

Two-Step Rate Distortion Optimization Algorithm for High Efficiency Video Coding

Kalyan Goswami^{1,*}, Dae Yeol Lee², Jongho Kim², Seyoon Jeong²,
Hui Yong Kim², Byung-Gyu Kim³

Abstract

High Efficiency Video Coding (HEVC) is the newest video coding standard for improvement in video data compression. This new standard provides a significant improvement in picture quality, especially for high-resolution videos. A quadtree-based structure is created for the encoding and decoding processes and the rate-distortion (RD) cost is calculated for all possible dimensions of coding units in the quadtree. To get the best combination of the block an optimization process is performed in the encoder, called rate distortion optimization (RDO). In this work we are proposing a novel approach to enhance the overall RDO process of HEVC encoder. The proposed algorithm is performed in two steps. In the first step, like HEVC, it performs general rate distortion optimization. The second step is an extra checking where a SSIM based cost is evaluated. Moreover, a fast SSIM (FSSIM) calculation technique is also proposed in this paper.

Key Words: High Efficiency Video Coding (HEVC), Rate Distortion Optimization (RDO), Video compression, Structural Similarity Matric (SSIM), Inter prediction, Intra prediction.

I. INTRODUCTION

High Efficiency Video Coding (HEVC) is the latest video coding standard proposed by the Joint Collaborative Team on Video Coding (JCT-VC). Increasing demand for HD and beyond HD video formats has created a need for HEVC and around 50% bit rate reduction is achieved using HEVC over the previous standard (H.264/AVC) with an equal perceptual video quality [1]-[2].

In the HEVC encoder, a frame is initially divided into slices and each slice could be processed in parallel. To achieve a flexible and efficient block partitioning structure in HEVC, each slice is partitioned into coding tree units (CTU) [3]. CTU is analogous to the macroblock (MB) structure of H.264/AVC, but with a flexible size. Moreover, to maintain a high degree of flexibility, a CTU can be split into multiple coding units (CU) with variable sizes of 64×64, 32×32, 16×16, and 8×8. For this purpose, each CTU contains quad-tree syntax or a coding tree that specifies sub-division into CUs.

For prediction of each coding unit (CU), HEVC generally allows 11 prediction modes [3] using appropriate prediction units (PU). In order to achieve the best prediction, the HEVC encoder calculates rate distortion costs (RD cost) for all possible combinations of PUs and CUs for a CTU. The combination that gives the minimum RD cost is considered as the best mode and CTU is encoded in that CU size with that PU mode. For this reason, the RD optimization (RDO) technique requires many RD cost calculations. This process significantly increases the encoding time of the HEVC encoder. The RD cost optimization (RDO) is one of the most curtail and decision-making phenomenon in terms of the perceptual quality of the encoded video. The traditional Lagrangian multiplied based cost function is used in the RDO. The RD cost value is calculated using (1), where D is the distortion (calculating using SSE), R is the rate value and λ is the corresponding Lagrangian multiplier.

$$J = D + \lambda R \quad (1)$$

Manuscript received December 20, 2017; Revised December 22, 2017; Accepted December 22, 2017. (ID No. JMIS-2017-0053)

Corresponding Author (*): Kalyan Goswami, MulticoreWare India (P) Ltd, E-mail: kalyan@multicorewareinc.com.

²Broadcasting and Telecommunications Convergence Media Research Department, Electronics and Telecommunications Research Institute, Republic of Korea.

³Dept. of IT Engineering, Sookmyung Women's University, Seoul, Korea, Tel: 82-2-2077-7293, E-mail: bg.kim@sm.ac.kr.

A simplified flow chart of the RDO technique in the HM is shown in the Fig. 1. In this diagram, we did not consider the CU level optimization. In HM, for a given CU structure, the RD cost values of every possible combination of PU modes are calculated. The first RD cost is calculated for the $2N \times 2N$. The PU mode with the lowest RD cost is considered as the best mode and the corresponding RD cost is J_{best} . In the Loop, every time the J_{best} is compared with the J_{temp} and the corresponding value is updated. Since, the RD cost value is a linear combination of rate and distortion, most of time due to little improvement in the RD cost value, highly distorted blocks are selected. In this paper, we have explored the RD cost calculation of HEVC encoder and come up with a novel two-step RD cost calculation technique. In the proposed method, along with the conventional RD cost calculation method, the structural similarity index (SSIM) is also used. Moreover, we propose an approximated version of SSIM, FSSIM, which has less computational complexity compared to SSIM.

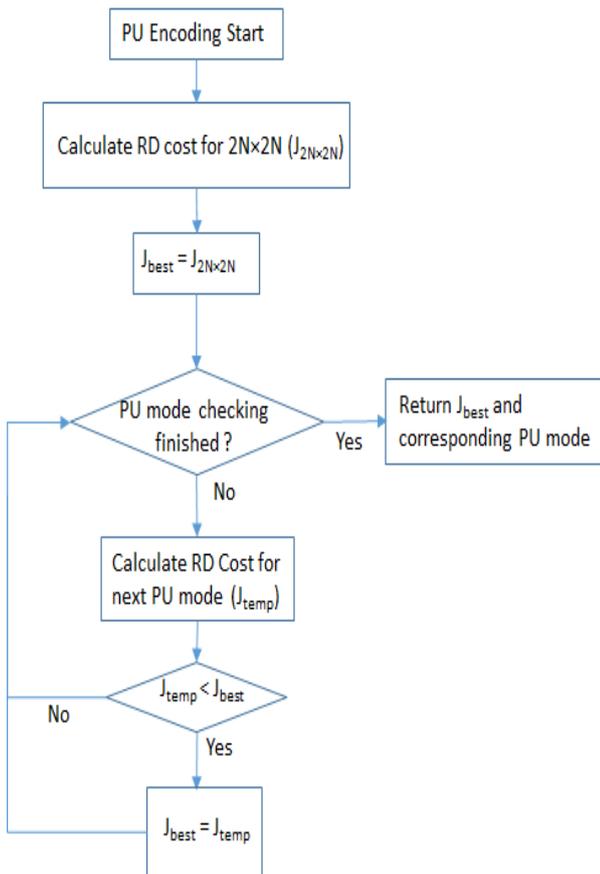


Fig. 1. Flow chart of the RDO in the HEVC encoder.

In the next section, an analysis of the RDO in the HEVC encoder is discussed. The proposed two-step algorithm described in section 3. Experimental results are shown in section 4. Finally, conclusions of the proposed work is drowned in the section 5.

II. ANALYSIS

The RD cost value of the HEVC encoder is a linear combination of rate and distortion. In the previous section, the RDO of the HEVC encoder is described. In this section, we analyze the number of selected CUs that is selected by RDO in the HEVC encoder, but have higher distortion value than the non-selected one. A very high-level architecture of this analysis is shown on the Fig. 2. In this experiment, we have checked the J_{best} and J_{temp} values if the condition $J_{temp} > J_{best}$ is satisfied (Figure 2). After that the distortion values of the best and temp modes (D_{best} and D_{temp} respectively) are checked. If the conditions (2) is satisfied, then only the count for the number of CUs is increased. In this expression, ΔRD represents $(J_{temp} - J_{best})$ and δ is threshold value.

$$D_{temp} > D_{best} \text{ and } \Delta RD\% < \sigma \quad (2)$$

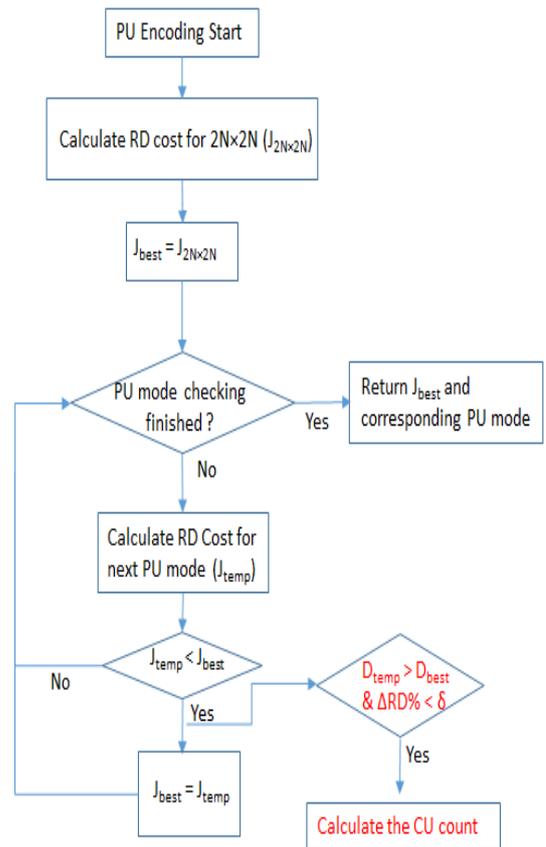


Fig. 2. Selected CUs which do not have lowest distortion values.

We have analyzed this phenomenon for 7 sequences (*BQTarrace*, *Cactus*, *Nabuta*, *ParkScene*, *PeopleOnStreet*, *SteamLocomotiveTrain*, and *Traffic*) with first 100 frames at target bitrate 4 Mbps. In this analysis we have plotted the number of CUs which are selected by the RDO of the HEVC encoder having higher distortion than the non-

selected one (Figure 2). In this context we have restricted the RD cost increment by using a parameter δ as shown in the flow chart in the Figure 3.

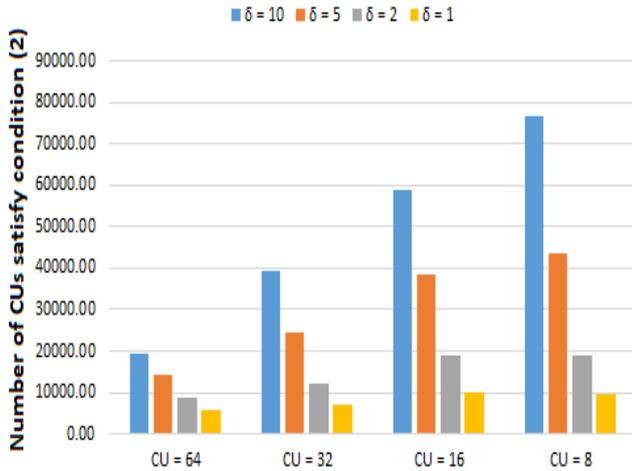


Fig. 3. Number of CUs which satisfy the condition (2) in the Fig. 2.

From this analysis, it is quite clear that the RDO in the HEVC encoder could not accurately select the lowest distorted block. Moreover, if the RDO condition is relaxed only 10%, then at the order of thousands of CUs can be selected with better subjective quality.

III. PROPOSED WORK

From the analysis discussed in the previous section, it is quite clear that good number of CUs selected as best mode in HEVC which has RD cost slightly better than the second best, but the distortion in the best CU is quite higher than other modes. One way to solve this problem is to introduce an extra checking for the distortion.

There are different ways to check the distortion or subjective quality of a CU. In the HEVC, only mean square error (MSE) or peak signal to noise ratio (PSNR) is used as the objective quality metric. However, PSNR is generally weakly correlates with subjective quality. Good amount of literatures provides the evidence that SSIM is stable and reliable perceptual quality metric than MSE. In [4], it is shown that PSNR is an unreliable video quality metric when different contents are considered in the test conditions. A relationship between SSIM and PSNR is shown in [5]. Moreover, in [6] a SSIM based RDO technique has been proposed for H.265/AVC and it has been shown that the proposed scheme can achieve significantly better rate-SSIM performance and provide better visual quality than conventional RDO coding schemes.

3.1. Two-Step RDO Algorithm

The abovementioned references, point out that if we can

use SSIM in RDO for HEVC, then it should enhance the overall visual quality. In our first attempt, we have developed an algorithm called two-step RDO algorithm. The block diagram of the proposed algorithm is shown in the Figure 4.

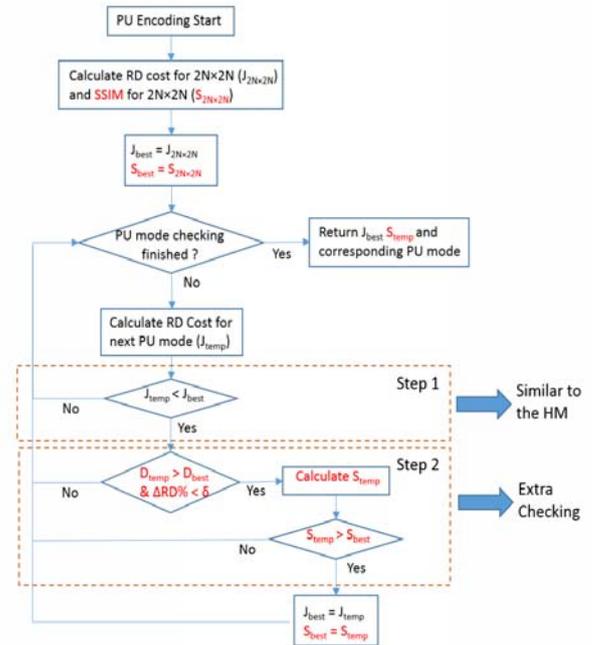


Fig. 4: Block diagram of the proposed two-step algorithm.

Like the HEVC rate-distortion optimization, initially the RD cost is calculated for the $2N \times 2N$ block. Along with that SSIM value is also calculated. As shown in the Fig. 4, in the step 1, the normal RDO is performed. In the second step, criteria stated in (2), is checked. If this criterion is satisfied, then only a SSIM based distortion condition is checked. This is the extra checking in this algorithm. Moreover, best RD cost and best SSIM values are updated for the CTU.

3.2. Fast SSIM (FSSM) Calculation Algorithm

In the above-mentioned algorithm, one of the most important block is the SSIM calculation. According to the literatures, SSIM can be calculated using the (3).

$$SSIM = \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1} \times \frac{2\delta_{xy} + c_2}{\delta_x^2 + \delta_y^2 + c_2} \quad (3)$$

All the abbreviations shown in (3) are described in [8]. In this equation μ_x and μ_y are the mean values of the reference and current blocks respectively. Standard deviations and covariance of the current and reference blocks are represented as δ_x , δ_y and δ_{xy} respectively. In this equation, c_1 and c_2 are the constants. However, the computational

complexity of (3) is quite high. Hence, a fast but accurate SSIM calculation algorithm is required. In the [7], a fast SSIM calculation technique is proposed (called dSSIM) using MSE which is given below:

$$dSSIM = \frac{1}{SSIM} \approx 1 + \frac{MSE}{2\sigma_x^2 + c_2} \quad (4)$$

Where δ_x^2 is the variance of the current block and c_2 is the constant. In the (4), MSE represents mean square error value. Now in this approach, too many assumptions are taken which makes it inaccurate. To overcome this problem, we have developed a novel fast SSIM calculation algorithm with less approximation. Our proposed Fast SSIM (FSSIM) calculation technique is given below:

$$FSSIM = \frac{2R}{1+R^2 + \frac{C_1}{E[X]^2}} \times \frac{C_2 + \varepsilon}{MSE - \Delta + C_2 + \varepsilon} \quad (5)$$

of mean square error (MSE). The MSE value is calculated in any modern hybrid video encoder. So, by reusing this in the FSSIM calculation, reduces the complexity of FSSIM calculation. Apart from the MSE, only mean values (μ_x and μ_y) of the current and the reference blocks are used here.

In the overall algorithm, we have used this proposed FSSIM calculation technique. Hence, in the proposed two-step RDO algorithm, the FSSIM value calculation is even optimized by the proposed technique.

IV. PERFORMANCE ANALYSIS

The proposed algorithm has been implemented on the HM reference software version 15.0 [9]. The main purpose of this work is to enhance the perceptual quality of the encoded video stream. One of the most challenging problems in this experiment is the evaluation of subjective quality of an encoded bitstream. However, the best way to evaluate the subjective quality is by using human

$$\text{Where } R = \frac{E[Y]}{E[X]} = \frac{\mu_y}{\mu_x}, \quad \varepsilon = \text{error function,}$$

$$\Delta = (E[X] - E[Y])^2 = (\mu_x - \mu_y)^2$$

The proof of the proposed FSSIM calculation is given in the appendix. In the (5), R is the ratio of mean values of current and reference blocks respectively. $E[x]$ is the expectation or mean of the reference block. Just like (3), c_1 and c_2 are the constants. Δ is the square of the difference of the mean values of current and reference blocks. The mean square error is represented here as MSE.

The FSSIM is an approximated version of SSIM value, which has less computational complexity than SSIM. In (5), C_1 and C_2 represent constant. The value of these constants is same as the SSIM expression (3). The most important part of the (5) is that, the FSSIM is expressed as a function

intervention, by checking the quality of the bitstream and then score manually. This technique is generally referred as mean opinion score (MOS), where five levels of accuracy is used.

This method of subjective quality evaluation is highly time consuming and expensive. Moreover, the score varies person to person. So, at least 20 scores are accumulated from different person and the mean of these points is use as a single entry. However, there is another way to calculate the subjective score, by using suitable tools. In this experiment, we have used Tektronix tool to evaluate the subjective quality. This tool provides Differential Mean Opinion Score (DMOS) value. In the DMOS value, the lower value is better.

In this experiment, we have used three videos: *BQ Terrace*, *Park Scene* and *Cactus* with resolution of 1920X1080. The target bitrate is 4 Mbps. We have compared our result with HM reference software version

Table 1. Overall performance of the proposed algorithm.

Sequence	Target bitrate	Resolution	Proposed Two-Step Algorithm				HM 15.0		
			δ	Encoding time	Bit rate	DMOS	Encoding time	Bit rate	DMOS
BQ Terrace (100 frames)	4 Mbps	1920X1080	1	4701.68	3981.07	9.64	4768.78	4018.86	9.80
			2	4759.44	4000.83	9.65			
			5	4768.78	4018.86	9.70			
			10	4689.47	3976.23	9.78			
Park Scene (100 frames)	4 Mbps	1920X1080	1	7475.20	4001.52	8.40	8820.74	4097.75	8.94
			2	7422.98	4125.13	8.71			
			5	7441.77	4065.76	9.09			
			10	7497.74	4000.03	9.22			
Cactus (100 frames)	4 Mbps	1920X1080	1	7999.79	4006.26	8.63	8083.72	4007.16	8.53
			2	7961.22	4006.26	8.76			
			5	8083.72	4007.16	8.57			
			10	8062.60	4006.40	8.73			

15.0. The results are shown in the Table 1. There are four δ values are used in this experiment.

From the Table 1, it is quite clear that when the δ value increases, DMOS value is also increases. In other words, subjective quality decreases when the δ increases. On the other hand, for *BQ Terrace, Park Scene* sequences the subjective quality of the proposed algorithm is always better than the HM 15.0. In the *Cactus* sequence quality wise, we do not get much improvement. However, in terms of encoding timing, proposed algorithm always provides superior results than the HM 15.0.

V. CONCLUSIONS

In this work we have proposed a novel approach to enhance the overall rate distortion optimization process of HM reference software. The proposed approach can improve the overall subjective quality of the HM. The proposed algorithm is performed in two steps. In the first step, like HEVC, it performs general rate distortion optimization. The second step is an extra checking where a SSIM based cost is evaluated. Moreover, a fast SSIM calculation technique is also proposed in this paper.

REFERENCES

- [1] G. J. Sullivan, J. R. Ohm, W. J. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649-1668, Dec. 2012.
- [2] J. R. Ohm, G. J. Sullivan, H. Schwarz, T. K. Tan, and T. Wiegand, "Comparison of the coding efficiency of video coding standards – including High Efficiency Video Coding (HEVC)," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1669-1684, Dec. 2012.
- [3] I. K. Kim, J. Min, T. Lee, W. J. Han, and J. H. Park, "Block partitioning structure in the HEVC standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1697-1706, Dec. 2012.
- [4] Q. Huynh-Thu, and M. Ghanbari, "Scope of validity of PSNR in image/video quality assessment," *IEEE Electronics letters*, Vol. 43, No. 13, pp.800-801, 2008.
- [5] A. Hore, and D. Ziou "Image quality metrics: PSNR vs. SSIM", *International Conference on Pattern Recognition (ICPR)*, 2010.
- [6] Shiqi Wang, Abdul Rehman, Zhou Wang, Siwei Ma, and Wen Gao, "SSIM-Motivated Rate-Distortion Optimization for Video Coding", *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 4, pp. 516-529, Sept. 2011.
- [7] Chuohao Yeo, Hui Li Tan, and Yih Han Tan, "On Rate Distortion Optimization Using SSIM", *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 7, pp. 1170-1181, Jan. 2013.
- [8] Zhou 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, April. 2004.
- [9] HM ref software, <https://hevc.hhi.fraunhofer.de/HM-doc/>, 2017.

Authors

APPENDIX

$$(A) \quad \delta_{xy} = E[XY]E[X]E[Y] \quad (a)$$

Proof

$$\begin{aligned} \delta_{xy} &= E[(X - \mu_x)(Y - \mu_y)] = E[(X - E[X])(Y - E[Y])] \\ &= E[XY - XE[Y] - YE[X] + E[X]E[Y]] \\ &= E[XY] - E[X]E[Y] \end{aligned}$$

$$(B) \quad \text{MSE} = \frac{1}{MN} \sum_{i=0}^M \sum_{j=0}^N (x_{ij} - y_{ij})^2 = \delta_x^2 + \delta_y^2 - 2\delta_{xy} + (\mu_x - \mu_y)^2 \quad (b)$$

Proof

$$\begin{aligned} \delta_x^2 + \delta_y^2 - 2\delta_{xy} &= \text{var}(x) + \text{var}(y) - 2\delta_{xy} \\ &= E[X^2] - E[X]^2 + E[Y^2] - E[Y]^2 - 2\delta_{xy} \\ &= E[X^2] - E[X]^2 + E[Y^2] - E[Y]^2 - \{2E[XY] - 2E[X]E[Y]\} \\ &\text{(using eq. a)} \\ &= E[X^2] + E[Y^2] - 2E[XY] - (E[X] - E[Y])^2 \\ &= \frac{1}{MN} \sum_{i=0}^M \sum_{j=0}^N (x_{ij} - y_{ij})^2 - (\mu_x - \mu_y)^2 \\ &= \text{MSE} - (\mu_x - \mu_y)^2 \end{aligned}$$

$$(C) \quad \text{SSIM} = \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1} \times \frac{2\delta_{xy} + c_2}{\delta_x^2 + \delta_y^2 + c_2} = P_1 \times P_2$$

$$P_1 = \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1} = \frac{2E[X]E[Y] + c_1}{E[X]^2 + E[Y]^2 + c_1} = \frac{2E[Y]/E[X] + c_1/E[X]^2}{1 + E[Y]^2/E[X]^2 + c_1/E[X]^2}$$

$$= \frac{2R + c_1/E[X]^2}{1 + R^2 + c_1/E[X]^2}, \text{ where } R = \frac{E[Y]}{E[X]}$$

$$\approx \frac{2R}{1 + R^2 + c_1/E[X]^2}$$

$$P_2 = \frac{2\delta_{xy} + c_2}{\delta_x^2 + \delta_y^2 + c_2}$$

$$\frac{1}{P_2} - 1 = \frac{\delta_x^2 + \delta_y^2 - 2\delta_{xy}}{2\delta_{xy} + c_2} = \frac{\text{MSE} - (\mu_x - \mu_y)^2}{2\delta_{xy} + c_2} \quad (\text{using eq. b})$$

$$= \frac{\text{MSE} - (E[X] - E[Y])^2}{2(E[XY] - E[X]E[Y]) + c_2} \quad (\text{using eq. a})$$

$$\approx \frac{\text{MSE} - \Delta}{c_2}$$

if X and Y are independent, then $E[XY] = E[X]E[Y]$

$$P_2 = \frac{c_2}{\text{MSE} - \Delta + c_2} \quad \text{where } \Delta = (E[X] - E[Y])^2$$

$$\text{SSIM} \approx \text{FSSIM} = \frac{2R}{1 + R^2 + \frac{c_1}{E[X]^2}} \times \frac{c_2 + \varepsilon}{\text{MSE} - \Delta + c_2 + \varepsilon}$$

$$R = \frac{E[Y]}{E[X]} = \frac{\mu_y}{\mu_x},$$

$$\varepsilon = \text{error function}, \quad \Delta = (E[X] - E[Y])^2 = (\mu_x - \mu_y)^2$$