# H.264 Rate-Distortion Analysis Using Subjective Quality Metric

Luís Teixeira[1] and Luís Corte-Real[2]

[1] UCP - E.Artes / CITAR, R. Diogo Botelho 1327,
4169 005 Porto, Portugal
lteixeira@porto.ucp.pt
[2] FE.UP / INESC Porto, R. Roberto Frias,
4250 005 Porto, Portugal
lreal@inescporto.pt

**Abstract.** In this paper we provide an analysis of rate-distortion (R-D) relationship in an H.264 codec using as quality metric Structural Similarly Information (SSIM). This study focus on the quantization parameter, namely rate-quantization (R-Q) functions and distortion-quantization (D-Q) functions. Together, these functions allow a better understanding of the rate-distortion (R-D) behaviour of an H.264 video codec, which is the key issue of optimum bit allocation. Initial results are presented and discussed.

**Keywords:** Rate-Distortion optimization, SSIM, video quality, H.264.

## 1 Introduction

The H.264/MPEG-4 Advanced Video Coding standard (H.264/AVC) [1] is the newest video coding standard jointly developed by the ISO/IEC Moving Picture Experts Group (MPEG) and the ITU-T Video Coding Experts Group (VCEG)[1]. H.264/AVC has accomplished a substantial progress regarding coding efficiency regarding its predecessors [2]. It covers common video applications such mobile services or Internet Protocol Television (IPTV) [2].

## 2 Video Quality Assessment

One of goals in the design of visual communication systems is to represent, broadcast and reproduce the information that the human eye can see and perceive. The most reliable way of assessing the quality of an image or video is subjective evaluation, because human beings are the final receivers in most applications [7]. Nevertheless, subjective evaluation is too complex (wide variety of possible methods and test elements) and provides too much variability in results.

However, since it is the observer's opinion of picture quality that counts, any objective measurement system must have good correlation with subjective results for the same video system and test scenes. To be able to incorporate HVS model into broadcasting encoding system could result in additional improve of the coding

A. Mauthe et al. (Eds.): FMN 2009, LNCS 5630, pp. 248–253, 2009.

efficiency and enhance video quality. Typically, to measure the video quality we can find, in the literature, metrics such as the Peak signal-to-noise ratio (PSNR) and the Mean Square Error (MSE), Sum of Squared Differences (SSD), Mean Absolute Difference (MAD), and Sum of Absolute Differences (SAD). These metrics can be determined by the following expressions:

$$PSNR = 10\log_{10}\frac{A^2}{MSE}, \tag{1}$$

$$MSE = \frac{1}{HW}SSD\,,\, SSD = \sum_{i=0}^{H-1}\sum_{j=0}^{W-1}\left(p(i,j) - \hat{p}(i,j)\right)^2$$

$$MAD = \frac{1}{HW}SAD\,,\ SAD = \sum_{i=0}^{H-1}\sum_{j=0}^{W-1}\left|p(i,j) - \hat{p}(i,j)\right| \tag{2}$$

where H and W denote the image dimension (height and width of the image), A represents the maximum grey level of the image (A = 255 for 8-bit representation), and $p(i,j)$ and $\hat{p}(i,j)$ represent the “original” and the processed image pixels at position (i, j). These metrics have been the target of an high number of critics for not correlating well with HVS [8][13] as they can not signify the exact perceptual quality as they are based on pixel to pixel difference calculation and ignore human perception and the viewing condition. Ongoing work regarding the impact of coding distortion on the subjective quality is still under investigation [9], [10].

### 2.1 Structural Similarity Index (SSIM)

A new scheme for a class of quality metrics, known as Structural Similarity (SSIM), has been proposed to model perception implicitly by taking into account the fact that the Human Visual System (HVS) is adapted for extracting structural information (relative spatial covariance) from images [8]. SSIM is an objective image quality assessment metric which attributes perceptual degradations to structural distortions [11]. The SSIM index has been demonstrated in [7] to be an effective measurement of perceptual global degradations in natural images. As describe in [7] and [11] SSIM index can be implemented using a set of equations defining SSIM quality metric in image space depending on luminance l, contrast c, and structure s between the reference and the distorted image.

Let us consider the reference image as x and the distorted image as y, and then each of these parameters can be determined by the expressions:

$$SSIM(x,y) = l(x,y)^{\alpha} \cdot c(x,y)^{\beta} \cdot s(x,y)^{\gamma} \tag{3}$$

where $\alpha, \beta, \gamma$ are positive constants used to weight each comparison function. The comparison functions are given by:

$$l(x,y)=\left(\frac{2\mu_x\mu_y+C_1}{\mu_x^2+\mu_y^2+C_1}\right), c(x,y)=\left(\frac{2\sigma_x\sigma_y+C_2}{\sigma_x^2+\sigma_y^2+C_2}\right), s(x,y)=\left(\frac{\sigma_{xy}+C_3}{\sigma_x\sigma_y+C_3}\right) \quad (4)$$

where $C_1, C_2$ and $C_3$ are constants to avoid instability and $\mu_x, \mu_y, \sigma_x, \sigma_y$ are computed as the mean and standard deviation of the reference and distorted image [11].

$$SSIM(x,y)=\frac{(2\mu_x\mu_y+C_1)(2\sigma_{xy}+C_2)}{(\mu_x^2+\mu_y^2+C_1)(\sigma_x^2+\sigma_y^2+C_2)} \quad (5)$$

In our simulation, we have followed the reference [8] and have selected the following parameters $\alpha=\beta=\gamma=1$ and $C_3=C_2/2$ to get the SSIM value.

## 3 Rate Distortion Models

In a TV broadcast system, the main goal is to optimize the video quality under the rate constraint. To achieve this goal, rate-distortion (R-D) theories have been developed to model the relationship between the coding bit rate and signal distortion. The use of R-D models allows to estimate the minimum number of bits needed in compression a source video file at a given distortion level. The complexity in mathematical modeling of both the video source characteristics and the compression mechanism generate a major different between the information-theory R-D analysis and practices in rate control and quality optimization for video compression.

In H.264 video standard, the quantization parameter of the quantizer controls both the coding bit rate, denoted by R, and the picture quality. The reconstruction error introduced during the encoding process, is regularly referred as distortion, and is denoted as D. The R-D behavior is characterized by its rate-quantizion (R-Q) and distortion-quantization (D-Q) function, denoted by R(q) and D(q), respectively [3]. Together, they are named the R-D functions or curves.

The extensive adoption of the new H.264/AVC video codec standard makes it necessary to study the video encoder's statistical characteristics and compression performance. To best of our knowledge the number of publications on this topic is still very limited [3], [6]. Also a thorough study regarding the application of these methods regarding H.264, including objective and subjective video quality assessment, appears to be missing. Work in this area applied to new codecs has recently started. J. Yang et al. propose in [5] an approach where the mean absolute difference (MAD) of the residual components is used as the complexity measure to adapt to the characteristics of H.264 video coding.

To achieve the target bit rate, the rate control scheme needs to choose the correct quantization parameter. For accuracy, it is of importance to exactly model or estimate the coding bit rate in terms of the quantization parameter, namely rate-quantization (*R-Q*) functions. Together with distortion-quantization (*D-Q*) functions, *R-Q* functions characterize the rate-distortion (*R-D*) behavior of video encoding, which is the key issue of optimum bit allocation. Many *R-Q* and *D-Q* functions have been reported in previous studies [13], in the literature.

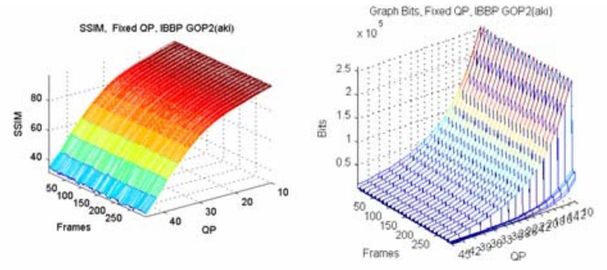


**Fig. 1.** SSIM and bit rate surfaces at various values of QP and frames for the "akiyo_cif" sequence

**Table 1.** GOP Patterns

| GOP Pattern | Intra Period | Nº of B Frames | Pattern |
|---|---|---|---|
| IBBP_GOP1 | 10 | 2 | IBBPBBPBBPBBPBB PBBPBBPBBPBBPBB |
| IBBP_GOP2 | 4 | 2 | IBBPBBPBBPBB |
| IPPP_GOP1 | 4 | 0 | IPPP |
| IPPP_GOP2 | 10 | 0 | IPPPPPPPPP |

In order to examine the behavior of the coding bit rate R regarding SSIM, we have encoded twelve CIF video sequences, 25 frames per second, with the duration of 10 seconds each (Akiyo, Coastguard, Deadline, Flower Garden, Football, Foreman, Hall, Mobile & Calendar, Mother & Daughter, News, Paris and Silence). We have used the JM reference software (version 10.2) of the H.264/AVC codec [12]. Different temporal prediction structures were simulated with Intra Period varying between of 4 and 10, with two B frames.

The sequences were encoded 40 times for fixed quantization parameter QP ranging from 5 to 45 (Figure 1). Figure 1 shows that the actual rate R is a decreasing function of the distortion D (a larger QP results in a lower value of picture quality that corresponds to a larger distortion D). Furthermore, the actual R is also an increasing function of SSIM under a fixed quantization parameter, as show in Fig.1. We can summarize the observations regarding bit rate R as the following: 1) the bit-rate R is a decreasing function of the distortion D; and 2) SSIM is an increasing function of R under a fixed quantization parameter.

To further reveal the relation between R, QP and SSIM we have analyzed results applying curve fitting technique. Before fitting data into a function that models the relationship between two measured quantities, it is a normal procedure to determine if a relationship exists between these quantities. Correlation is a method to confirm the degree of probability that a linear relationship exists between two measured quantities.

$$r = \frac{\sum\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)}{\sqrt{\sum\left(x_i - \bar{x}\right)^2\left(y_i - \bar{y}\right)^2}} \ . \qquad (6)$$

where r is a matrix of correlation coefficients. The sample correlation always lies in the interval from -1 to 1.

**Table 2.** Correlation coefficients between Bits Rate and Quality Metric (SSIM)

| | IBBP-GOP1 | | | IBBP-GOP2 | | |
|---|---|---|---|---|---|---|
| Sequence | I Type | P Type | B Type | I Type | P Type | B Type |
| aki | 0,9017 | 0,8895 | 0,9618 | 0,8975 | 0,8854 | 0,9612 |
| fot | 0,9738 | 0,9688 | 0,9807 | 0,9687 | 0,9657 | 0,9799 |
| new | 0,9164 | 0,8621 | 0,9111 | 0,8622 | 0,8668 | 0,9131 |
| par | 0,9178 | 0,9143 | 0,9426 | 0,9178 | 0,9137 | 0,9439 |
| sil | 0,9335 | 0,9287 | 0,9466 | 0,9288 | 0,9283 | 0,9457 |

A value of r near of positive one or negative one, it is interpreted as indicating a relatively strong linear relationship and r near zero is inferred as indicating a lack of linear relationship. The sign of r indicates whether y tends to increase or decrease with increase x. Results shows a strong correlation between SSIM e Bit Rate.

**Table 3.** Average Square Error, per frame, in 12 video sequences for Rate-SSIM curve fitting

| Fit Method | IBBP GOP2 | | |
|---|---|---|---|
| | I Type | P Type | B Type |
| Linear fit | 3,20E+08 | 2,85E+08 | 3,64E+08 |
| Logarithmic fit | 5,82E+08 | 4,65E+08 | 8,89E+08 |
| Power Regression | 3,17E+08 | 2,91E+08 | 3,41E+08 |
| LNP fit | 1,39E+09 | 9,18E+08 | 2,55E+09 |

Table 3 shows results for curve fitting using Least Squares Methods of Linear Regression Analysis and different fit methods. Simulations in Table 3 are from the twelve video sequences. As mention previously, quantization parameter QP was fixed. Results are consistent and linear relation presents the best results. More experimental results on the other GOP patterns also show similar results. As a consequence, we can draw a conclusion that the relation between R and SSIM can be taken as linear in H.264/AVC.

## 4 Concluding Remarks

The analysis and estimation of the R-D functions have relevant use in H.264 video encoding. First, with the estimated R-D functions we can regulate the quantization settings of the encoder and control the generate bit rate or picture quality according to channel bandwidth, user requirements or storage capacity. Second, based on the estimated R-D functions, optimum bit allocation, as well as other R-D optimization procedures, can be performed to improve the efficiency of the coding algorithm and, consequently, to improve the image quality. The result of the application of these techniques depends on the specific application of the video application. Best scenario is offline video compression where a two passes strategy can result in an increase of video quality.

## Acknowledgements

This work has been supported by “Fundação para a Ciência e Tecnologia” and “Programa Operacional Ciência e Inovação 2010” (POCI 2010), co-funded by the Portuguese Government and European Union by FEDER Program.

## References

[1] ITU-T and ISO/IEC JTC 1, Advanced Video Coding for Generic Audiovisual Services, ITU-T Rec. H.264 & ISO/IEC 14496-10, Version 1 (May 2003); Version 2 (January 2004); Version 3 (with High family of profiles) (September 2004); Version 4 (july 2005)

[2] Wiegand, T., Sullivan, G.J.: The H.264/AVC Video Coding Standard {Standards in a nutshell}. IEEE Signal Processing Magazine 24(2) (March 2007)

[3] He, Z., Mitra, S.K.: A linear source model and a unified rate control algorithm for DCT coding. IEEE Transaction on Circuits ans System on Video Technology 12, 970–982 (2002)

[4] Yang, J., Fang, X., Xiong, H.: A Joint Rate Control Scheme for H.264 Encoding of Multiple Video Sequences. IEEE Transactions on Consumer Electronics 51(618, 2) (May 2005)

[5] Teixeira, L., Corte-Real, L.: Statistical Multiplexing of H.264 programms. In: International Conference on Intelligent and Advanced Systems (ICIAS 2007), Kuala Lumpur, Malaysia, November 25-28 (2007)

[6] Sheikh, H.R., Bovik, A.C.: Image Information and Visual Quality. IEEE Transactions on Image Processing 15(2), 430–444 (2006)

[7] Zhou, W., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image Quality Assessment: From Error Measurement to Structural Similarity. IEEE Transactions on Image Processing 13(4), 600–613 (2004)

[8] Nemethova, O., Ries, M., Siffel, E., Rupp, M.: Subjective Evaluation of Video Quality for H.264 Encoded Sequences. In: Symposia TIC, Bratislava, Slovakia, October 24-26 (2004)

[9] Wolff, T., Ho, H.-H., Foley, J.M., Mitra, S.K.: H.264 coding artifacts and their relation to perceived annoyance. In: European Signal Processing Conference (2006)

[10] Brooks, A.C., Pappas, T.N.: Structural similarity quality metrics in a coding context: exploring the space of realistic distortions. In: IEEE International Conference on Acoustics, Speech, and Signal Processing, Honolulu, Hawaii, April 15-20, pp. 869–872 (2007)

[11] Joint Video Team (JVT), H.264/Advanced Video Coding reference software version 10.2 (2006), `http://iphome.hhi.de/suehring/tml/`

[12] Yin, P., Boyce, J.: A new rate control scheme for H.264 video coding. In: Proc. IEEE Intl. Conf. Image Processing, October 2004, pp. 449–452 (2004)