

# Improved Best Prediction Mode(s) Selection Methods Based on Structural Similarity In H.264 I-Frame Encoder

**Zhi-Yi Mai**

School of electronic  
and information engineering,  
South China University of  
Technology  
Guangzhou, China  
kathymaizy@yahoo.com.cn

**Chun-Ling Yang**

School of electronic  
and information engineering,  
South China University of  
Technology  
Guangzhou, China  
eeclyang@scut.edu.cn

**Sheng-Li Xie**

School of electronic  
and information engineering,  
South China University of  
Technology  
Guangzhou, China  
adshlxie@scut.edu.cn

**Abstract** - In H.264 I-frame encoder, the best intra prediction modes are chosen by utilizing the rate-distortion (R-D) optimization whose distortion is the sum of the squared differences (SSD, means the same as MSE) between the reconstructed and the original blocks. 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, we propose two improved prediction modes selection methods based on SSIM for H.264 I-frame encoder. The first one is the SSIM-based R-D optimization (SBRDO) method, the other is the fast mode selection method based on SSIM (FMSBS). Experiments show that both the proposed method can improve the coding efficiency while maintaining the same perceptual reconstructed image quality.

**Keywords:** Structural Similarity (SSIM), intra prediction, rate-distortion optimization, SSIM-based R-D optimization (SBRDO), fast mode selection based on SSIM (FMSBS).

## 1 Introduction

With 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 latest 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 by using Lagrangian optimization techniques with respect to their rate-distortion efficiency, just as 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 the sum of the squared differences (SSD), which has the same meaning with MSE, between the reconstructed and the original blocks [2]. Although Peak Signal-to-Noise Ratio (PSNR) and MSE [3] are currently the most widely used objective metrics due to their low complexity and clear physical meaning, they have been also widely criticized for not correlating well with Human Visual System (HVS) for a long time [4]. In the past several decades, a great deal of effort has been made to develop new image quality assessment based on error sensitivity theory of HVS, but only limited success has been achieved by the reason that the HVS is rather complex and has not been well comprehended. Thus SSD is still employed as the distortion metric in H.264.

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 says that a measure of structural information change can provide a good approximation to perceived image distortion [4,5]. In that philosophy, an item called Structural Similarity (SSIM) index which includes three comparisons is introduced to measure the structural information change. Experiments showed that the SSIM index method which is easy to be implemented can correspond with human perceived measurement better than PSNR (or MSE). Therefore, in this paper we propose two algorithms that employ the SSIM index in the H.264 I-frame encoder to choose the best prediction mode(s).

The remainder of this paper is organized as follows. In section 2, the I-frame coding of H.264 and the idea of SSIM index is summarized. The detail of our proposed methods is given in section 3. Section 4 presents the experimental results to demonstrate the advantage of the SSIM index method. Finally, section 5 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 integrally transformed, quantized and transmitted using entropy coding. The latest JVT reference software version (JM92) of H.264 [6] provides three types of intra prediction denoted as intra\_16x16, intra\_8x8 and intra\_4x4. The intra\_16x16 which supports four prediction modes performs prediction of the whole macroblock and is suited for smooth area, while the intra\_8x8 and intra\_4x4 which performs prediction on 8×8 or 4×4 block support nine prediction modes respectively and are suited for detailed parts of the picture. The best prediction modes are chosen by utilizing the R-D optimization [2] which is described as:

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

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

### 2.2 Structural Similarity (SSIM)

The new idea of SSIM index is to introduce the measure of structural information degradation, which include three comparisons: luminance, contrast and structure [5]. It's defined as

$$SSIM(x, y) = l(x, y) \cdot c(x, y) \cdot s(x, y) \quad (2)$$

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

$$l(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \quad (3)$$

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

$$s(x, y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \quad (5)$$

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

$$C_1 = (K_1L)^2, C_2 = (K_2L)^2, C_3 = C_2/2 \quad (6)$$

where  $K_1, K_2 \ll 1$  and  $L$  is the dynamic range of the pixel values (255 for 8-bit grayscale images). In addition, the higher the value of  $SSIM(x, y)$  is, the more similar the image  $x$  and  $y$  are.

## 3 Improved best prediction mode(s) selection methods based on SSIM in H.264 I-frame encoder

### 3.1 The SSIM-based R-D optimization (SBRDO) in H.264 I-Frame encoder

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

$$D(s, c, MODE | QP) = 1 - SSIM(s, c) \quad (7)$$

where  $s$  and  $c$  are the original and the reconstructed image block respectively.

Due to the change of distortion measure, the Lagrangian multiplier should be modified correspondingly. In conformity to the relation between  $SSIM(s, c)$  and  $R(s, c, MODE | QP)$  and motivated by the theory in [7] and [8], the new Lagrangian multiplier in our algorithm becomes:

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

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

$$J(s, c, \text{MODE} | \text{QP}) = 1 - \text{SSIM}(s, c) + \lambda_{\text{MODE}} R(s, c, \text{MODE} | \text{QP}) \quad (9)$$

In this method, we use the SSIM index instead of SSD as the distortion measure in  $\text{RDCost\_for\_4x4IntraBlock}$ ,  $\text{RDCost\_for\_8x8IntraBlock}$  and  $\text{RDCost\_for\_macroblocks}$ , but the decisions of finding the best mode for Intra\_16x16 which using Hadamard transform remain unchanged.

Although the SBRDO can decrease the bit number for H.264 I-frame encoding, its computation cost becomes higher than the SSD-based R-D optimization when QP is large, e.g.  $\text{QP} \geq 30$ . Thus, a fast prediction mode selection method based on SSIM is proposed according to the properties of SSIM in the following subsection.

### 3.2 A fast mode selection method based on SSIM (FMSBS) for intra prediction

H.264 provides multiple spatial prediction modes in order to achieve high compression efficiency. The best modes in H.264 intra prediction are determined after all the prediction residual images are transformed, quantized, and entropy coded. As intra\_16x16 supports four prediction modes, intra\_4x4 and intra\_8x8 support nine prediction modes respectively, the best mode selection for each macroblock demands rather heavy computation cost.

On the other hand, as the SSIM expresses the structural similarity of two images, the prediction block having larger SSIM implicates that it's more similar to the original one, and then produce lower frequency residual image which can be easily encoded. According to the above analysis, we propose a fast mode selection method based on SSIM (FMSBS) for intra prediction. The major steps for each macroblock selecting the best prediction mode are summarized as follow:

Step 1: Find the best intra\_16x16 prediction mode.

- Generate the four prediction blocks respectively according to the four intra\_16x16 prediction modes.
- Perform Hadamard transform for the residual blocks and then sum up the absolute values of all the Hadamard transform coefficients as the cost.
- The mode that has the smallest cost is chosen as the best intra\_16x16 prediction mode.

Step 2: Find the best intra\_4x4 prediction mode

Divide the macroblock into sixteen 4x4 non-overlapped blocks. For each 4x4 block:

- Generate nine prediction blocks based on the nine intra\_4x4 prediction modes.
- Compute the SSIM between the 4x4 prediction block and the original one.
- The mode that has the largest SSIM is chosen as the best mode.

Step 3: Find the best intra\_8x8 prediction mode

Divide the macroblock into four 8x8 non-overlapped blocks. For each 8x8 block:

- Generate nine prediction blocks based on the nine intra\_8x8 prediction modes
- Compute the SSIM between the 8x8 prediction block and the original one.
- The mode that has the largest SSIM is chosen as the best mode.

Step 4: Find the best prediction mode for the macroblock

- Compute the rate-distortion cost using function (8) and (9) for the best intra\_16x16 mode, the best intra\_4x4 modes and the best intra\_8x8 modes respectively.
- The mode with the minimum cost will be chosen as the best prediction mode of the macroblock.

In a word, only the residual blocks generated by the best intra\_16x16 mode, the best intra\_4x4 modes, and the best intra\_8x8 modes need transforming, quantizing and entropy coding. Hence the FMSBS can greatly reduce the computation cost and save a lot of time compare to the original intra prediction mode selection process in H.264.

## 4 Experiments

### 4.1 Experimental environment

Experiments are carried out using several 8 bit/pixel grayscale images of various sizes. They are Salesman and Coastguard of 176x144, Moon surface and Chemical plant of 256x256, Lena and Baboon of 512x512, San Francisco and Airport of 1024x1024.

All of our experiments are based on the JVT reference software JM92 program [6]. The results are performed on a P4/2.0GHz personal computer with 256MB RAM and Microsoft Windows 2000 as the operation system.

The SSIM index for a 4x4IntraBlock or 8x8IntraBlock is computed directly while the SSIM index for a macroblock is calculated within sixteen 4x4 non-overlapped square windows and then averaged as a MSSIM. Also 16x16 slide window is used to compute the whole reconstructed image quality MSSIM. Furthermore, the following parameter settings is used in the SSIM measure:  $K_1=0.01$ ,  $K_2=0.03$ ,  $L=255$ .

### 4.2 Experiment results

Results of SBRDO in terms of total bits of the compressed image, MSSIM of the whole reconstructed image, coding time and the comparison between our method and H.264 are listed in Table 1, 3, 5 with the Quantization Parameter (QP) equal to 10, 20 and 30 respectively. Table 2, 4, 6 show the results of FMSBS also with the QP equal to 10, 20 and 30 respectively.

Results in Table 1, 3 and 5 show that the SBRDO can achieve about 2.6~5.3% bit savings 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 decrement 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 SBRDO (Fig.1 c) can hardly be found. That means the new R-D optimization algorithm can achieve about 2.6~5.3% bit savings while maintaining almost the same perceptual quality. Results also show that the SBRDO retain the

computation complexity as H.264 for small QP (QP=10), but it costs 3~6.5% more coding time than H.264 when QP is large (QP=30).

Results in Table 2, 4 and 6 show that the fast mode selection method, FMSBS, can achieve about 60% time savings with no more than 0.51% MSSIM decrement and 2.5% output bits increment. Moreover, the output bits even decrease when the QP is small, e.g. QP is 10. That's because the residuals usually become lower frequency signals when implementing the SSIM measure for prediction mode selection.

Table 1. Results of comparison between H.264 and SBRDO with QP=10

Image	H.264_JM92			SBRDO			Comparison (%)		
	Bits	MSSIM	Time (ms)	Bits	MSSIM	Time (ms)	Bit inc.	MSSIM dec.	Time inc.
Salesman	94760	0.9994	931	91928	0.9992	925	-2.99	0.02	-0.70
Coastguard	104120	0.9988	945	99776	0.9984	933	-4.17	0.04	-1.30
Moon surface	281752	0.9988	2480	272904	0.9983	2511	-3.14	0.05	1.24
Chemical plant	296000	0.9994	2575	286040	0.9992	2592	-3.36	0.02	0.68
Lena	874480	0.9982	9155	836048	0.9975	9133	-4.39	0.07	-0.24
Baboon	1331024	0.9993	11570	1295736	0.9991	11537	-2.65	0.02	-0.28
San Francisco	4419912	0.9979	42372	4262800	0.9972	42377	-3.55	0.07	0.01
Airport	4796408	0.9989	44017	4633032	0.9984	43647	-3.41	0.05	-0.84

Table 2. Results of comparison between H.264 and FMSBS with QP=10

Image	H.264_JM92			FMSBS			Comparison (%)		
	Bits	MSSIM	Time (ms)	Bits	MSSIM	Time (ms)	Bit incr.	MSSIM dec.	Time saving
Salesman	94760	0.9994	931	94328	0.9992	278	-0.46	0.02	70.13
Coastguard	104120	0.9988	945	101232	0.9984	283	-2.77	0.04	70.08
Moon surface	281752	0.9988	2480	278760	0.9982	758	-1.06	0.06	69.44
Chemical plant	296000	0.9994	2575	291576	0.9992	775	-1.49	0.02	69.90
Lena	874480	0.9982	9155	850240	0.9974	2875	-2.77	0.08	68.59
Baboon	1331024	0.9993	11570	1321480	0.9990	3539	-0.72	0.03	69.41
San Francisco	4419912	0.9979	42372	4329048	0.9971	13424	-2.06	0.08	68.32
Airport	4796408	0.9989	44017	4699096	0.9983	13781	-2.03	0.06	68.69

Table 3. Results of comparison between H.264 and SBRDO with QP=20

Image	H.264_JM92			SBRDO			Comparison (%)		
	Bits	MSSIM	Time (ms)	Bits	MSSIM	Time (ms)	Bit inc.	MSSIM dec.	Time inc.
Salesman	51984	0.9951	674	50248	0.9942	680	-3.34	0.09	0.94
Coastguard	53984	0.9886	684	52256	0.9877	688	-3.20	0.09	0.47
Moon surface	158560	0.9875	1856	150200	0.9839	1874	-5.27	0.36	0.94
Chemical plant	165072	0.9945	1899	159336	0.9935	1924	-3.47	0.10	1.30
Lena	366624	0.9813	6350	351120	0.9794	6530	-4.23	0.19	2.83
Baboon	821424	0.9928	8719	789872	0.9912	8761	-3.84	0.16	0.49
San Francisco	2411544	0.9787	30948	2324352	0.9764	31474	-3.62	0.24	1.70
Airport	2775680	0.9876	33155	2650048	0.9848	33374	-4.53	0.28	0.66

Table 4. Results of comparison between H.264 and FMSBS with QP=20

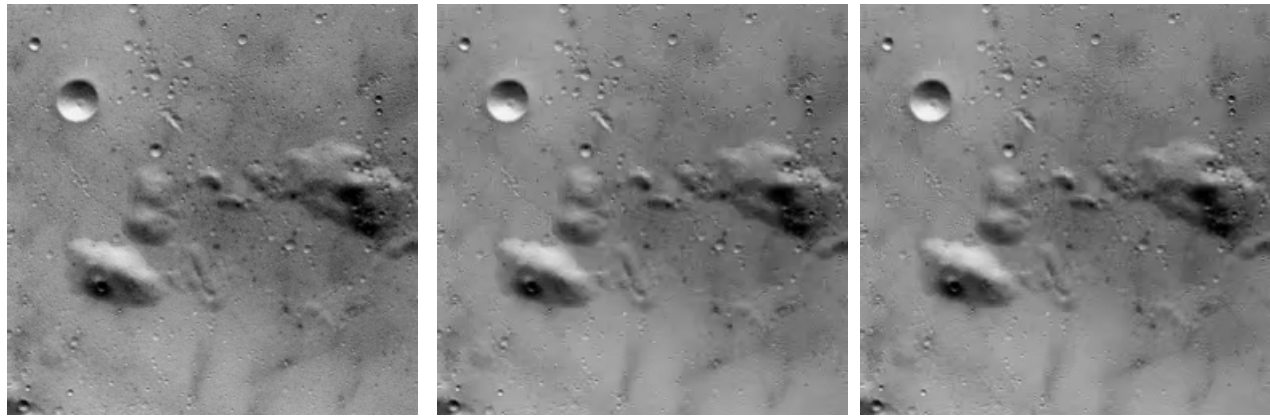
Image	H.264_JM92			FMSBS			Comparison (%)		
	Bits	MSSIM	Time (ms)	Bits	MSSIM	Time (ms)	Bit inc.	MSSIM dec.	Time saving
Salesman	51984	0.9951	674	52096	0.9941	211	0.22	0.10	68.66
Coastguard	53984	0.9886	684	53472	0.9877	216	-0.95	0.09	68.48
Moon surface	158560	0.9875	1856	153488	0.9837	586	-3.20	0.38	68.43
Chemical plant	165072	0.9945	1899	163832	0.9937	595	-0.75	0.08	68.66
Lena	366624	0.9813	6350	362032	0.9799	2191	-1.25	0.14	65.50
Baboon	821424	0.9928	8719	808192	0.9912	2797	-1.61	0.16	67.92
San Francisco	2411544	0.9787	30948	2380888	0.9768	10663	-1.27	0.19	65.55
Airport	2775680	0.9876	33155	2695752	0.9851	11091	-2.88	0.25	66.55

Table 5. Results of comparison between H.264 and SBRDO with QP=30

Image	H.264_JM92			SBRDO			Comparison (%)		
	Bits	MSSIM	Time (ms)	Bits	MSSIM	Time (ms)	Bit inc.	MSSIM dec.	Time inc.
Salesman	21416	0.9647	492	20704	0.9612	511	-3.32	0.36	3.82
Coastguard	19240	0.9325	481	18600	0.9286	503	-3.33	0.42	4.53
Moon surface	42968	0.8807	1206	41184	0.8724	1258	-4.15	0.94	4.28
Chemical plant	68752	0.9648	1353	66328	0.9609	1403	-3.53	0.40	3.70
Lena	102568	0.9468	4636	99200	0.9434	4928	-3.28	0.36	6.30
Baboon	361696	0.9457	6217	346112	0.9403	6416	-4.31	0.57	3.19
San Francisco	959120	0.9263	22180	911592	0.9199	23217	-4.96	0.69	4.68
Airport	927216	0.9041	22289	889952	0.8981	23391	-4.02	0.66	4.94

Table 6. Results of comparison between H.264 and FMSBS with QP=30

Image	H.264_JM92			FMSBS			Comparison (%)		
	Bits	MSSIM	Time (ms)	Bits	MSSIM	Time (ms)	Bit inc.	MSSIM decr.	Time saving
Salesman	21416	0.9647	492	21856	0.9611	170	2.05	0.37	65.40
Coastguard	19240	0.9325	481	19672	0.9281	169	2.25	0.47	64.89
Moon surface	42968	0.8807	1206	42800	0.8762	434	-0.39	0.51	63.99
Chemical plant	68752	0.9648	1353	69576	0.9617	463	1.20	0.32	65.81
Lena	102568	0.9468	4636	104904	0.9443	1800	2.28	0.26	61.17
Baboon	361696	0.9457	6217	358288	0.9410	2201	-0.94	0.50	64.59
San Francisco	959120	0.9263	22180	940296	0.9209	8563	-1.96	0.58	61.39
Airport	927216	0.9041	22289	929856	0.8996	8600	0.28	0.50	61.42



(a) Moon surface (original)

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

(c) Encoded by SBRDO with QP=30

Figure 1. The reconstructed image by H.264 and our first proposed method

## 5 Conclusion

In this paper, we propose two improved algorithms for the H.264 I-frame encoder. One is a new R-D optimization using the structural similarity (SSIM) instead of SSD as the quality assessment (SBRDO). The other is a fast mode selection method for intra prediction by SSIM directly (FMSBS). Experiments show that the SBRDO can reduce approximately 2.6~5.3% bit rate while maintaining the same perceptual quality and costing almost the same time for encoding with small QP, but a little more for large QP, and the FMSBS can achieve about 60% time savings while maintaining almost the same compression rate and equivalent perceptual image quality. The improvement of the coding efficiency is not very large for SBRDO, but the time saving is obvious for the FMSBS. This new idea and the beginning results are inspiring, and better results maybe obtained by further studying. Furthermore, the proposed R-D optimization can be transplanted easily into motion estimation of inter frame coding.

## Acknowledgments

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

## Reference

[1] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "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] S.W. Ma, W. Gao, P. Gao, and Y. Lu, "Rate control for advance video coding (AVC) standard," in *Proc. ISCAS'03*, vol.2, pp.II-892-II-895, May 2003.

[3] VQEG, "Final report from the video quality experts group on the validation of objective models of video quality assessment", Mar.2000. <http://www.vqeg.org/>

[4] Z. Wang, A. C. Bovik, and L. Lu, "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.

[5] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Processing*, vol. 13, no.4, pp. 600-612, Apr. 2004.

[6] <http://bs.hhi.de/~suehring/tml/download>

[7] T. Wiegand and B. Girod, "Lagrangian multiplier selection in hybrid video coder control," in *Proc. ICIP 2001*, Thessaloniki, Greece, Oct. 2001.

[8] G. J. Sullivan and T. Wiegand, "Rate-Distortion Optimization for Video Compression", *IEEE Signal Processing Magazine*, vol. 15, no. 6, pp. 74-90, Nov. 1999