Comparison of H.264/AVC, H.264 with AIF, and AVS based on different video quality metrics

S. Pasqualini*, F. Fioretti*, A. Andreoli* and P. Pierleoni*
*Dipartimento di Ingegneria Biomedica, Elettronica e Telecomunicazioni Università Politecnica delle Marche
Via Brecce Bianche 1, 60131 Ancona (Italy)
Phone/Fax number:+0039 071 2204847,
e-mail: s.pasqualini@univpm.it, f.fioretti@univpm.it, a.andreoli@univpm.it, p.pierleoni@univpm.it

***Abstract*—Video communication has a very important role nowadays. Also very efficient video coding techniques are developed in order to provide high quality video streams using all the available bandwidth. In this work we perform a comparison between three different video codecs: the ITU-T H.264, an improved ITU-T H.264 version including the AIF (Adaptive Interpolation Filter), and the AVS (Audio and Video Coding Standard) that is the Chinese standard video codec. The realized tests involve eight standard video sequences that are coded end decoded using the mentioned standards. Then, for each sequence we evaluate the video quality using four objective metrics: PSNR (Peak Signal-to-Noise Ratio), VQM (Video Quality Metric), SSIM (Structural SIMilarity), and the HSM (HVS - Human Visual System - Similarity Measure). While PSNR and VQM take into account the mathematical features of the video sequences, SSIM and HSM are based on the HVS theory. So it is possible to evaluate also the human eye perceived quality. The realized analysis is related to the relationship between the bitrate and the quality indexes. The conclusions of our analysis can give an evaluation about several codec performances and can be useful for developers to plan and manage their network. Also, taking into account the HSM scores, the AVS has better performance than the ITU-T video codecs.**

***Index Terms*—Adaptive Interpolation Filter, AVS, H.264, HSM, PSNR, SSIM, VQM**

I. INTRODUCTION

Telecommunication developments brought to an important diffusion of multimedia services in which the video transmission assumes a very important role. In all real situations it is necessary to introduce video compression techniques that give a good trade off between transmissive resources and perceived video quality after the co-decoding operations. H.264\AVC [1] is the standard able to cover different applications thanks to its several configuration parameters. In this article we compare this codec with the under develop ITU-T new features for H.264, the Adaptive Interpolation Filter [2][3], and the Chinese AVS codec [4].

This article examines and describes H.264/AVC, Adaptive Interpolation Filter in H.264, and AVS which represent video coding standard already established in the sector of telecommunications. The PSNR, VQM [5], SSIM [6], and a new video quality parameter HSM [7][8] methods have been used for the analysis of the three encoding techniques.

Several studies have been presented with the comparison between video codec performances [9][10], now also considering the new Chinese codec [11]. These works always consider only one video quality metric applied to the comparison of two video codec standards, or they do not consider differences in quality indexes based on perceived quality.

This paper is organized as follow: section II describes the main features of all the codec involved in the tests. Section III gives an overview on all the metrics used to evaluate video quality, mainly on a new technique, the HSM. Settings and conditions for tests are described in section IV, while test results are examined in section V. Finally conclusions are explained in section VI of this paper.

II. UNDER STUDY VIDEO CODECS

A. H.264

H.264/AVC standard was developed and standardized by the ITU-T Video Coding Experts Group (VCEG) and ISO/IEC Moving Experts Group (MPEG). It shows great improvements in compression efficiency and error robustness in comparison with old video standards such as MPEG-4 Part 2, MPEG-2, H.263 and H.261 for a number of applications ranging from mobile services, videoconferencing, HD broadcasting, and hard disk storage. Essentially, it uses block based spatial prediction, motion estimation and compensation for the reduction of temporal redundancy and residual entropy coding.

Allowed macroblock divisions are 16x16, 16x8, 8x16, 8x8, 8x4, 4x8, and 4x4 block size for Intra and Inter prediction. There are 13 possible prediction modes for Intra prediction process and the accuracy of the motion vectors in Inter prediction reaches the 1/4 of pixel resolution. The maximum number of reference frames are 16.

In-loop deblocking filtering function compares samples near vertical and horizontal block boundaries and smooths the block edges if necessary. It has great significance because it reduces blocking distortion while preserving image structure, object characteristics and subjective quality.

Redundant picture is another important feature of the codec. It represents part or entire coded picture which is discarded in normal operation. This functionality is very useful in case of missing or damaged primary pictures normally used for reconstruction in decoder. Therefore redundant picture can replace damaged or missing primary picture.

Using Arbitrary Slice Order (ASO) means that slices inside video frame could not be in precise decoding order. If first macroblock in a slice of decoded frame has a smaller address

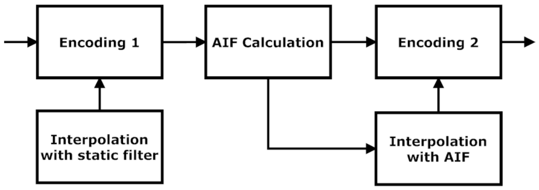

Figure 1. Flowchart of video encoder with adaptive interpolation.

comparing to first macroblock in a previously decoded slice of the same frame, ASO is in use.

Flexible Macroblock Ordering (FMO) allows grouping of macroblocks to certain slices inside the frame.

For entropy coding of the bit stream syntax H.264/AVC uses advanced functionality called Context-Based Adaptive Variable Length Coding (CAVLC).

B. H.264 with AIF

The interpolation filter defined in H.264/AVC is designed to minimize the effects of aliasing present in the input image sequence. This algorithm requires two encoding passes for each P- or B-frame as illustrated in Fig. 1. In the first coding pass (Encoding 1), the current frame is coded with the reference frame interpolated with static filter. Motion Estimation (ME) and rate distortion optimization (RDO) methods are performed during this pass and motion prediction information are found. Then encoder calculates the coefficients of the adaptive filter using this information, by solving system of linear equations that minimize the prediction error energy. Finally, frame is re-encoded including ME and RDO using the computed adaptive interpolation filter. It is clear from Fig. 1 that this algorithm bring to an increase in complexity compared to encoding with static filter. This is due to three additional steps: the AIF calculation steps, the interpolation of the reference frame with the new filter, and the second encoding pass.

C. AVS

AVS video standard is developed by the audio video coding standard working group of China (AVS working group in short), which was approved by the Chinese science and technology department of ministry of information industry in June 2002.

Fig. 2 depicts the block diagram of AVS video encoder. Due to the same modules, the architecture of AVS looks like the H.264/AVC encoder. However, considering the target applications, backward compatibility with MPEG-2 and decoding complexity, the technique of the AVS video codec in every module is more and less different from that used in H.264/AVC.

The Intra prediction is based only on 8x8 block, there are five luminance Intra prediction modes, and four chrominance Intra prediction modes. There are four macroblock partition types for Inter prediction, 16x16, 16x8, 8x16 and 8x8, and the maximum number of reference frames is 2. The resolution of the motion vectors in Inter prediction is 1/4 pixel. The unit size of the of spatial prediction is 8x8 because of 8x8 integer transform.

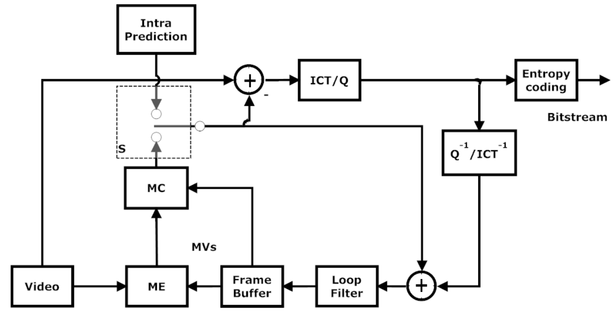

Figure 2. AVS encoder block diagram.

III. Methods used for video quality assessment

To determine the level of compression used in encoding process and its influence on video quality, different methods for video quality assessment are used. Furthermore, quality evaluation and comparison of encoded video content is a very difficult process because quality of certain video sequences is determined by subjective experience of viewer. Currently, contributing efforts are being made to develop and standardize objective measurement method that would align with subjective measurement methods. The VQEG group and ITU-T organization have developed recommendations for the objective measurement of video quality [12]. For our codec tests PSNR, VQM, SSIM, and HSM objective quality measurement methods have been used. All four quality assessment methods belong to full reference method.

A. PSNR

Peak Signal to Noise Ratio is the classical and more used objective measurement parameter expressed as a logarithmic ratio of maximum amplitude of picture pixels and Mean Squared Error (MSE). PSNR is defined as:

$$PSNR = 20 \log_{10} \frac{(2^n - 1)^2}{\sqrt{MSE}} \tag{1}$$

Picture elements are in the range from 0 to $2^n - 1$ where n is number of quantized bits per pixel of original picture. MSE is defined as:

$$MSE = \frac{1}{M * N} \sum_{i=0}^{M-1} * \sum_{j=0}^{N-1} (x_{ij} - y_{ij})^2 \tag{2}$$

where $M * N$ is frame size of original video signal. x_{ij} and y_{ij} represents original frame elements while a ij represents compressed frame element indexes. MSE is squared difference between original picture and compressed picture.

B. VQM

The VQM model [5] contains seven independent parameters. Four parameters are based on features extracted from spatial gradients of the Y luminance component such as $si\_loss, hv\_loss, hv\_gain, si\_gain$, two parameters are based on features extracted from the vector formed by the two (Cb, Cr) chrominance components such as $chroma\_spread, chroma\_extreme$, and one parameter is

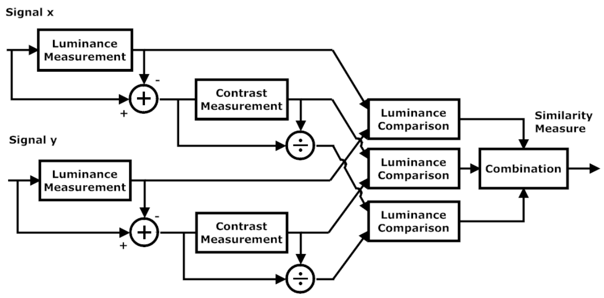

Figure 3. Diagram of the structural similarity (SSIM) measurement system.

based on the product of features that measure contrast and motion, both of which are extracted from the Y luminance component such as ct_ati_gain.

Equation 3 show the optimization of general VQM model to achieve maximum objective to subjective correlation using a wide range of video quality and bitrate. So VQM score consists of a weighted linear combination of the seven parameters listed above.

$$\begin{aligned} VQM \; = \; & -0.2097 * si\_loss \\ & +0.5969 * hv\_loss \\ & +0.2483 * hv\_gain \\ & +0.0192 * chroma\_spread \\ & -2.3416 * si\_gain \\ & +0.0431 * ct\_ati\_gain \\ & +0.0076 * chroma\_extreme \end{aligned} \tag{3}$$

The only parameter that can decrease VQM is si_gain while all the others increase the VQM (si_loss is always less than or equal to zero).

VQM computed in the above manner will have values greater than or equal to zero (no perceived impairment) and a nominal maximum value of one (maximum perceived impairment). VQM may exceed one for video scenes that are extremely distorted.

C. SSIM

SSIM is objective video quality assessment method which compares information about luminance, contrast and structural similarity between original and processed picture [6]. The luminance of objects being observed is the result of the illumination and the reflectance, but the structures of the objects in the scene are independent of the illumination. Consequently, we wish first separate the influence of the illumination to explore the structural information in an image. Structural information of an image are the attributes that represent the structure of objects in the scene, regardless of the average luminance and contrast. Since luminance and contrast can vary across a scene, we use the local luminance and contrast. The system diagram of the proposed quality assessment system is shown in Fig. 3. For example, consider two image signals, which have been aligned with each other (e.g., spatial patches extracted from each image). If we evaluate one of the signals to have perfect quality, then the similarity measure can serve as a quantitative measurement of the quality of the second signal. The system separates the task of similarity measurement into three comparisons: luminance, contrast and structure. First, the luminance of each signal is compared. Assuming discrete signals, this is estimated as the mean intensity.

$$\mu_x = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{4}$$

We use the standard deviation (the square root of variance) as an estimate of the signal contrast. An unbiased estimate in discrete form is given by

$$\sigma_x = \sqrt{\left(\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\mu_x)^2\right)} \tag{5}$$

The contrast comparison $c(\mathbf{x},\mathbf{y})$ (Equation 8) is then the comparison of σ_x and σ_y. Third, the signal is divided by its own standard deviation, so that the two signals being compared have unit standard deviation. The structure comparison $s(\mathbf{x},\mathbf{y})$ is conducted on these normalized signals $(\mathbf{x}-\mu_x)\setminus\sigma_x$ and $(\mathbf{y}-\mu_y)\setminus\sigma_y$.

Finally, three components are combined to yield an overall similarity measure:

$$S\left(x,y\right) = f\left(l(x,y)\cdot c(x,y)\cdot s(x,y)\right) \tag{6}$$

Where for luminance comparison is given by:

$$l(x,y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2+\mu_y^2+C_1} \tag{7}$$

The contrast comparison function takes a similar form:

$$c(x,y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2+\sigma_y^2+C_2} \tag{8}$$

The structure comparison function as follows:

$$s(x,y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \tag{9}$$

The constants C_1, C_2, C_3 are introduced to avoid instability of the functions. It is possible to adjust the relative importance of the three components, in fact:

$$SSIM(x,y) = [l(x,y)]^{\alpha}[c(x,y)]^{\beta}[s(x,y)]^{\gamma} \tag{10}$$

where parameters $\alpha,\beta,\gamma > 0$.

D. HSM

HSM is an objective method to assess the video quality based on the intensive use of HVS theory [7][8]. This method does not evaluate video quality comparing pixel of original and coded frames, but it takes into account of the psycho-visual basis: it considers spatial and temporal correlation described into the HVS. This algorithm can be split in several functional blocks as shown in Fig. 4.

After the acquisition step, the original and the distorted video sequences are both transformed into the so called PCD

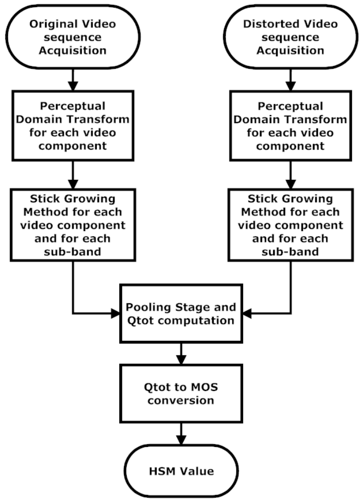

Figure 4. HSM algorithm flow diagram.

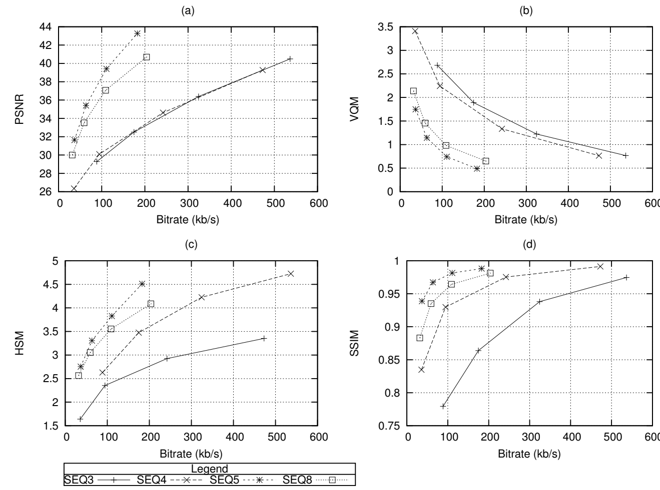

Figure 5. Quality evaluation graphs for H.264 codec: (a) PSNR scores vs bitrate; (b) VQM scores vs bitrate; (c) HSM scores vs bitrate; (d) SSIM scores vs bitrate.

Table I
SEQUENCES UNDER TEST

Sequence
SEQ1 - City
SEQ2 - Crew
SEQ3 - Football
SEQ4 - Harbour
SEQ5 - Ice
SEQ6 - Mobile
SEQ7 - Soccer
SEQ8 - Foreman

(Perceptual Channel Domain). This domain is very close to the human eye image representation; in fact, it is based on two different kinds of filters: a spatial and a temporal one, both described into the HVS theory. Regarding the spatial filter CSF (Contrast Sensitivity Functions) are used. These curves explain the human eye sensitivity for targets characterized by different spatial frequencies measured in cycles per degrees (cpd). The curves show that the human eye acts as a pass-band filter, and it is more sensitive for targets at 4 cpd. The temporal filtering has a low-pass trend: all the high frequencies variations are discarded by the human eye and so they must not be considered into the quality evaluation.

Then the structural features of each video sequences and of each spatial sub-band (derived from the CSF spatial filtering application) are extracted using the Stick Growing Method and used to compute the similarity functions. Then, a pooling stage allows the merge of all the functions into an unique parameter called Q_{tot}. Finally, a Q to MOS regression curve is applied in order to produce the HSM score. In this way, the scores are bounded into a scale from 1 to 5 as the subjective scores, where 1 denotes the worst quality and 5 an excellent video sequence.

IV. TEST

According to [13] we used the suggested sequences and parameters for our test settings. So the used sequences are listed in Table I, the QCIF (Quarter Common Intermediate Format) frame format (176x144) has been used. Framerate of all the sequences is 25 frame per second. All sequences have been coded with the three encoders described in Section II, namely H.264, H.264 with AIF and the Chinese codec AVS. The coded bitstream structure is IPPP and, as detailed in [13], we used fixed QP values for our tests. We used four different QP couples, that is (22,23), (27,28), (32,33), and (37,38). The first number is the value of the quantization parameter used for I frames, the second value enclosed in parenthesis has been used for P frames. Only the first frame has been coded as I frame. Two are the number of reference frames used by all the three codecs. Performances are evaluated in term of rate-distortion according to the indexes of the four different metrics in section III for all the codec presented in section II.

V. EXPERIMENTAL RESULTS AND ANALYSIS

This section shows the experimental results involving the video sequences produced following the procedure described in section IV. Two kinds of results will be taken into account: first, four quality indexes are evaluated for each video codec, whereas in the the second we consider the HSM scores in order to provide a comparison between the three video codecs based on perceived quality evaluations.

In Fig. 5 the quality curves for the H.264 codec are shown. As explained in the previous sections the graphs (a) and (b) refer to two objective methods that consider the variations between mathematical features of the two sequences. On the other side, the plots (c) and (d) involve two objective methods based on HVS. In fact, the quality for SEQ 3 and SEQ 4 is almost the same regarding PSNR and VQM scores, while the same sequences have different quality levels using HSM and SSIM algorithms. This is due to the intrinsic features of the under study sequences: the spatial and temporal filtering used in HSM and SSIM lead to an analysis focused on targets where the human eyes are more sensitive.

The same analysis can be performed for the other two codecs (see Fig. 6 and 7). All the metrics show the same trend of the H.264 codec; this means that quality evaluation is consistent for all the four indexes.

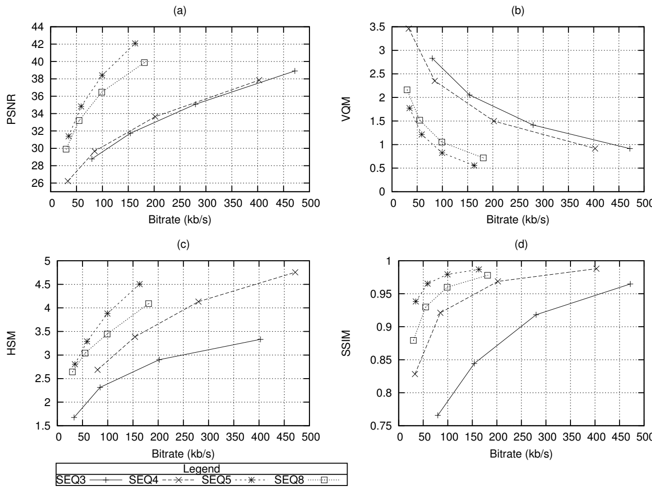

Figure 6. Quality evaluation graphs for H.264 with AIF codec: (a) PSNR scores vs bitrate; (b) VQM scores vs bitrate; (c) HSM scores vs bitrate; (d) SSIM scores vs bitrate.

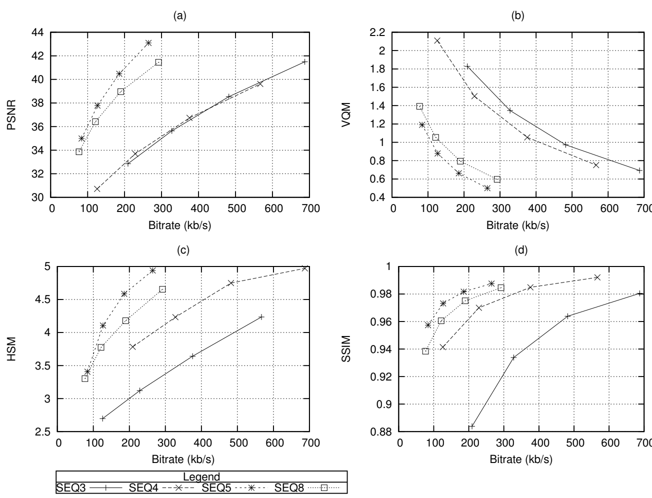

Figure 7. Quality evaluation graphs for AVS codec: (a) PSNR scores vs bitrate; (b) VQM scores vs bitrate; (c) HSM scores vs bitrate; (d) SSIM scores vs bitrate.

Other relevant issues follow the analysis of the Fig. 8 where the remaining four sequences of Table I are plotted. Considering all the eight sequences analyzed, we notice that for bitrate below 100 kb/s the three codecs produce almost the same video quality. Instead for higher bitrate, AVS codec has better performance due to higher quality scores. For instance, looking at Fig. 8 (a), a HSM score of 3.5 requires around 150 kb/s regarding H.264, while H.264 with AIF needs about 135 kb/s, and AVS only 125 kb/s. The same behavior can be observed in Fig. 8 (b) , (c), and (d): given a fixed value of bitrate the AVS codec produces the highest quality score, whereas H.264 and H.264 with AIF work at the same manner. Moreover for a fixed quality level the AVS codec requires the lowest bitrate relating to H.264 and H.264 with AIF. This is due to the fact that this version of AVS has been developed to optimized TV stream, that normally require a lot bandwith, so for high bitrate its coding efficiency is better than other video standards. We can notice also that at the moment the version

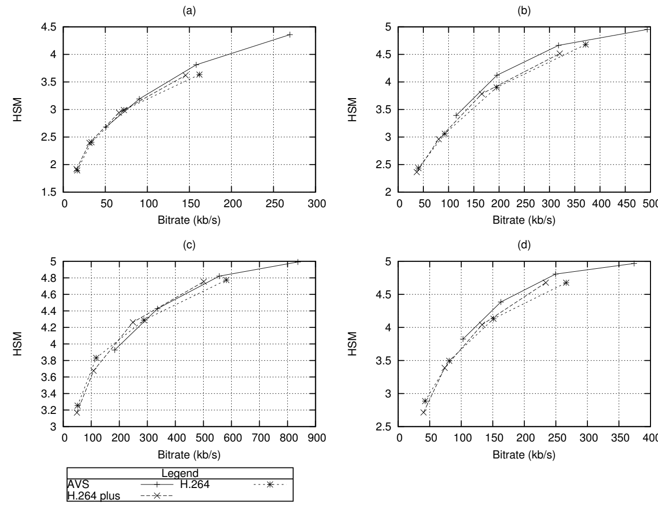

Figure 8. HSM quality scores comparison for different test sequences: (a) for SEQ 1; (b) for SEQ 2; (c) for SEQ 6; (d) for SEQ 7.

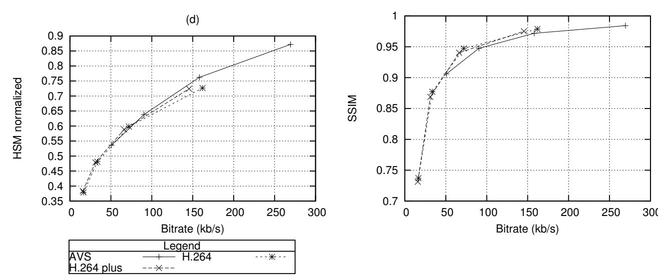

Figure 9. HSM normalized versus SSIM for SEQ 1.

of H.264 with AIF does not provide enough improvement to justify the birth of a new ITU-T video codec standard.

We can now consider Fig. 9 and 10 that show the perceived quality using HSM and SSIM algorithms for all the under analysis codecs. The aim of these plots is to provide a performance comparison among the three standards by using a non standard video quality index, that is HSM and investigate features of indexes that consider the perceived quality. Although the range of HSM value is between 0 and 5, we have normalized these values to allow an easy comparison with SSIM index. For all the tested sequences, we can observe a significant increase on the perceived quality of the SSIM index for low encoding rate range. A less perceptible increase is observed in the middle bitrate range, untill an imperceptible increase is detected beyond a threshold. This is a typical behaviour of indexes that consider perceived quality, that is to reach a saturation point in quality evaluation. For HSM this trend is not so evident and this can be part of our future work.

VI. Conclusion

This paper compares the performances of three of the most innovative video codecs: the H.264, the H.264 with AIF and the AVS. The performances are evaluated taking into account four different video quality metrics: PSNR, VQM, HSM, and

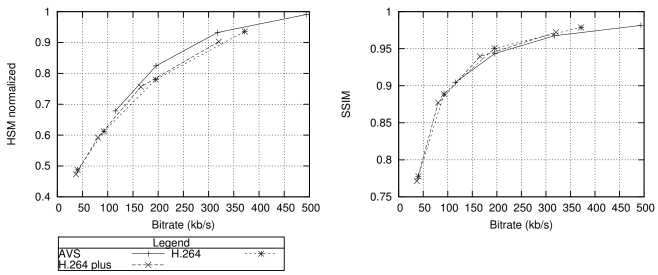

Figure 10. HSM normalized versus SSIM for SEQ 2.

SSIM. These four algorithms belong to the objective quality methods category. The obtained scores take into account the mathematical features of the video sequences in the case of PSNR and VQM, such as energy losses, chrominance spread, gains. The scores obtained by HSM and SSIM algortihms are more related to subjective tests because they are based on the HVS theory.

The tests are executed according to the ITU reference documents and the results of the codec benchmark are plots of quality scores in function of the required bitrate. We considered the HSM scores produced by the three codecs relating to three different video sequences. In the presented analysis we noticed that at lower bitrate the three video act at the same manner, whereas at higher bitrate these plots lead us to assert that AVS performs better than the other two ITU-T codecs because for a fixed quality level requires a lower bitrate, and for a fixed bitrate it produces the highest quality level. This is the result of this version of AVS design project; in fact it has been developed for HDTV, so it is optimized to work fine with large bandwith. Also our tests are performed using the baseline profile of the ITU-T video codecs, we can also assert that for video communication all the three codecs can be used and they produce satisfactory quality levels, whereas if an higher bitrate is available AVS codec is the best choice, producing high quality video sequences. Improvements of H.264 with AIF are not sufficient at the moment for the birth of a new ITU-T video standard, and finally, authors would give an overview on the most recent video codec standard to provide developers a useful aid in planning and managing thier development projects.

REFERENCES

[1] ITU-T Video Coding Experts Group (VCEG) and ISO/IEC Moving Picture Experts Group (MPEG), "Recommendation H.264 - advanced video coding for generic audiovisual services," March 2005.

[2] D. Rusanovskyy, K. Ugur, and M. Gabbouj, "Fast encoding algorithms for video coding with adaptive interpolation filters," *Multimedia Signal Processing, 2008 IEEE 10th Workshop on*, pp. 317–321, Oct. 2008.

[3] D. Rusanovskyy, K. Ugur, M. Gabbouj, and J. Lainema, "Video coding with pixel-aligned directional adaptive interpolation filters," *Circuits and Systems, 2008. ISCAS 2008. IEEE International Symposium on*, pp. 704–707, May 2008.

[4] L. Fan, S. Ma, and F. Wu, "Overview of AVS video standard," *Multimedia and Expo, 2004. ICME '04. 2004 IEEE International Conference on*, vol. 1, pp. 423–426 Vol.1, June 2004.

[5] M. Pinson and S. Wolf, "A new standardized method for objectively measuring video quality," *Broadcasting, IEEE Transactions on*, vol. 50, no. 3, pp. 312–322, Sept. 2004.

[6] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *Image Processing, IEEE Transactions on*, vol. 13, no. 4, pp. 600–612, April 2004.

[7] M. Carnec and D. Barba, "Simulating the human visual system: towards objective measurement of visual annoyance," *Systems, Man and Cybernetics, 2002 IEEE International Conference on*, vol. 6, pp. 6 pp. vol.6–, Oct. 2002.

[8] A. Andreoli, F. Fioretti, S. Pasqualini, P. Pierleoni, and S. Tumini, "A new objective parameter for video quality assessment," *E-Activity and Leading Technologies*, Madrid, Spain, 2008.

[9] G. Gvozden, M. Gosta, and S. Grgic, "Comparison of H.264/AVC and MPEG-4 ASP coding techniques designed for mobile applications using objective quality assessment methods," *ELMAR, 2007*, pp. 51–54, Sept. 2007.

[10] J. L. Martinez, P. Cuenca, F. Delicado, and L. Orozco-Barbosa, "On the capabilities of quality measures in video compresion standards," *Electrical and Computer Engineering, 2006. CCECE '06. Canadian Conference on*, pp. 527–532, May 2006.

[11] X.-F. Wang and D.-B. Zhao, "Performance comparison of AVS and H.264/AVC Video Coding Standards," *J. Comput. Sci. Technol.*, vol. 21, no. 3, pp. 310–314, 2006.

[12] "Final report from the video quality experts group on the validation of objective models of video on the validation of objective models of videoquality assessment, phase II," August, 2003.

[13] T. Tan, G. Sullivan, and T. Wedi, "Recommended simulation common conditions for coding efficiency experiments," *ITU-T Q.6/SG16, 34th VCEG Meeting*, Antalya, Turkey, 2008.