

TREE STRUCTURED HYBRID INTRA PREDICTION

Satoshi Kondo, Hisao Sasai and Shinya Kadono

Matsushita Electric Industrial Co., Ltd.
1006 Kadoma, Kadoma, Osaka, 571-8501 Japan.
{kondo.satoshi, sasai.hisao, kadono.shinya}@jp.panasonic.com

ABSTRACT

This paper proposes a new method to generate intra prediction images in intra coding, Tree Structured Hybrid Intra Prediction. When generating intra prediction images, the proposed method selects the most suitable one from intra prediction images generated using two methods, pixel-based and block-based intra prediction. This selection is performed on macroblock or sub macroblock level based on rate distortion optimization scheme. In addition, information indicating which prediction method is selected is described in the bitstream using a tree structure. The proposed method is initially described, and then the results of evaluation experiments are shown by comparison to conventional methods. Evaluation is carried out by implementing the proposed method in an H.264-based intra picture codec. The results of evaluation confirm that the proposed method improves PSNR by up to nearly 0.7 dB compared to conventional methods over a wide range of bit rates.

1. INTRODUCTION

Intra prediction is used to improve coding efficiency of intra coding. It utilizes the spatial correlation in an image to predict the block being encoded from its surrounding blocks. Intra prediction is introduced in the video coding standards such as MPEG-4 and H.264 | MPEG-4 AVC [1] (below referred to as “H.264”). In these standards, intra prediction is performed for frequency (DCT) regions and spatial (pixel) regions, respectively.

In H.264, intra prediction is performed in units of a 4×4 pixel block or a 16×16 pixel block. Frequency transform of the residual image between the image being encoded and the intra prediction image is then carried out. In the case of intra prediction with a 4×4 pixel block, one direction of prediction is selected from pre-defined 9 types of directions. The prediction image is then generated using pixels correlating to the selected prediction direction from the already encoded pixels in 13 pre-determined positions in proximity to the block being encoded. Information on which prediction direction is chosen for generation of the intra prediction

image is described in the bitstream as supplementary information. In the following, the intra prediction method used in H.264 is referred to as the Pixel-based Intra Prediction (PIP) method.

In PIP, the prediction image is generated by extrapolating pixel values neighboring the block being encoded into the block itself. Also, the position of pixels used to generate the prediction image is fixed.

As an alternative to PIP, another intra prediction method, the Block-based Intra Prediction (BIP) method has been proposed [2]. The concept of BIP is inspired by a motion compensation method using a block matching technique in inter coding. In BIP, the reference image most suited to the block being encoded is acquired from arbitrary positions in the already encoded area of the picture being encoded. And this reference image is taken as the intra prediction image. In addition, the difference between position of the block being encoded and the position of the reference image is described in the bitstream as supplementary information.

Because of these nature, the efficiency of prediction in PIP is supposed to be high for images with a large number of low frequency elements, and images in which the edges exist in one certain direction. And BIP is highly efficient in its prediction for images with 2-D patterns, for which there arises a problem in PIP that the prediction efficiency is not good. On the other hand, there is a problem that efficiency with images that contain a large number of low frequency elements is supposed to be not always high in BIP.

To address these problems, we propose the Tree Structured Hybrid Intra Prediction method (TSHIP), as a new method for the generation of intra prediction images. In this method, the most suitable intra prediction image is chosen from intra prediction images generated with the PIP and BIP methods, with a macroblock or a sub macroblock level as base unit, on the basis of rate distortion optimization. In addition, information denoting how the intra prediction image is selected in the macroblock is described using a tree structure in the bitstream as prediction mode information.

This paper is organized as follows. The proposed TSHIP method is explained in detail in Section 2. Experimental results of the proposed method by comparing with conven-

Table 1. Prediction mode information in tree structured hybrid intra prediction.

Macroblock Mode		Sub Macroblock Mode	
0	PIP	0	PIP
1	16×16 BIP	1	8×8 BIP
2	16×8 BIP	2	8×4 BIP
3	8×16 BIP	3	4×8 BIP
4	Sub macroblocks	4	4×4 BIP

tional methods are given in Section 3.

2. TREE STRUCTURED HYBRID INTRA PREDICTION

2.1. Overview

The proposed method, TSHIP, is a new method for intra prediction in intra coding. TSHIP selects the most suitable intra prediction image from intra prediction images generated with the PIP and BIP methods on the basis of rate distortion optimization scheme. The unit for selection is a macroblock of 16×16 pixels, or a sub macroblock of 8×8 pixels. In addition, prediction mode information on how the intra prediction image is selected for each macroblock is described using a tree structure in the bitstream.

The prediction mode for macroblock is shown on the left side of Table 1. When the macroblock prediction mode is 0, the prediction image for the macroblock is generated using PIP only. When the macroblock prediction mode is 1-3, the prediction image for the macroblock is generated using BIP only. And in case that the macroblock prediction mode is 2 or 3, the macroblock is divided into two blocks. For example, when the macroblock prediction mode is 2, the macroblock is divided into 2 blocks of 16 pixels in width by 8 pixels in height. And an intra prediction image is obtained from separate positions for each block. When the macroblock prediction mode is 4, PIP and BIP methods can be mixed on the sub macroblock level. In this case, the sub macroblock prediction mode is further described for each sub macroblock, as shown on the right side of Table 1. When the sub macroblock prediction mode is 0, the prediction image for the sub macroblock is generated with PIP only. When the sub macroblock prediction mode is 1-4, the prediction image is generated with BIP only. And in case that the sub macroblock prediction mode is 2-4, the sub macroblock is further divided into two or four blocks.

2.2. Selection of Prediction Mode

In this section, we describe details of the method to determine the prediction mode. First, the evaluation value J is calculated for when the macroblock prediction mode M is

0, i.e. when the prediction image is generated using PIP only, using Eq. (1). For macroblocks and sub macroblocks that use PIP, intra prediction is carried out in units of 4×4 pixel blocks in the same way as intra prediction in H.264.

$$J(M | Q, \lambda) = \min_P [D(s, d | M, P, Q) + \lambda\{R(s - r | M, P, Q) + R(M, P)\}] \quad (1)$$

Here, Q is a quantization parameter, λ is a Lagrangian multiplier. s is the original image, d is the decoded image, the first term on the right side D is the sum of the squared error between the original image s and the decoded image d . r denotes the intra prediction image, and the 2nd and 3rd terms on the right side R denotes the number of generating bits for the residual image $s - r$, and macroblock prediction mode M and the PIP prediction direction P , respectively.

Next, the evaluation value J is calculated for when the macroblock prediction mode M is 1-3, i.e. when the prediction image is generated using BIP only. The method to calculate is described below.

When BIP is used, first, pixel values for areas not yet encoded are padded by using pixel values from areas that have already been encoded. Then, intra reference image for blocks being encoded is searched by using the padded image as a reference image. When macroblocks are divided, regions that have already been encoded in the current macroblock are also used as a reference image.

The search for intra reference blocks is carried out by searching for the value m that minimizes the evaluation value L from Eq. (2).

$$L(m | \lambda) = D(s, r | m) + \lambda R(m - p) \quad (2)$$

Here, m is the difference between the position of the block being encoded and the position of the intra reference block, expressed as a 2-D vector, which we call the intra vector. The first term on the right side D denotes the sum of the squared error between the original image s and the intra reference image r when the intra vector is m . p is the prediction value for the intra vector and the 2nd term on the right side R denotes the number of generating bits of the differential of the intra vector $m - p$. Here, the prediction value of the intra vector p is obtained as the median value of the intra vectors used in the blocks neighboring the block being coded which generated prediction images with BIP. When encoding the intra vector, for example when the macroblock prediction mode M is 2, a different intra vector is encoded for each of the two 16×8 pixel blocks in the macroblock.

Once the search for intra reference blocks has finished, the evaluation value J is calculated using Eq. (3).

$$J(M | Q, \lambda) = D(s, d | M, m, Q) + \lambda\{R(s - r | M, m, Q) + R(M, m - p)\} \quad (3)$$

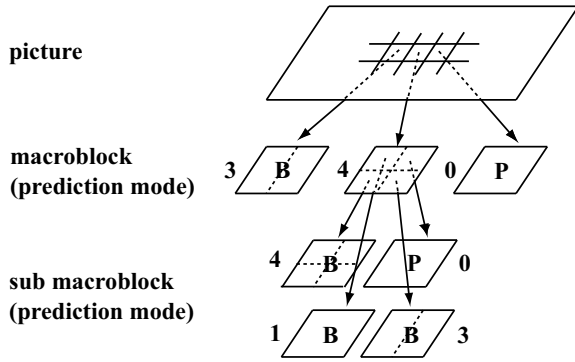


Fig. 1. An example of TSHIP prediction mode.

When the macroblock prediction mode M is 4, i.e. when PIP and BIP are mixed at the sub macroblock level, the evaluation value J is obtained from Eq. (4).

$$\begin{aligned}
 J(M | Q, \lambda) = & \sum_{b=1}^4 \min_S [D(s, d | S, P, m, Q) \\
 & + \lambda \{R(s - r | S, P, m, Q) + R(S, P, m - p)\}] \\
 & + \lambda R(M)
 \end{aligned} \quad (4)$$

Here, b denotes the sub macroblock index and S denotes the sub macroblock prediction mode. Equation (4) means that it determines the most suitable prediction mode for each sub macroblock and makes the sum of the evaluation values the macroblock evaluation value J . In addition, the prediction direction P is encoded in sub macroblocks that use PIP, and the differential of the intra vector $m - p$ is encoded in sub macroblocks that use BIP. When BIP is used, the method to search the reference block is the same way as in case that the macroblock prediction mode is BIP as explained above using Eq. (2). And sub macroblocks which have already encoded are also used as reference images.

Finally, the prediction mode for which the evaluation value J is smallest is selected and this is described in the bitstream as the macroblock prediction mode or sub macroblock prediction mode.

An example of prediction mode is shown in Fig. 1. In Fig. 1, "P" and "B" denote macroblocks and sub macroblocks for which PIP and BIP are selected, respectively. Dotted lines denote the block size in macroblocks and sub macroblocks for which BIP is selected. And the numbers indicated next to blocks are the prediction modes shown in Table 1. These prediction modes are expressed in a tree structure as shown in Fig. 1 and denoted in the bitstream.

3. EXPERIMENTAL RESULTS

In this section, we describe the results of experiments that exhibit the performance of our proposed method. In the ex-

Table 2. Experimental conditions.

Image size	SIF (352×240 pixels)
Frame rate	10 Hz
No. of frames per test sequence	100 Frames
Search range in BIP and TSHIP	$\pm 32 \times \pm 32$ pixels
Intra vector precision in BIP and TSHIP	1/4 pixel
Deblocking filter	No
Bit rate control	No (Constant quantization parameter)

periments, we used H.264 intra picture coding as the base encoding method, and the two conventional methods (PIP and BIP) and the proposed method (TSHIP) were implemented as the intra prediction method. Then, we compare the conventional methods and the proposed method. The 2-D Run/Level encoding method used in H.26L JM-1 [3] was utilized for variable length encoding of frequency transform coefficients. And Exponential Golomb code [4] was used as the variable length code for frequency transform coefficients after the 2-D Run/Level encoding, prediction modes, prediction direction in PIP, intra vectors in BIP, and so on. The experimental conditions are shown in Table 2.

Examples of the results of the experiments are shown in Figs. 2 and 3. These figures show rate distortion curves showing the change in the number of generated bits and the PSNR of the luminance component as the quantization parameters are altered. Although the results shown in Figs. 2 and 3 show PIP outperforms BIP, BIP outperforms PIP for some other test sequences. And it can also be seen that by using the proposed method, TSHIP, in comparison to the conventional methods of PIP and BIP, PSNR can be improved by up to nearly 0.7 dB over a broad range of bit rates.

We also compared the subjective quality of intra prediction images generated with the conventional methods of PIP and BIP, and evaluated the selection distribution of PIP and BIP in the proposed method. Although details of the results are not shown, we also confirm that PIP and BIP give high prediction efficiency especially for images with a large number of low frequency elements and images with 2-D texture patterns, respectively. In addition, we confirm that the proposed method selects the intra prediction image which give higher subjective quality from prediction images generated with PIP and BIP.

Figure 4 shows the ratio of macroblocks selected for PIP and BIP by the proposed method for the test sequence "Mobile and Calendar" and "Foreman". In Fig. 4, average values of the ratio over the entire sequence with several quantization parameters are shown. Note that each sub macroblock is calculated as 1/4 of a macroblock when the sub macroblock prediction mode is used.

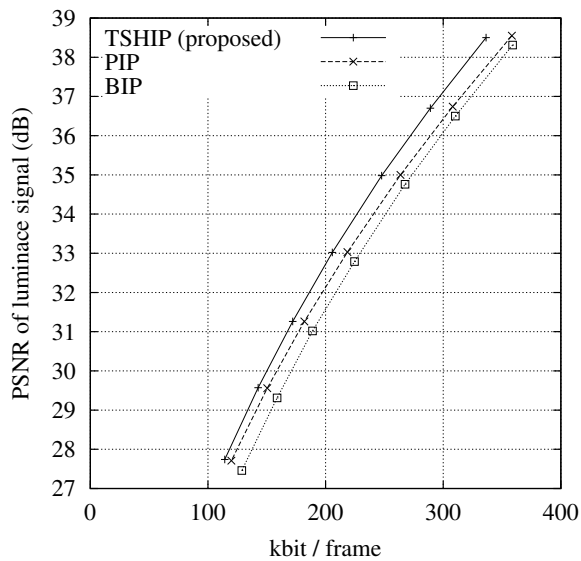


Fig. 2. Rate distortion curve for test sequence "Mobile and Calendar".

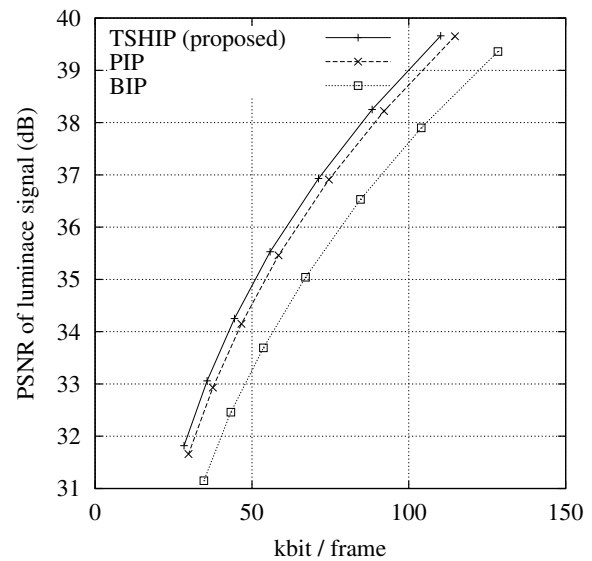


Fig. 3. Rate distortion curve for test sequence "Foreman".

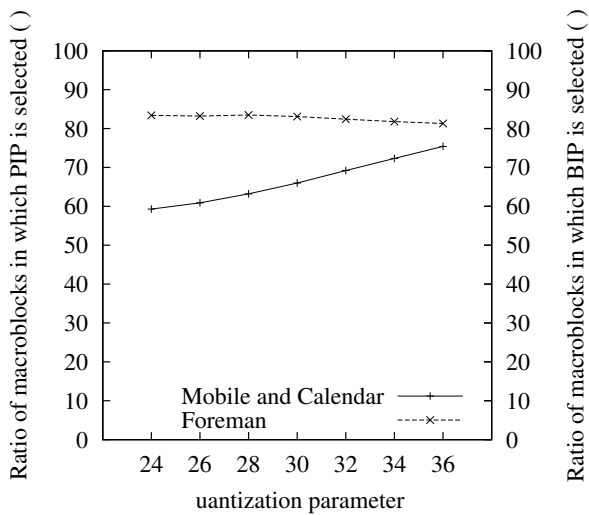


Fig. 4. Ratio of macroblocks in which PIP(BIP) is selected for each test sequence in proposed method.

From the results of test sequences shown in Fig. 4 and other test sequences, we confirm that the proposed method has the characteristic that the likelihood of which prediction method is selected between PIP and BIP depends on the sequence. Moreover, it is confirmed that the design of images is a superior deciding factor to the selection of the prediction method rather than the quantization parameter value and bit rate in the proposed method, except for the images having complex design as test sequence "Mobile and Calendar".

4. CONCLUSIONS

In this paper we proposed the Tree Structured Hybrid Intra Prediction method, a new method for generating intra prediction images in intra coding. Evaluation of the proposed method was carried out by implementing it in an H.264-based intra picture coding. From the results of the experiments, it is confirmed that the proposed method can improve PSNR by up to nearly 0.7 dB in comparison with conventional methods, over a broad range of coding rates. Directions for future work include evaluation of the capabilities of the proposed method when used as a prediction image generation method in intra macroblocks for inter coding picture.

5. REFERENCES

- [1] "Information technology — Coding of audio-visual objects — Part 10: Advanced video coding," ISO/IEC 14496-10, Dec. 2003.
- [2] S. L. Yu and C. Chrysafis, "New intra prediction using intra-macroblock motion compensation," JVT-C151, , Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG Meeting, May 2002.
- [3] "Joint model number 1 (JM-1)," JVT-A003, Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG Meeting, Jan. 2002.
- [4] J. Teuhola, "A compression method for clustered bit-vectors," Information Processing Letters, vol. 7, no. 6, pp. 308-311, Oct. 1978.