A New Rate-Distortion Optimization Using Structural Information in H.264 I-Frame Encoder

Zhi-Yi Mai[1], Chun-Ling Yang[1], Lai-Man Po[2], and Sheng-Li Xie[1]

[1] School of Electronic and Information Engineering, South China University of Technology,
Guangzhou, Guangdong, 510640, China
[2] Department of Electronic Engineering, City University of Hong Kong,
83 Tat Chee Avenue, Kowloon Tong, Hong Kong, China
kathymaizy@yahoo.com.cn, eeclyang@scut.edu.cn,
eelmpo@cityu.edu.hk

Abstract. Rate-distortion optimization is the key technique in video coding standards to efficiently determine a set of coding parameters. In the R-D optimization for H.264 I-frame encoder, the distortion (D) is measured as the sum of the squared differences (SSD) between the reconstructed and the original blocks, which is same as MSE. Recently, a new image measurement called Structural Similarity (SSIM) based on the degradation of structural information was brought forward. It is proved that the SSIM can provide a better approximation to the perceived image distortion than the currently used PSNR (or MSE). In this paper, a new rate-distortion optimization for H.264 I-frame encoder using SSIM as the distortion metric is proposed. Experiment results show that the proposed algorithm can reduced 2.2~6.45% bit rate while maintaining the perceptual quality.

1 Introduction

As the rapid development of digital techniques and increasing use of Internet, image and video compression plays a more and more important role in our life. The newest international video coding standard H.264 adopts many advanced techniques, such as directional spatial prediction in I-frame encoder, variable and Hierarchical block transform, arithmetic entropy coding, multiple reference frame motion compensation, deblocking etc. All these novel and advanced techniques make it provide approximately a 50% bit rate savings for equivalent perceptual quality relative to the performance of prior standards [1]. Except for the new techniques, the operational control of the source encoder is still a key problem in H.264, and it is still optimized with respect to the rate-distortion efficiency using Lagrangian optimization techniques, just like the prior standards, MPEG-2, H.263 and MPEG-4. In the R-D optimization function for H.264 intra prediction, distortion is measured as SSD between the reconstructed and the original blocks, which has the same meaning with MSE. Although Peak Signal-to-Noise Ratio (PSNR) and MSE are currently the most widely used objective metrics due to their low complexity and clear physical meaning, they were also widely criticized for not correlating well with Human Visual System (HVS) for a long time [2]. During past several decades a great deal of effort has been made to

J. Blanc-Talon et al. (Eds.): ACIVS 2005, LNCS 3708, pp. 435 – 441, 2005.

develop new image quality assessment based on error sensitivity theory of HVS, but only limit success has been achieved by the reason that the HVS has not been well comprehended.

Recently a new philosophy for image quality measurement was proposed, based on the assumption that the human visual system is highly adapted to extract structural information from the viewing field. It follows that a measure of structural information change can provide a good approximation to perceived image distortion [3]. In this new theory, an item called Structural Similarity (SSIM) index including three comparisons is introduced to measure the structural information change. Experiments showed that the SSIM index method is easy to implement and can better corresponds with human perceived measurement than PSNR (or MSE). Thus, in this paper we propose to employ SSIM in the rate-distortion optimizations of H.264 I-frame encoder to choose the best prediction mode(s).

The remainder of this paper is organized as follows. In section II, the I-frame coding of H.264 and the idea of SSIM is summarized. The detail of our proposed method is given in section III. Section IV presents the experimental results to demonstrate the advantage of the SSIM index method. Finally, section V draws the conclusion.

2 H.264 I-Frame Encoder and SSIM

2.1 H.264 I-Frame Encoder

In H.264 I-frame encoder, each picture is partitioned into fixed-size macroblocks (MB) that cover a rectangular area of 16×16 samples of the luma component and 8×8 samples of each chroma component. Then each macroblock is spatially predicted using its neighbouring samples of previously coded blocks which are to the left and/or above the block, and the prediction residual is integer transformed, quantized and transmitted using entropy coding. The latest JVT reference software version (JM92) of H.264 [4] provides three types of intra prediction denoted as intra_16x16, intra_8x8 and intra_4x4. The intra_16x16 which supports 4 prediction modes performs prediction of the whole macroblock and is suited for smooth area, while the intra_8x8 and intra_4x4 which performs 8×8 and 4×4 block respectively support 9 prediction modes and are suited for detailed part of the picture. The best prediction mode(s) are chosen utilizing the R-D optimization[5] which is described as:

$$J(\mathbf{s},\mathbf{c},MODE\,|\,QP) = D(\mathbf{s},\mathbf{c},MODE\,|\,QP) + \lambda_{MODE} R(\mathbf{s},\mathbf{c},MODE\,|\,QP) \ . \tag{1}$$

In the above formula, the distortion D(***s,c***,MODE|QP) is measured as SSD between the original block ***s*** and the reconstructed block ***c***, and QP is the quantization parameter, MODE is the prediction mode. R(***s,c***,MODE|QP) is the bit number coding the block. The modes(s) with the minimum J(***s,c***,MODE|QP) are chosen as the prediction mode(s) of the macroblock.

2.2 Structural Similarity (SSIM)

The new idea of SSIM index is to introduce the measure of structural information degradation, which includes three comparisons: luminance, contrast and structure [3]. It's defined as

$$SSIM(\mathbf{x},\mathbf{y}) = l(\mathbf{x},y)\cdot c(\mathbf{x},\mathbf{y})\cdot s(\mathbf{x},\mathbf{y}) \ . \tag{2}$$

where $\mathrm{l}(\boldsymbol{x},\boldsymbol{y})$ is Luma comparison, $\mathrm{c}(\boldsymbol{x},\boldsymbol{y})$ is Contrast comparison and $\mathrm{s}(\boldsymbol{x},\boldsymbol{y})$ is Structure comparison. They are defined as:

$$\mathrm{l}(\boldsymbol{x},\boldsymbol{y}) = \frac{2\mu_{\mathrm{x}}\mu_{\mathrm{y}} + \mathrm{C}_1}{\mu_{\mathrm{x}}^2 + \mu_{\mathrm{y}}^2 + \mathrm{C}_1} \ . \tag{3}$$

$$c(\mathbf{x},\mathbf{y}) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \ . \tag{4}$$

$$\mathrm{s}(\boldsymbol{x},\boldsymbol{y}) = \frac{(\sigma_{\mathrm{xy}} + \mathrm{C}_3)}{\sigma_{\mathrm{x}}\sigma_{\mathrm{y}} + \mathrm{C}_3} \ . \tag{5}$$

where $\boldsymbol{x}$ and $\boldsymbol{y}$ are two nonnegative image signals to be compared, μ_{x} and μ_{y} are the mean intensity of image $\boldsymbol{x}$ and $\boldsymbol{y}$ respectively, σ_{x} and σ_{y} are the standard deviation of image $\boldsymbol{x}$ and $\boldsymbol{y}$ respectively, σ_{xy} is the covariance of image x and y. In fact, without C_3, the equation (5) is the correlation coefficient of image $\boldsymbol{x}$ and $\boldsymbol{y}$, and C_1, C_2 and C_3 are small constants to avoid the denominator being zero. It's recommended by [3]:

$$\mathrm{C}_1 = (\mathrm{K}_1\mathrm{L})^2,\ \mathrm{C}_2 = (\mathrm{K}_2\mathrm{L})^2,\ \mathrm{C}_3 = {}^{\mathrm{C}_2}\!/_{2} \ . \tag{6}$$

where $\mathrm{K}_1,\mathrm{K}_2 << 1$ and L is the dynamic range of the pixel values (255 for 8-bit grayscale images). In addition, the higher the value of $SSIM(\mathbf{x},\mathbf{y})$ is, the more similar the image $\boldsymbol{x}$ and $\boldsymbol{y}$ are.

3 The R-D Optimization Using Structural Similarity in H.264

As the SSIM index method performs better as image quality measurement than MSE (SSD), we propose to replace the SSD with the SSIM index in the R-D optimization of H.264 I-frame encoder. The quality of the reconstructed picture is higher when its SSIM index is greater while the SSD performs the other way. Therefore the distortion in our method is measured as:

$$\mathrm{D}(\boldsymbol{s},\boldsymbol{c},\mathrm{MODE}|\mathrm{QP}) = 1 - \mathrm{SSIM}(\boldsymbol{s},\boldsymbol{c}) \ . \tag{7}$$

where $\boldsymbol{s}$ and $\boldsymbol{c}$ are the original and reconstructed image block respectively.

Due to the change of distortion measure, the Lagrangian multiplier should be modified correspondingly. According to the relation between SSIM($\boldsymbol{s}$,$\boldsymbol{c}$) and R($\boldsymbol{s}$,$\boldsymbol{c}$,MODE|QP) and motivated by the theory in [6] and [7], the new Lagrangian multiplier in our algorithm is

$$\lambda_{MODE} = 1.11 * 2^{(QP-60)/5} \ . \tag{8}$$

where QP denotes the quantization parameter. Consequently, the new R-D cost function can be written as:

$$J(\mathbf{s}, \mathbf{c}, MODE \mid QP) = 1 - SSIM(\mathbf{s}, \mathbf{c}) + \lambda_{MODE} R(\mathbf{s}, \mathbf{c}, MODE \mid QP) \ . \tag{9}$$

Our new algorithm is using SSIM index instead of SSD as the distortion measure in RDCost_for_4x4IntraBlock, RDCost_for_8x8IntraBlock and RDCost_for_macro-blocks, but the decisions of finding the best mode for Intra_16x16 which uses Hadamard transform remain unchanged. The SSIM indexes of all types of prediction blocks are computed within 4×4 nonoverlapping square windows, while slide window, which is of 16×16, is used to compute the whole reconstructed image quality MSSIM (mean SSIM). Furthermore, the parameter setting here is chosen as follows: K_1=0.01, K_2=0.03, L=255.

4 Experimental Results

Experiments are carried out using several 8 bit/pixel grayscale images of various sizes. They are Apple, Claire, MissA and Salesman of 176×144, Bridge and Camera of 256×256, Airplane, Baboon, Lena and Sailboat of 512×512, Pentagon and Man of 1024×1024. All the modifications are based on the JVT reference software JM92 program [4]. Results in terms of total bits of the compressed image, MSSIM of the whole reconstructed image and the comparison between the two methods are listed in Table 1~3 under the Quantization Parameter (QP) equal to 10, 20 and 30 respectively.

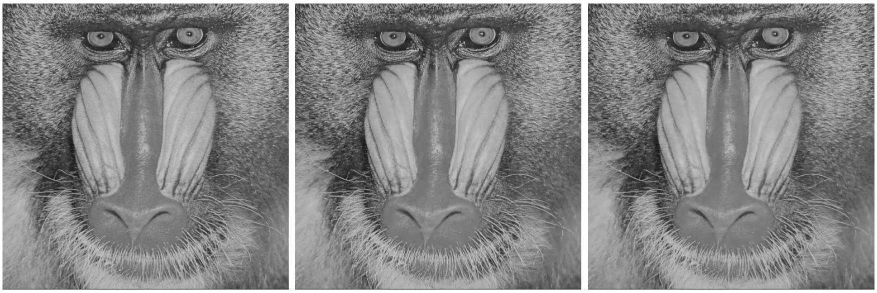
Table 1. Simulation results with QP=10

Image	H.264-JM92		Our method		Comparison (%)	
	Bits	MSSIM	Bits	MSSIM	Bit decrement	MSSIM decrement
Apple	53664	0.9980	50200	0.9973	6.45	0.07
Claire	39056	0.9976	37480	0.9973	4.04	0.03
MissA	42072	0.9965	40160	0.9959	4.54	0.06
Salesman	94760	0.9994	91800	0.9991	3.12	0.03
Bridge	335464	0.9997	327456	0.9995	2.39	0.02
Camera	227768	0.9976	218104	0.9968	4.24	0.08
Airplane	722888	0.9973	687392	0.9963	4.91	0.10
Baboon	1331024	0.9993	1294408	0.9990	2.75	0.03
Lena	874480	0.9982	835024	0.9973	4.51	0.09
Sailboat	1042040	0.9984	1003040	0.9978	3.74	0.06
Man	4068144	0.9986	3911080	0.9980	3.86	0.06
Pentagon	4589568	0.9991	4437472	0.9987	3.31	0.04

Results in Table 1 to 3 show that the proposed algorithm can achieve about 2.2~6.45% bits saving while maintaining almost the same MSSIM index. In order to illustrate the perceptual quality of the reconstructed image, this paper shows the original and reconstructed images with the largest MSSIM decreased in Figure 1, from which it's clear that the visual difference between the two reconstructed images using H.264 JM92 (Fig.1 b) and our proposed algorithm (Fig.1 c) can hardly be found. That means the new R-D optimization algorithm can achieve about 2.2~6.45% bit saving while maintaining almost the same perceptual quality.

Table 2. Simulation results with QP=20

Image	H.264-JM92		Our method		Comparison (%)	
	Bits	MSSIM	Bits	MSSIM	Bit decrement	MSSIM decrement
Apple	16728	0.9889	15984	0.9879	4.45	0.10
Claire	17800	0.9941	17088	0.9934	4.00	0.07
MissA	16088	0.9898	15296	0.9885	4.92	0.13
Salesman	51984	0.9951	50192	0.9938	3.45	0.13
Bridge	209880	0.9968	203096	0.9958	3.23	0.10
Camera	108824	0.9818	104976	0.9802	3.54	0.16
Airplane	293744	0.9833	280152	0.9815	4.63	0.18
Baboon	821424	0.9928	789032	0.9907	3.94	0.21
Lena	366624	0.9813	349608	0.9790	4.64	0.23
Sailboat	548400	0.9858	524272	0.9834	4.40	0.24
Man	2039408	0.9859	1938360	0.9829	4.95	0.30
Pentagon	2595528	0.9906	2477960	0.9878	4.53	0.28



(a) Baboon (original)

(b) Encoded by H.264 I-frame encoder with QP=30

(c) Encoded by our method with QP=30

Fig. 1. The reconstructed image by the two methods

Table 3. Simulation results with QP=30

Image	H.264-JM92		Our method		Comparison (%)	
	Bits	MSSIM	Bits	MSSIM	Bit decrement	MSSIM decrement
Apple	5808	0.9762	5680	0.9731	2.20	0.32
Claire	8056	0.9829	7824	0.9812	2.88	0.17
MissA	6176	0.9718	5848	0.9681	5.31	0.38
Salesman	21416	0.9647	20528	0.9587	4.15	0.62
Bridge	97352	0.9714	92240	0.9647	5.25	0.69
Camera	48240	0.9561	46864	0.9512	2.85	0.51
Airplane	102904	0.9599	97920	0.9555	4.84	0.46
Baboon	361696	0.9457	343904	0.9368	4.92	0.94
Lena	102568	0.9468	98648	0.9420	3.82	0.51
Sailboat	173304	0.9362	163920	0.9306	5.41	0.60
Man	587952	0.9326	560816	0.9251	4.62	0.80
Pentagon	812728	0.9243	775392	0.9164	4.59	0.85

5 Conclusion

In this paper, we propose a new R-D optimization using the structural similarity (SSIM) instead of SSD for quality assessment in H.264 I-frame encoder. Experiments show that it can reduce approximately 2.2~6.45% bit rate while maintaining the same perceptual quality. The improvement of coding efficiency is not very large, but the new idea and the beginning results are inspiring. Thus, even better results maybe obtained by deeply studying. Furthermore, the proposed R-D optimization can be transplanted easily into motion estimation of inter frame coding.

Acknowledgement

The work described in this paper was substantially supported by research projects from National Natural Science Foundation of China. [Project No. 60402015, No.60325310]

References

1. Wiegand T., Sullivan G. J., Bjontegaard G., and Luthra A., “Overview of the H.264/AVC Video coding Standard,” IEEE Trans. on CAS for video Technology, no.7, Vol. 13, pp.560-576, July 2003.
2. Wang Z., Bovik A. C., and Lu L., “Why is image quality assessment so difficult,” in Proc. IEEE Int. Conf. Acoustics, speech, and Signal Processing, vol. 4, Orlando, FL, May 2002, pp.313–3316.
3. Wang Z., Bovik A. C., Sheikh H. R., and Simoncelli E. P., “Image quality assessment: from error visibility to structural similarity,” IEEE Trans. Image Processing, vol. 13, no.4, pp. 600–612, Apr. 2004.

4. http://bs.hhi.de/~suehring/tml/download
5. Ma S.W., Gao W., Gao P., and Lu Y., “Rate control for advance video coding (AVC) standard,” in Proc. ISCAS’03, vol.2, pp.II-892-II-895, May 2003.
6. Wiegand T. and Girod B., “Lagrangian multiplier selection in hybrid video coder control,” in Proc. ICIP 2001, Thessaloniki, Greece, Oct. 2001.
7. Sullivan G. J. and Wiegand T., “Rate-Distortion Optimization for Video Compression”, IEEE Signal Processing Magazine, vol. 15, no. 6, pp. 74-90, Nov. 1999