.

Although analyzed metrics showed good performances on databases of non scalable coded video sequences, for evaluation of scalable coded video sequences neither of metrics shows consistently good results. One of the reasons for that lies in the fact that these metrics do not include temporal features of evaluated sequences.

IV. Conclusion

Since the video transmission with the best possible quality in certain moment over the time variant network is required, scalable video coding recently is increasingly used. Before the transmission, for ensuring the best Quality of Service (QoS), coded materials have to be evaluated. Scalable coded video materials are mostly subjective evaluated until now and it is also known that subjective tests are expensive and time consuming. Therefore in this paper scalable coded video sequences with known subjective evaluation results are evaluated with 5 objective algorithms. To establish which of those objective algorithms has the best match to subjective results, Pearson's correlation coefficient between subjective and objective results is computed. Measurements were done for all sequences together, and after that the sequences were divided firstly by codec and than by content, frame rate and resolution. Taking into consider all off 5 tested objective algorithms it can be noticed that the highest correlation to subjective results is presented for SSIM an MS-SSIM algorithm, and VQM is also noteworthy. Anyway, none of the analyzed metrics shows consistently good results across different contents, resolutions and frame rates. Therefore, objective quality metrics which better suit scalable coded videos should be developed in the near future.

Acknowledgment

The Croatian Ministry of Education, Science and Sports support this research through the project No. 165-0361630-1636.

References

[1] A. K. Moorthy, K. Seshadrinathan, R. Soundararajan and A. C. Bovik, "Wireless video quality assessment: A study of subjective scores and objective algorithms", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 20, No. 4, pp. 587-599, April 2010.

[2] D. M. Chandler and S. S. Hemami, "VSNR: A wavelet-based visual signal-to-noise ratio for natural images", IEEE Transactions on Image Processing, Vol. 16, No. 9, pp. 2284-2298, Sep. 2007

[3] M. H. Pinson and S. Wolf, "A new standardized method for objectively measuring video quality", IEEE Transactions on Broadcasting, Vol. 50, No. 3, pp. 312-313, Sep. 2004

[4] H. Sheikh and A.Bovik, "Image information and visual quality", IEEE Transactions on Image Processing, Vol. 15, No.2, pp 430-444, Feb. 2006

[5] Z. Wang, E. Simoncelli and A. Bovik, "Multiscale structural similarity for image quality assessment", in Conference Recommendation. 37th Asilomar Conference Signals, Systems and Computers, Vol. 2, pp 1398-1402, 2003.

[6] K. Seshadrinathan, R. Sundararajan, A. C. Bovik and L. K. Cormack, "Study of subjective and objective quality assessment of video", IEEE Transactions on Image Processing, Vol. 19, No. 6,pp. 1427-1441, June 2010.

[7] K. Seshadrinathan, R. Soundararajan, A. C. Bovik and L. K. Cormack, "LIVE Video Quality Database", Available at:
http://live.ece.utexas.edu/research/quality/live_video.html

[8] K. Seshadrinathan and A.C. Bovik, "Motion-based perceptual quality assessment of video", in [Proc. SPIE – Human Vision and Electronic Imaging], 2009.

[9] S. Chikkerur, V. Sundaram, M. Reisslein and L. J. Karam, "Objective video quality assessment methods: A classification, review and performance comparison", IEEE Transactions on Broadcasting, Vol. 57, No.2, pp. 165-181, June 2011.

[10] J. S. Lee, F. De Simone and T. Ebrahimi, "Subjective quality evaluation via paired comparison: Application to scalable video coding", IEEE Transactions on Multimedia, Vol. 13, No. 5, pp. 882-893, Oct. 2011.

[11] J. S. Lee, F. De Simone and T. Ebrahimi, "Multimedia signal processing group: Scalable video database", Available at: http://mmspg.epfl.ch/svd

[12] J. Reichel, H. Scwarz and M. Wien, Joint Sclable Video Model 11 (JSVM 11), Joint Video Team, 2007, doc. JVT-X202

[13] N. Ramzan, T. Zgaljic and E. Izquierdo, "An Efficient optimisation scheme for scalable surveillance centric video communiacations", Signal Processing: Image Communication., Vol. 24, No. 6, pp. 510-523, 2009.