Simplification of Intra Prediction Mode

Soon-kak Kwon[†], Amal Punchihewa[††], and Donald G. Bailey[††]

[†]Dongeui University, Korea, [††]Massey University, New Zealand

Summary
An intra prediction method that uses neighbor pixels within current picture has been adopted for improving coding efficiency, but the intra prediction having a large number of prediction modes increases the implementation operation at the encoder side. In order to reduce the operation, we first investigate the usage distribution of the intra prediction modes according to the coded picture characteristics. Then a simplification method of the intra prediction modes is proposed. The prediction modes that are utilized more than 8% are considered important but the others are excluded in the intra prediction operation. The exclusion modes are updated on a picture by picture basis after the scene change detection. From the simulation results, we can verify that approximately 50% reduction is possible in operation complexity for the intra prediction through the simplification of intra prediction modes without significant losses in picture quality.
Key words:
Intra prediction, Prediction mode, Simplification, Video coding.

1. Introduction

The standard of H.264/AVC [1,2] was established by the consultation between the VCEG (Video Coding Expert Group) of the ITU-T and the MPEG (Moving Picture Group) of the ISO/IEC. By using various coding algorithms, it is possible to reduce twice a coding bitrate in the same picture quality compared to that of the MPEG-2. The H.264 intra prediction method has been adopted in order to remove the spatial redundancy of signals within the current picture [3,4]. Also a much extended prediction algorithm is now being considered in the part of the high efficiency video coding (HEVC) [5], which will be established by the joint collaborative team on video coding (JCT-VC) of ITU-T and ISO/IEC.
H.264 supports 4 prediction modes for 16x16 blocks and 9 prediction modes for 4x4 blocks in the case of coding for a macroblock unit. HEVC may support up to much more prediction modes [6,7,8]. In the intra prediction process, the best matching block can be chosen by calculating and comparing the predictions from all of the possible modes. Although the intra prediction proposed by the international standard represents excellent picture quality, it may cause complex calculations and decrease its processing efficiency.
Some specific methods are used to reduce the complexity of intra prediction coding, such as a method that reduces the candidate of the prediction modes based on the correlation between neighboring blocks and the sum of absolute transformed difference (SATD) between the original block and the intra-predicted block [9], a method that presents a fast mode decision algorithm based on local edge information [10], and a method that uses jointly both spatial and transform domain features of the target block [11].
However, it is necessary that the simplification should be controlled according to the video characteristics and can be acceptable of much large number of the prediction modes as in HEVC, and can be applied to the different intra prediction algorithms.
For achieving the goal of the simplification of the intra prediction, this paper implements the H.264 intra coding algorithm and then investigates the usage distribution of the intra prediction modes according to the distance of intra-coded pictures of various video characteristics. Also, by analyzing the usage distribution of the prediction modes, this paper determines which modes are frequently used and which are rarely used depending on the video characteristics. Although the H.264 video coding standard has proposed 9 different prediction modes for a block, using all of these modes significantly increases the complexity of prediction processing.
Thus, this paper examines the coding efficiency for the case that excludes the rarely used modes through analyzing the prediction mode characteristics. Then, this paper proposes a desirable simplification method for the prediction modes according to video characteristics.
The remainder of this paper consists of five sections. Section 2 describes the H.264 intra coding method and prediction modes. Section 3 analyzes the usage distribution of prediction modes according to the video characteristics. Section 4 describes an algorithm that simplifies the method of selecting the prediction modes according to the video characteristics and then investigates the changes in picture quality after simplifying the prediction modes. Finally, Section 5 concludes this paper and describes future research.

2. H.264 Intra Prediction

H.264 uses the methods of predicting intra-coded macroblocks to reduce the high amount of bits coded by

Manuscript received August 5, 2011
Manuscript revised August 20, 2011

original input signal itself. For encoding a block or macroblock in Intra-coded mode, a prediction block is formed based on previously reconstructed blocks. The residual signal between the current block and the prediction is finally encoded. For the luma samples, the prediction block may be formed for each 4 x 4 subblock, each 8 x 8 block, or for a 16 x 16 macroblock. One case is selected from a total of 9 prediction modes for each 4 x 4 and 8 x 8 luma blocks; 4 modes for a 16 x 16 luma block; and 4 modes for each chroma blocks [4].

Fig. 1 shows a 4 x 4 luma block that is to be predicted. For the predicted samples [a, b, …, p] for the current block, the above and left previously reconstructed samples [A, B, …, M] are used according to direction modes. The arrows in Fig. 1 indicate the direction of prediction in each mode.

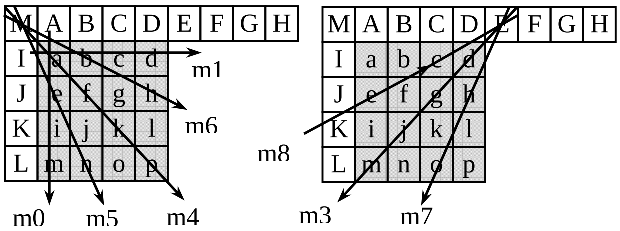

Fig. 1 Intra 4 x 4 prediction mode directions (vertical : 0, horizontal : 1, DC : 2, diagonal down left : 3, diagonal down right : 4, vertical right : 5, horizontal down : 6, vertical left : 7, horizontal up : 8).

For mode 0 (vertical) and mode 1 (horizontal), the predicted samples are formed by extrapolation from upper samples [A, B, C, D] and from left samples [I, J, K, L], respectively. For mode 2 (DC), all of the predicted samples are formed by mean of upper and left samples [A, B, C, D, I, J, K, L]. For mode 3 (diagonal down left), mode 4 (diagonal down right), mode 5 (vertical right), mode 6 (horizontal down), mode 7 (vertical left), and mode 8 (horizontal up), the predicted samples are formed from a weighted average of the prediction samples from A to M. For example, samples a and d are respectively predicted by round(I/4 + M/2 + A/4) and round(B/4 + C/2 + D/4) in mode 4, also by round(I/2 + J/2) and round(J/4 + K/2 + L/4) in mode 8. The encoder may select the prediction mode for each block that minimizes the residual between the block to be encoded and its prediction.

For prediction of each 8 x 8 luma block, one mode is selected from the 9 modes, similar to the (4x4) intrablock prediction. For prediction of all 16 x 16 luma components of a macroblock, four modes are available. For mode 0 (vertical), mode 1 (horizontal), mode 2 (DC), the predictions are similar with the cases of 4 x 4 luma block. For mode 4 (Plane), a linear plane function is fitted to the upper and left samples.

Each chroma component of a macroblock is predicted from chroma samples above and/or to the left that have previously been encoded and reconstructed. The chroma prediction is defined for three possible block sizes, 8 x 8 chroma in 4:2:0 format, 8 x 16 chroma in 4:2:2 format, and 16 x 16 chroma in 4:4:4 format. The 4 prediction modes for all of these cases are very similar to the 16 x 16 luma prediction modes, except that the order of mode numbers is different: mode 0 (DC), mode 1 (horizontal), mode 2 (vertical), and mode 3 (plane).

3. Usage Distribution of Intra Prediction Modes

In order to investigate the usage distribution of the intra prediction modes, the videos that show the various spatially directional pixels are selected. The coding was performed using the JM encoder [12] for a total of 45 pictures with a bitrate of 1Mbps as a baseline profile in which 352 pixels for the horizontal direction and 288 pixels for the vertical direction were used for each video.

3.1 Usage Frequency of Prediction Modes According to the Intra-Coded Picture Distance

The intra-coded picture (I-picture) is periodically inserted for the random access within video sequence. Normally the I-picture distance is not more than 15 by considering the coding efficiency and random access. An I-picture distance was chosen less than 15 to investigate the usage distribution of the prediction modes. The usage frequency of the prediction modes according to the I-picture distance was obtained using all of 9 modes for each 4x4 luma block. Table 1 shows the simulation results of the prediction modes for the Bus video. In the case of I-period 14, it means that 14 pictures are coded by temporally prediction between both neighbored I-pictures. Average value of usage frequency of each prediction mode is not largely distributed differently according to the I-picture distance.

Table 1: Average usage frequency of prediction modes for Bus video

I-period	mode0	mode1	mode2	mode3	mode4	mode5	mode6	mode7	mode8
2	1705	1487	1514	259	267	254	267	247	336
3	1789	1537	1621	221	230	217	228	209	284
4	1726	1507	1643	232	243	229	241	222	293
5	1707	1544	1349	267	283	278	290	262	356
6	1704	1576	1216	281	308	291	303	281	376
7	1729	1558	1107	300	323	292	325	292	410
8	1709	1607	1003	303	332	323	336	311	412
9	1711	1586	905	304	359	338	366	319	448
10	1712	1621	888	314	362	319	346	328	446
11	1709	1624	899	318	349	326	344	314	453
12	1651	1657	849	319	361	353	353	346	447
13	1711	1602	841	341	373	335	360	322	451
14	1689	1628	828	325	372	342	359	329	464

As shown in Table 1, the prediction modes 0, 1, and 2 were used most frequently for all of I-periods. We can also see that the usage frequency of the prediction mode is distributed without large dependence of the I-picture distance in the same video sequence.

3.2 Usage Distribution of Prediction Modes According to Video Characteristics

Table 2 shows the frequency of the prediction modes according to video characteristics.

Table 2: Usage frequency (%) of prediction modes in video characteristics

video/mode	0	1	2	3	4	5	6	7	8
Bus	26	25	18	5	5	5	5	5	7
Flower	14	11	43	5	4	4	6	4	9
Foreman	16	15	31	4	12	5	6	3	8
Waterfall	13	15	24	7	10	6	10	5	9

Prediction modes 0 and 1 were used extensively in the Bus video. The other three videos showed a high frequency in the use of prediction mode 2, that uses average values. In particular, the Flower video applied the average values (DC) more than the other videos because it showed an even distribution in similar colors while there were no particular moving objects in video. However, the Foreman and Waterfall videos showed large movements in the center area with little movement in the background, resulting in a more even distribution among prediction modes 0, 1, and 2. In all of the test video sequences, prediction modes 3, 5, and 7 were used less than 8% of the time.
As a result, although there were some differences in the use of the prediction modes, modes 0, 1, and 2 showed the highest usage frequencies for intra prediction. About half of all prediction modes showed less than 8% usage regardless of their video characteristics. This implies that the method of selecting the prediction modes could be simplified.

4. Simplification Method of Intra Prediction Modes and Simulation Results

From the previous analysis, it was recognized that the prediction modes were different according to the video characteristics. Thus, this paper proposes a simplification method for the intra prediction. The proposed algorithm is described in Fig. 2.

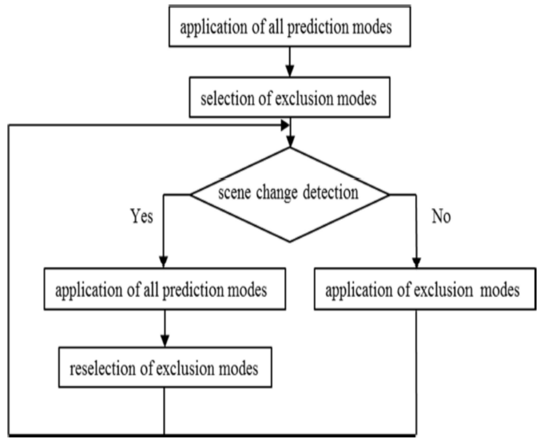

Fig. 2 Proposed algorithm for the simplification of intra prediction modes

Initial picture is applied by all of modes for the intra prediction. Then the usage distribution for each mode is calculated and small frequently modes are excluded for applying next picture. However, because the video can show a significant change in its characteristics after scene change, a detection method of scene change is used. Various methods for the scene change detection have been conducted. This paper adopts the method [13] that detects the scene change by considering the motion vector in coded bit streams and the characteristics in a discrete cosine transform region in order to perform a simple implementation. In the case of the decision of scene change, all of prediction modes are applied and then small frequently modes are reselected to exclude the prediction modes for applying to the next pictures. For convenience, the prediction mode will be excluded if it was used less than 8%.

For the simulation, a total of 180 pictures were used as a merged video sequence which composes of four different types of 45 pictures within Bus, Flower, Foreman, and Waterfall videos. The coding was performed using the JM encoder with a baseline profile of 1Mbps on the 352x288 resolution video sequence. Each picture has a total of 6336 4x4 luma blocks. Table 3 shows the results of comparing the objective picture quality, the PSNR (Peak Signal to Noise Ratio) values for the reconstructed video with and without the simplification method proposed in this paper. Based on these results, excluding the prediction modes which used less than 8% of the time showed a very small effect on the reconstruction fidelity, less than 0.1dB in all cases compared to the conventional method. In the searching operation for the intra prediction, 50% reduction was achieved.

Table 3: PSNR values after the simplification of prediction modes (merged video)

I-period	conventional method (dB)	proposed method (dB)
14	37.021	36.934
13	36.933	36.823
12	36.783	36.702
11	36.648	36.576
10	36.456	36.383
9	36.253	36.156
8	35.532	35.439
7	34.808	34.709
6	34.096	34.013
5	33.425	33.335
4	31.488	31.406
3	30.904	30.811

In the Bus video within the merged sequence, six prediction modes 3 to 8 were excluded by using the proposed method. It is possible to implement a simplification prediction process that reduces a 66% in the search operations, due to the exclusion of such 6 modes from all of 9 prediction modes. Also, in the case of the Flower video, the simplification was achieved by 55% through excluding 5 prediction modes, modes 3 to 7. In the case of the Foreman video, the simplification was achieved by 44% through excluding 4 prediction modes, modes 3, 5, 6, and 7. In addition, in the case of the Waterfall video, the simplification was achieved by 33% through excluding the prediction modes 3, 5, and 7. It can be considered that it is possible to achieve the simplification of the calculation in the prediction modes, minimum 33% to maximum 66%, by applying the exclusion prediction modes, which is used less than 8% of the time, to the next intra prediction.

5. Conclusion

This paper demonstrates that it is possible to simplify the calculation of the intra prediction modes. Those modes with frequency below 8% are neglected when predicting of the next intra coded picture. This processing is adaptive to the content of the video, and results in a reduction in the computational complexity of the prediction process by approximately 50%. This simplification of processing does not have a significant impact on the picture quality. This would significantly reduce the coding delay enabling the real-time continuous transmission or streaming of videos through a network. Therefore, it is expected that this simplification which reduces the complexity of the intra coding in the encoder side will be applied to HEVC [5-8] and various applications for region of interest [14-19].

Acknowledgments

This work was supported by Dongeui University Grant. (2010AA180)

References

[1] ITU-T Rec. H.264 / ISO/IEC 11496-10, “Advanced video coding,” Final Committee Draft, Document JVTF100, 2002.

[2] I. E. G. Richardson, H.264 and MPEG-4 video compression : video coding for next-generation multimedia, Wiley, 2003.

[3] T. Wiegand, G. Sullivan, G. Bjøntegaard, and A. Luthra, “Overview of the H.264/AVC video coding standard,” *IEEE Trans. Circuit Syst. Video Technology*, Vol. 13, No. 7, pp. 560-576, 2003.

[4] S.-k. Kwon, A. Tamhankar, and K. R. Rao, “Overview of H.264/MPEG-4 part 10,” *Journal of Visual Communications and Image Representation*, Vol. 17, No. 2, pp. 186-216. 2006.

[5] Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, WD3: working draft 3 of high-efficiency video coding, March 2011.

[6] K. Ugur, K. Andersson, A. Fuldseth, G. Bjøntegaard, L. P. Endresen, J. Lainema, A. Hallapuro, J. Ridge, D. Rusanovskyy, C. Zhang, A. Norkin, C. Priddle, T. Rusert, J. Samuelsson, R. Sj¨oberg, and Z. Wu, “High performance, low complexity video coding and the emerging HEVC standard,” *IEEE Trans. Circuit Syst. Video Technology*, Vol. 20, No. 12, pp. 1688-1697, 2010.

[7] W.-J. Han, J. Min, I.-K. Kim, E. Alshina, A. Alshin, T. Lee, J. Chen, V. Seregin, S. Lee, Y. M. Hong, M.-S. Cheon, N. Shlyakhov, K. McCann, T. Davies, and J.-H. Park, “Improved video compression efficiency through flexible unit representation and corresponding extension of coding tools,” *IEEE Trans. Circuit Syst. Video Technology*, Vol. 20, No. 12, pp. 1709-1720, 2010.

[8] F. Bossen, V. Drugeon, E. Francois, J. Jung, S. Kanumuri, M. Narroschke, H. Sasai, J. Sole, Y. Suzuki, T. K. Tan, T. Wedi, S. Wittmann, P. Yin, and Y. Zheng, “Video coding using a simplified block structure and advanced coding techniques,” *IEEE Trans. Circuit Syst. Video Technology*, Vol. 20, No. 12, pp. 1667-1675, 2010.

[9] M. G. Sarwer, L.-M. Po, and Q. M. J. Wu, “Fast sum of absolute transformed difference based 4×4 intramode decision of H.264/AVC video coding standard,” *Signal Processing: Image Communication*, Vol. 23, No. 8, pp. 571-580, 2008.

[10] F. Pan, X. Lin, S. Rahardja, K. P. Lim, Z. G. Li, D. Wu, and S. Wu, “Fast mode decision algorithm for intraprediction in H.264/AVC video coding,” *IEEE Trans. Circuit Syst. Video Technology*, Vol. 15, No. 7, pp. 813-822, 2005.

[11] C. Kim, H.-H. Shih, and C.-C. J. Kuo, “Fast H.264 intra-prediction mode selection using joint spatial and transform

domain features," *Journal of Visual Communications and Image Representation*, Vol. 17, No. 2, pp. 291-310, 2006.
[12] JM Reference Software, http://iphome.hhi.de/suehring/tml/download.
[13] K. Tse, F. Wei, and S. Panchanathan, "A scene detection algorithm for MPEG compressed video sequences," *Electrical and Computer Engineering*, Vol. 2, pp. 827-830, 1995.
[14] H. Song and C.-C. J. Kuo, "A region-based H.263+ codec and its rate control for low VBR video," *IEEE Trans. Multimedia*, Vol. 6, No. 3, pp. 489-500, 2004.
[15] L. Tong and K. R. Rao, "Region of interest based H.263 compatible codec and its rate control for low bit rate video conferencing," *International Symposium on Intelligent Signal Processing and Communication Systems*, pp. 249-252, Dec. 2005.
[16] Y. Liu, Z. G. Li, and Y. C. Soh, "Region-of-interest based resource allocation for conversational video communication of H.264/AVC," *IEEE Trans. Circuit Syst. Video Technology*, Vol. 18, No. 1, pp. 134-139, 2008.
[17] M.-C. Chi, C.-H. Yeh, and M.-J. Chen, "Robust region-of-interest determination based on user attention model through visual rhythm analysis," *IEEE Trans. Circuit Syst. Video Technology*, Vol. 19, No. 7, pp. 1025-1038, 2009.
[18] Y. Zhang, G. Jiang, M. Yu, Y. Yang, Z. Peng, and K. Chen, "Depth perceptual region-of-interest based multiview video coding," *Journal of Visual Communication and Image Representation*, Vol. 21, No. 5-6, pp. 498-512, 2010.
[19] N. Doulamis, A. Doulamis, D. Kalogeras, and S. Kollias, "Low bit-rate coding of image sequences using adaptive regions of interest," *IEEE Trans. Circuit Syst. Video Technology*, Vol. 8, No. 8, pp. 928-934, 1998.

Soon-kak Kwon This author received the B.S. degree in Electronic Engineering from Kyungpook National University, in 1990, the M.S. and Ph.D. degrees in Electrical Engineering from Korea Advanced Institute of Science and Technology (KAIST), in 1992 and 1998, respectively. From 1998 to 2000, he was a team manager at Technology Appraisal Center of Korea Technology Guarantee Fund. Since 2001, he has been a faculty member of Dongeui University, where he is now an associate professor in the Department of Computer Software Engineering. From 2003 to 2004, he was a visiting professor of the Department of Electrical Engineering in the University of Texas at Arlington. Since 2010, he has been an international visiting research associate in the School of Engineering and Advanced Technology in Massey University. Prof. Kwon received the awards, Leading Engineers of the World 2008 and Foremost Engineers of the World 2008, from IBC, respectively. His biographical profile has been included in the 2008, 2009, 2010, and 2011 Editions of *Marquis Who's Who in the World* and the 2009/2010 Edition of *IBC Outstanding 2000 Intellectuals of the 21st Century*. His research interests are in the areas of image processing, video processing, and video transmission.

Amal Punchihewa This author obtained his Bachelor of Engineering specializing into Electronics and Telecommunication Engineering from The University of Moratuwa, in Katubedda, Sri Lanka with honours in 1986. He specialized in digital video signal processing for his Master of Electronics Engineering completed in 1991 at the Technical University of Eindhoven, The Netherlands and PhD from Massey University in 2009. His current research interests are objective assessment of compression artefacts, multi-media communication, sensor networks, image processing for surveillance and eldercare. He is a Fellow of the Institution of Engineering and Technology. In 2000, he was awarded the "Wimalasurendra award" in recognition of the contribution made to the broadcasting development by the Institution of Engineers Sri Lanka.

Donald G. Bailey This author received the B.E. (Hons) degree in Electrical Engineering in 1982, and the Ph.D. degree in Electrical and Electronic Engineering from the University of Canterbury, New Zealand in 1985. From 1985 to 1987, he applied image analysis to the wool and paper industries within New Zealand. From 1987 to 1989 he was a Visiting Research Engineer at University of California at Santa Barbara. Dr. Bailey joined Massey University in Palmerston North, New Zealand as Director of the Image Analysis Unit at the end of 1989. He is currently an Associate Professor in the School of Engineering and Advanced Technology, and leader of the Image and Signal Processing Research Group. His primary research interests include applications of image analysis, machine vision, and robot vision. One area of particular interest is the application of FPGAs to implementing image processing algorithms.