Impact of H.264 Advanced Video Coding Inter-Frame Block Sizes on Video Quality

Harilaos Koumaras[1*], Michail-Alexandros Kourtis[2], Drakoulis Martakos[2] and Christian Timmerer[3]

[1]NCSR Demokritos, Institute of Informatics and Telecommunications, Aghia Paraskevi, Greece
[2]Department of Informatics, University of Athens, Athens, Greece
[3]Multimedia Communication, Alpen-Adria-Universität Klagenfurt ,Klagenfurt, Austria
** Corresponding author: koumaras@iit.demokritos.gr*

Keywords: Video Quality, AVC, SSIM, Inter-frame prediction, Block size.

Abstract: In this paper, we present a perceptual-based encoding benchmarking of the H.264 Advanced Video Coding (AVC) inter-frame prediction variable block sizes for various spatial and temporal contents. This paper in order to quantify the impact on the video quality of the AVC inter-frame variable block sizes and the responsible prediction algorithm has disabled the motion estimation mechanism of the encoder and manually each block size is selected. Thus each time only one available block size out of the total seven is available to be searched for each MB and it is possible to examine the video quality impact of each block size independently to the remaining ones. The scope of this paper is to study if the use of sophisticated predictions algorithms and variable block sizes enhance the perceived quality of the encoded video signal or if there is not any significant quality degradation when the option of variable size is disabled.

1 INTRODUCTION

Among the various video encoding standards, Advanced Video Coding (AVC)/H.264 has become the most popular and widely used in many multimedia applications and services, ranging from digital TV, mobile video, video streaming to high-definition media (ISO/IEC, 2006). AVC has been jointly developed by the ITU-T Video Coding Experts Group (VCEG) and the ISO/IEC MPEG and is currently considered as the state-of-the-art video coding standard since it is able to save up to 50% in bandwidth consumption while maintaining similar quality levels as compared to existing standards.

This enhancement of AVC in the encoding performance is the result of many new features. Although as a block-based motion-compensated predictive coder, AVC is similar to its prior standards in the general framework, however it contains significant improvements, such as variable block-size motion estimation, ¼-pel motion accuracy, allows the use of multiple reference frames, intra-frame prediction, in-loop deblocking filter and context-adaptive arithmetic coding.

In the inter-frame prediction mechanism, the most important improvement is the use of variable block sizes (shown in Fig. 1) that can be chosen dynamically for each motion-compensated macroblock in the AVC inter-frame prediction mechanisms in comparison to previous video encoding standards. Partitions with luminance block sizes of 16×16, 16×8, 8×16, or 8×8 samples, called the macroblock types or M types (Sullivan and Wiegand, 2005) are mainly used for low dynamic video and homogeneous contents, while smaller blocks intend to better characterize the motion behavior of high dynamic and heterogeneous contents. The 8×8 block size in AVC consists of additional syntax elements for its further division into smaller blocks of 8×4, 4×8, or 4×4 luma samples, called the sub-macroblock types or 8×8 types. There is also a special case of the 16×16 block, which is called SKIP mode, where the best reference frame, motion vector and transform coefficients are identical to the predicted values. The innovative use of variable block size in the inter-frame prediction mechanisms of AVC reduces the prediction error and enhances the encoding efficiency due to the higher precision of the motion vector representation.

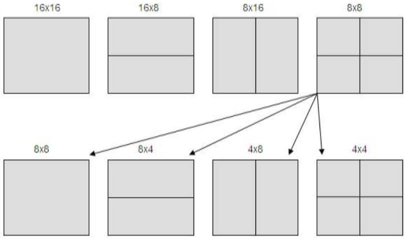


Figure 1: AVC Variable Block Sizes.

In this paper, we benchmark the video quality impact of each AVC/H.264 inter-frame prediction block size for various spatial and temporal contents. This study follows a discrete methodology, meaning that the testing and evaluation of each block size is performed independently and not in parallel with the remaining block sizes. For the purpose of this paper, in the motion estimation for AVC, there is each time only one available block size out of the total seven to be searched for each MB. By this way this paper researches if the sophisticated inter-frame predictions algorithms and variable in size blocks have an impact on the deduced video quality or if the expected perceptual enhancement is negligible. If the perceptual performance of the variable block sizes and motion compensation algorithms is very low then it means that the required high processing power for the execution of such algorithms in the encoding process is practically wasted without significant perceptual outcome.

The remainder of this paper is organized as follows. Section 2 performs a brief description of the inter-frame prediction, Section 3 describes the video quality metric that used in this paper, Section 4 presents the test signals of the experimental section, Section 5 discusses the evaluation results, and, finally, Section 6 concludes this paper.

2. INTER-FRAME PREDICTION

For completeness of the paper, this section provides some basic background information for the inter-frame and motion compensation algorithm that is implemented in the AVC reference encoder.

Motion compensation is an algorithmic technique employed in the encoding of video data for video compression, which describes a frame in terms of the transformation of a reference frame to the current picture. Variable block-size motion compensation is the use of block motion compensation with the ability for the encoder to dynamically select the size of the used blocks. Thus, the use of larger blocks can reduce the number of bits needed to represent the motion vectors (better compression), while the use of smaller blocks can result in a smaller amount of prediction residual information to encode (better prediction). Older designs such as H.261 and MPEG-1 video typically use a fixed block size while newer ones such as AVC give the encoder the ability to dynamically choose which block size fits better to a specific region. The overall procedure, which practically exploits the temporal redundancy of a video sequence for compression purposes, is called inter-frame prediction. In the motion estimation for AVC, there are in total seven possible block sizes to be searched for each MB (modes 1–7 denote block sizes of 16 × 16, 16 × 8, 8 × 16, 8 × 8, 8 × 4, 4 × 8, and 4 × 4, respectively).

The simplest algorithmic implementation for the inter-frame prediction is the Full Search (FS) algorithm, which checks every displacement inside the designated search window in order to specify the best block size out of the seven available. The FS algorithm, which evaluates Mean Absolute Difference (MAD) at all possible regions of a frame, has very high computational requirements, making necessary the development of most sophisticated algorithms providing a better trade-off between computational complexity and prediction efficiency.

In this paper, we have modified the inter-frame prediction mechanism in the reference encoder of AVC/H.264 in order to perform the search and match as it normally does, but each time looking for the better match of only one specific block size (out of the total seven available). Thus, the inter-frame prediction continues to run but modified for a fixed block size each time. So practically with this modification the processing power consumption and requirements have been reduced to the minimum.

The scope is to examine the perceptual impact of the inter-frame prediction algorithms in conjunction with variable in size blocks in relevance to the spatiotemporal dynamics of the content.

3. VIDEO QUALITY ASSESSMENT

The evaluation of the video quality is a matter of objective and subjective procedures, which take place after the encoding process. Subjective quality evaluation processes of video streams require large amount of human resources, establishing it as a time-consuming process (Pereira and Alpert, 1997), (ITU, 2000). Objective methods, on the other hand, can provide perceived QoS evaluation results faster, but require sophisticated apparatus configurations.

The majority of the existing objective methods requires the undistorted source video sequence as a reference entity in the quality evaluation process, and due to this, these methods are characterized as Full Reference (FR) (Tan and Ghanbari, 2000), providing a good benchmarking tool.

For this reason, in order to quantify the perceptual difference between the different block sizes, the use of objective instead of subjective procedures was preferred. Since the perceptual difference between the seven available block sizes is expected to be rather small, the subjective assessments could not be able to provide a reliable result due to the relative high statistical error (Pinson and Wolf, 2003), which will encompass the corresponding evaluation. Keeping this restriction in mind, we decided to use the FR SSIM metric as an objective metric which will benchmark the encoding efficiency of different block sizes in relevance to the spatiotemporal activity level of the video content. SSIM is a FR metric for measuring the structural similarity between two image sequences, exploiting the general principle that the main function of the human visual system is the extraction of structural information from the viewing field. If x and y are two video frames, then the SSIM is defined in Equation (1):

$$SSIM(x,y)=\frac{(2\mu_x\mu_y+C_1)(2\sigma_{xy}+C_2)}{(\mu_x^2+\mu_y^2+C_1)(\sigma_x^2+\sigma_y^2+C_2)} \qquad (1)$$

Where μ_x, μ_y are the mean of x and y, σ_x, σ_y, σ_{xy} are the variances of x, y and the covariance of x and y, respectively. The constants C_1 and C_2 are defined in Equation (2):

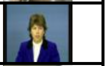
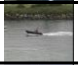
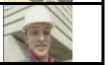
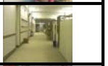
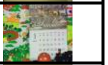
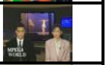
$$\begin{aligned} C_1 &= (K_1L)^2 \\ C_2 &= (K_2L)^2 \end{aligned} \qquad (2)$$

Where L is the dynamic pixel range and $K_1 = 0.01$ and $K_2 = 0.03$, respectively (Wang and Lu, 2004), (Wang and Bovik, 2004b).

4. TEST VIDEO SIGNALS

In order to examine the perceptual efficiency of each block size, considering identical rest encodings settings, we use 11 reference test signals of various spatiotemporal activity levels.

Table 1: Test Signals

Signal	Frames	
Akiyo	300	
Suzzie	150	
Claire	494	
Carphone	382	
Coastguard	300	
Container	300	
Foreman	300	
Hall	300	
Mobile	300	
News	300	
Silent	300	

The test signals used in this paper are of QCIF resolution and their data and spatiotemporal activity level are presented in the Table 1.

As it can be observed, the selected test signals cover different range of the spatiotemporal plane, are of QCIF spatial resolution and 25 fps of temporal resolution. The next section presents the results of the experimental section.

5. EVALUATION RESULTS

In order to examine the perceptual efficiency of each block size, the aforementioned 11 reference test signals were used as an input in the reference AVC encoder, which was properly modified to use each time only one specific block size in the inter-frame prediction mechanism.

The encoding parameters were set to be similar to the original source file in terms of spatial and temporal resolution, while the encoding bit rate was selected to be variable with the highest possible quantization parameter in order to produce the best possible encoding result, without considering the encoding efficiency. Afterwards each encoded signal was used along with the original source file in the SSIM algorithm in order evaluate the deduced video quality level for each block size as an average for the whole in duration video signal. The results of the procedure are depicted in Figure 2.

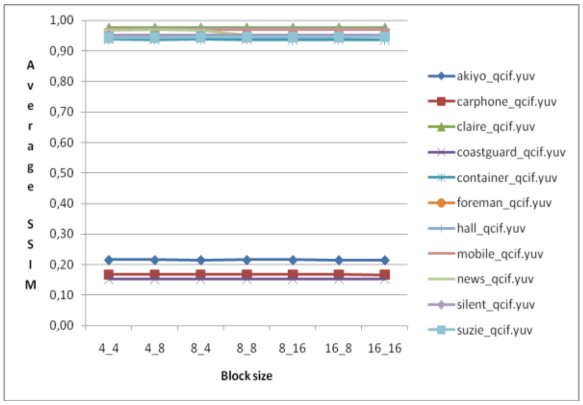


Figure 2: Average SSIM curves per block size and signal.

According to the depicted results of Figure 2, it can be deduced two important observations: i) for a specific video content (i.e., test signal) the variation of the deduced video quality is not significant for each block size, and ii) for specific video contents (mainly low dynamic ones) it is noticed that the limitation of the block sizes in the inter-frame prediction process causes significant perceptual degradation.

Considering the first observation and in order to examine thoroughly the perceptual difference among the different block sizes, we provide the arithmetic results of the evaluation procedure in Table 2

Table 2: Average SSIM values per block size and signal

	Inter block search mode						
Average SSIM	4_4	4_8	8_4	8_8	8_16	16_8	16_16
akiyo_qcif.yuv	0,215950	0,215780	0,215540	0,215850	0,215990	0,215570	0,215180
suzie_qcif.yuv	0,942910	0,942910	0,944110	0,945040	0,945730	0,945590	0,946240
claire_qcif.yuv	0,975250	0,975870	0,975660	0,976130	0,975820	0,975630	0,975210
carphone_qcif.yuv	0,167450	0,167240	0,167560	0,167310	0,167360	0,167510	0,167000
coastguard_qcif.yuv	0,152820	0,152970	0,152910	0,153090	0,152850	0,153440	0,152810
container_qcif.yuv	0,937450	0,937310	0,937520	0,937200	0,936420	0,937070	0,936380
foreman_qcif.yuv	0,950940	0,951800	0,952180	0,952730	0,952730	0,951970	0,952040
hall_qcif.yuv	0,971720	0,971480	0,971830	0,971650	0,971200	0,970980	0,971130
mobile_qcif.yuv	0,970550	0,970050	0,970410	0,969880	0,969340	0,969170	0,969110
news_qcif.yuv	0,968150	0,968290	0,968150	0,951880	0,951140	0,950830	0,950850
silent_qcif.yuv	0,952300	0,951860	0,952090	0,951880	0,951140	0,950830	0,950850
Average SSIM	0,745954	0,745960	0,746178	0,744785	0,744520	0,744417	0,744255

Table 2 demonstrates the slight difference in the perceptual performance due to different block size.

Figure 3 provides a graphical representation of the arithmetic data listed in Table 2. Based on the depicted results, it can be observed that the perceptual impact of the variable block size selection is content dependent, while the fluctuation of the quality among the block sizes in significantly low in all cases. Although it was expected that the use of smaller block sizes would generate better perceptual results, but inefficient compression, however for specific video signals (i.e., Suzie, Coastguard, Foreman), a reverse analogous behavior is noticed. More specifically, it is observed that for these signals the video quality is enhanced when bigger blocks are used and not small ones.

Based on these results, the suggestion that the spatiotemporal activity of the content should be considered as an input in the inter-frame prediction and motion compensation algorithms is further supported towards next generation power efficiency encoders with low carbon footprint.

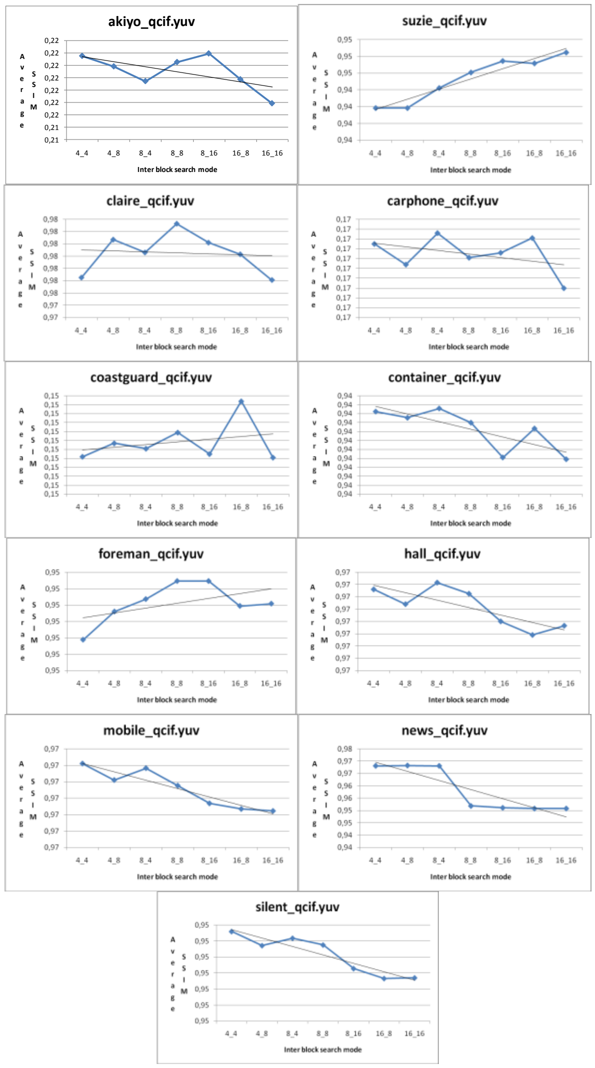


Figure 3: Detailed Curves per Block Size and Signal.

Moreover, it is important to be also noted that the average quality level remains at satisfactory levels for specific type of video contents even if during the encoding process is used one block size. However, for specific contents severe degradation has been observed in the one block size mode.

6. CONCLUSIONS

This paper has presented a perceptual-based encoding benchmarking of the AVC inter-frame prediction variable block sizes for various spatial and temporal contents. Preliminary results have been provided showing that the perceptual efficiency of each block size is dependent on the content dynamics. Future work includes the expansion of the experimental section using more objective metrics in order to extent the sensitivity of the measured perceptual efficiency. Also detailed measurement of the spatiotemporal dynamics will be performed in order to provide a mapping between the content dynamics and the efficiency of each block type.

ACKNOWLEDGMENTS

The authors would like to thank A. Papaioannou and Ch. Strologa for their contribution at the experimental part of this paper. This work was performed in the framework of the ALICANTE project (FP7-ICT-248652).

REFERENCES

ISO/IEC, 2006. *International Standard 14496-10, Information Technology – Coding of Audio-Visual Objects – Part 10: Advanced Video Coding*, third ed.

Sullivan G., Wiegand, T., 2005. *"Video compression – from concepts to the H.264/AVC standard"*. In proceedings of the IEEE, vol. 93, no. 1, pp. 18-31.

Pereira, F., Alpert, T., 1997. *"MPEG-4 Video Subjective Test Procedures Results"*. In IEEE Trans. on Circuits and Systems for Video Tech. Vol.7(1), pp.32-51.

ITU 2000. *"Methodology for the subjective assessment of the quality of television pictures"*. In Recommendation ITU-R BT.500-10.

Tan, K., Ghanbari, M., 2000. *"A Multi-Metric Objective Picture Quality Measurements Model for MPEG Video"*. In IEEE Transactions on Circuits and Systems for Video Technology, Vol.10(7), pp. 1208-1213.

Pinson M., Wolf, S, 2003. *"Comparing subjective video quality testing methodologies"*. In Proceedings of the SPIE Visual Communications and Image Processing, Vol 5150, pp. 573-582.

Wang, Z., Lu, L., Bovik, A., 2004. *"Video Quality Assessment Based on Structural Distortion Measurement"*. In Image Communication, Vol. 19(2), pp. 121-132.

Wang, Z., Bovik, A. C., Sheikh H. R., Simoncelli, E. P., 2004. *"Image quality assessment: From error visibility to structural similarity"*. In IEEE Transactions on Image Processing, Vol. 13(4), pp. 1-14.