A Deblocking Method using Wavelet Transform for MPEG-II format

Tomio GOTO, Tatsuya YAMAZAKI and Masaru SAKURAI
Nagoya Institute of Technology
Department of Electrical and Computer Engineering
Gokiso-cho, Showa-ku, Nagoya 466–8555, Japan.
E-mail: {t.goto@, yama@apple.elcom., sakurai.masaru@}nitech.ac.jp

***Abstract*— JPEG and MPEG International Standards have been used as formats for compressing and storing still images and moving pictures, respectively. In these standards, DCT (Discrete Cosine Transform) has been being used for compression of data because of compression perfomance. However, there is a problem with blocky noise appearing in reconstructed images when a higher compression rate is set. This is because there are no correlations between DCT blocks as a result of quantizing DCT coefficients, and as the horizontal and vertical dimensions of reconstructed images are scaled up, blocky noise becomes correspondingly more apparent.**

There have been a lot of studies on reducing blocky noise. One of the most well-known methods is a deblocking edge filter (DEF), which is defined by the international standard of H.263. However, this method increases the computational cost and does not decrease blocky noise at low bit rates.

In this study, we propose a new decoding method of reducing blocky noise using inverse wavelet transform for moving pictures. In the proposed method, MPEG-II images, which are low resolution and include blocky noise, are transformed to wavelet transform coefficients. A masking process using these coefficients is also done. As a result of applying the proposed method to a decoder, blocky noise is decreased and the resulting image is clear.

I. INTRODUCTION

JPEG and MPEG International Standards [1], [2] have been used as formats for compressing and storing still images and moving images, respectively. In these standards, DCT (Discrete Cosine Transform) has been adopted because of compression performance and image quality. In DCT coding, the original image is divided into blocks, and the DCT is processed in each block, integrating signals in all blocks into low bands and quantizing the DCT coefficients so that compression performance can be increased. However, the higher the compression rate is set, the lower the quality of reconstructed images because of increased blocky noise. This is because there are no correlations between blocks as a result of quantizing low band signals, and since the high band signals are quantized, reconstructed images become blurry and include mosquito noise. Therefore, wavelet transform coding [3] is an attractive field because these noises can be decreased, and this coding has been adopted as the JPEG2000 International Standard [4]. However, there were many images which were encoded with low resolution devices with blocky noise in the past, and the removal of those images is intolerable. There were also many images which were encoded and shown with high resolution devices. In this case, blocky noise appears in the magnified image.

For international JPEG and MPEG standards, a block transform coding such as DCT has been used so a reconstructed image includes blocky noise at a low bit rate at these standards.

Y.Yang, N. P. Galatsanos and A. K. Katsaggelos proposed a noise removal techniques which is called POCS (Projection Onto Convex Sets) [5], [6]. This proposed method repeats the filtering processes to reduce blocky noise. However, the process takes a long time because of the repetition, so this technique is not applicable to a moving image.

There are other approaches in which the data is transformed to the DCT coefficients in a coder with significant reduction in blocky noise. This approach, which is called LOT (Lapped Orthogonal Transform) [7], can reduce the blocky noise by overlapping the blocks. Other approaches, which involve pre-filtering and post-filtering [8], have been proposed. However, these approaches cannot reduce the blocky noise in the reconstructed image. This means that a new approach to improve the image quality using coded data in a decoder is needed.

In this study, we propose a new method of reducing blocky noise using inverse wavelet transform for moving pictures. In the proposed method, MPEG images, which are low resolution and include blocky noise, are transformed to wavelet transform coefficients, and as a result, reconstructed images which are clearer than the original MPEG images and which have no blocky noise are obtained.

In this paper, we argue for the noise removal technique of the conventional and proposed methods in chapter II. We make estimates using computer simulation to check the usefulness of our method in chapter III. And we explain the conclusion in chapter IV.

II. NOISE REMOVAL

A. DEF (Deblocking Edge Filter)

DEF is one of the most well-known methods of reducing blocky noise, which is defined on ITU-T H263–Annex J [9]. In DEF, by filtering both sides of 2 pixels between the blocks in the reconstructed images with blocky noise, blocky noise is smoothed out. However, it is not enough to reduce blocky noise at low bit rates because processing is only on two sides of a two-pixel area. And every side of an average and a deviation in each DCT block needs to be calculated, which increases the computational cost.

B. Proposed Method

In this study, we propose a new decoding method of reducing blocky noise using inverse wavelet transform. We will show that DCT coefficients are the same as wavelet transform coefficients, and we will try to use the band classified coefficients to reduce blocky noise.

1) Sub-band Expression of DCT Coefficients: A basis with a one-dimensional N point DCT can be presumed as an N division analysis filter bank.

In a two-dimensional signal such as an image, the signal which is transformed to $N \times N$ DCT coefficients can be expressed as a sub-band signal (for an example of 4×4 DCT see Fig. 1). In Fig. 1 (a), each pixel is a DCT coefficient, and pixels 1 through 9 are DC components in each block. These coefficients are mapped as shown in Fig. 1 (b), and these coefficients can be considered as sub-band coefficients.

2) Decoding of DCT Coefficients using Inverse Wavelet Transform Coding: A proposed decoding flow is illustrated in Fig. 2. In JPEG and MPEG, the block size that is generally used is 8 × 8 pixels. When data is compressed, quanting low band signals, discontinuity is generated between the blocks. As a result, blocky noise is included in reconstructed images. In the proposed method, each block of 8 × 8 pixels is linearly transformed into 4 × 4 blocks of 2 × 2 pixels (see Fig. 3) using the transformation

$$\boldsymbol{X}_0' = \boldsymbol{A}\boldsymbol{X}_0\boldsymbol{A}^t \quad (1)$$

$$= \begin{pmatrix} \boldsymbol{F}_{11} & \boldsymbol{F}_{12} & \boldsymbol{F}_{13} & \boldsymbol{F}_{14} \\ \boldsymbol{F}_{21} & \boldsymbol{F}_{22} & \boldsymbol{F}_{23} & \boldsymbol{F}_{24} \\ \boldsymbol{F}_{31} & \boldsymbol{F}_{32} & \boldsymbol{F}_{33} & \boldsymbol{F}_{34} \\ \boldsymbol{F}_{41} & \boldsymbol{F}_{42} & \boldsymbol{F}_{43} & \boldsymbol{F}_{44} \end{pmatrix} \quad (2)$$

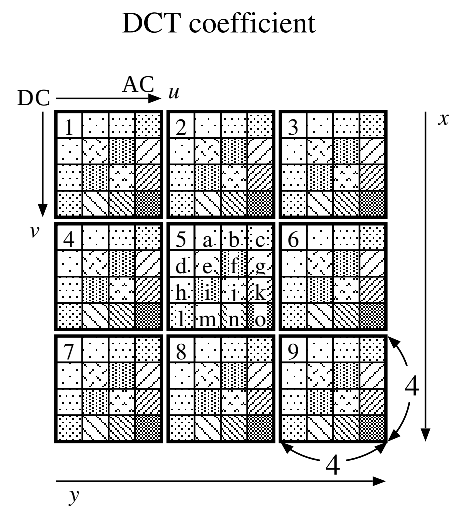
where $\boldsymbol{X}_0$ is input 8 × 8 DCT coefficients, $\boldsymbol{A}$ is 8 × 8 matrix:

$$\boldsymbol{A} = \begin{pmatrix} 0.50 & 0.64 & 0.46 & 0.22 & 0 & -0.15 & -0.19 & -0.13 \\ 0 & 0.05 & 0.19 & 0.36 & 0.50 & 0.54 & 0.46 & 0.27 \\ 0.50 & 0.27 & -0.46 & -0.54 & 0 & 0.36 & 0.19 & -0.05 \\ 0 & 0.13 & 0.19 & -0.15 & -0.50 & -0.22 & 0.46 & 0.64 \\ 0.50 & -0.27 & -0.46 & 0.54 & 0 & -0.36 & 0.19 & 0.05 \\ 0 & 0.13 & -0.19 & -0.15 & 0.50 & -0.22 & -0.46 & 0.64 \\ 0.50 & -0.64 & 0.46 & -0.22 & 0 & 0.15 & -0.19 & 0.13 \\ 0 & 0.05 & -0.19 & 0.36 & -0.50 & 0.54 & -0.46 & 0.27 \end{pmatrix} \quad (3)$$

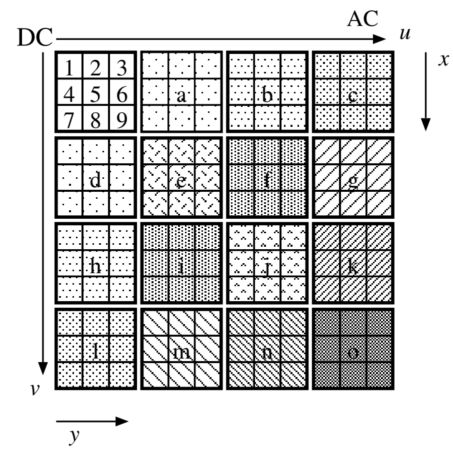
and $\boldsymbol{F}_{\mathrm{ij}}$ is 2 × 2 DCT coefficients:

$$\boldsymbol{F}_{\mathrm{ij}} = \begin{pmatrix} f_{11}^{\mathrm{i,j}} & f_{12}^{\mathrm{i,j}} \\ f_{21}^{\mathrm{i,j}} & f_{22}^{\mathrm{i,j}} \end{pmatrix}, \qquad \mathrm{i,j} = 1, \cdots, 4. \quad (4)$$

These 2 × 2 DCT coefficients are mapped as shown in section II-B.1, so it is possible to consider these coefficients as sub-band coefficients.



(a) DCT expression



(b) Sub-band expression

Fig. 1. Relationship between block expression and sub-band expression

3) Masking Process: An LL band as shown in Fig. 2, which is a low frequency component both horizontally and vertically, is filtered to reduce the discontinuity so that blocky noise can be decreased. In this process, the edge of the images is blurry because of filtering, so masking that corresponds to the high frequency components is done. This masking involves the following process: First, the thresholds ($\tau_{\mathrm{LH,HL,HH}}$) are set. If the value of the high frequency components (LH, which has a low frequency horizontal component and a high frequency vertical component; HL, which has a high frequency horizontal component and a low frequency vertical component; and HH, which has both a high frequency horizontal and vertical component (see Fig. 2)) are higher than that of the threshold, the pixels are not filtered because these pixels are the edge component. This results in keeping the edge so that a clear image is maintained. These thresholds will be chosen based on experimental results.

4) Inverse Wavelet Transform: The DCT coefficients expressed as a sub-band signal are processed for 1-level inverse

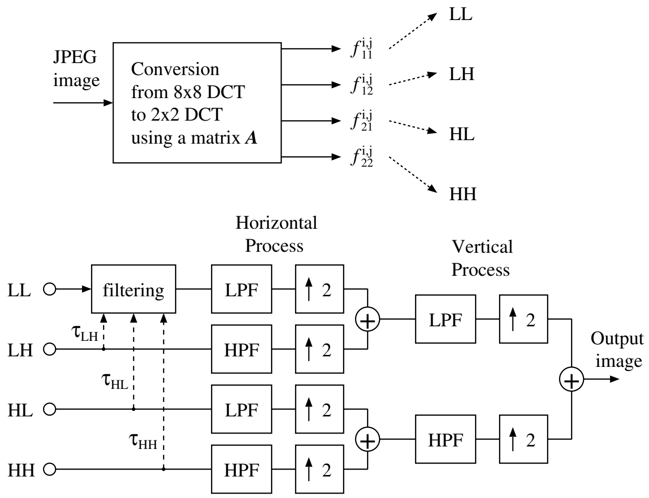


Fig. 2. Proposed decoding flow

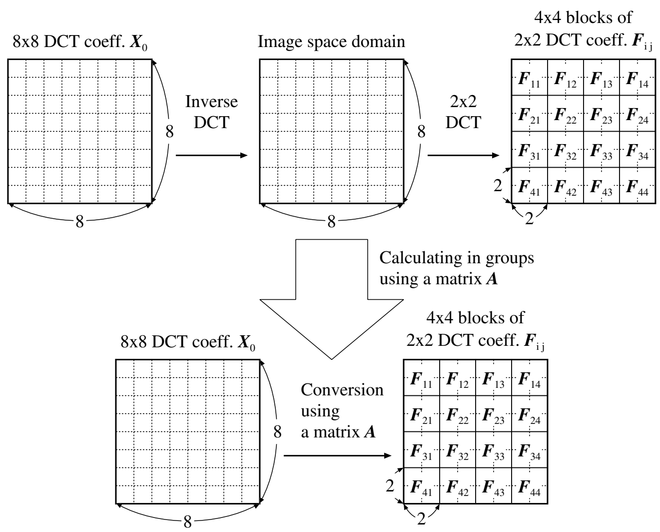


Fig. 3. Transformation from $\boldsymbol{X}_0$ to $\boldsymbol{X}'_0$

wavelet transform, and by adjusting the brightness, the reconstructed images are obtained. In this process, there are a number of filters for inverse wavelet transform. In this study, haar filters are used for reversible transform, so that the reconstructed images are the same as those of inverse DCT. 5/3 filters, which are used in JPEG2000, are used to remove more blocky noise.

5) Processing between frames: A typical MPEG sequence consists of three separate parts: a series of intraframes (called I-frames), which are image frames coded individually without any temporal prediction; a series of forward predicted frames (called P-frames), interspersed between the intraframes; and bidirectionally predicted frames (called B-frames) interspersed between the forward predicted frames and the intraframes. The bidirectionally predicted frames can be considered to be motion-compensated interpolation between the predicted and the intraframes, with the quantizer coefficients being different in each type of frame. Therefore, doing the masking process independently in each frame, the number of correlations between frames diminishes, and then flicker appears in the reconstructed moving pictures. To solve this problem, the masking area and value for P- and B-frames following an I-frame are the same as those for an I-frame in this study.

III. Experimental Results

To check the usefulness of the proposed method, we experimented using C-language. The Peak Signal to Noise Ratio (PSNR) was calculated by using the original MPEG image and the reconstructed image, and the GBIM [10], which is a measurement of the quantity of blocky noise, was calculated.

We experimented with some moving pictures, and in this paper, the experimental images encoded by the MPEG-II are the Foreman, which is 352 $\times$ 288 pixels and 300 frames, the Hall monitor, which is 352 $\times$ 288 pixels and 125 frames, the Football, which is 352 $\times$ 240 pixels and 300 frames, and the Stefan, which is 352 $\times$ 240 pixels and 300 frames as shown in Fig. 4 $\sim$ 7 (a). Where the first image is 15th frame at 403 kbits/s, and the second one is 45th frame at 403 kbits/s, and the third one is 15th frame at 604 kbits/s, and the last one is 15th frame at 623 kbits/s.

A. Subjective Evaluation

The reconstructed images are shown in Fig. 4 $\sim$ 7. The images reconstructed with the DEF method and the proposed method are shown in (b) and (c) of Fig. 4 $\sim$ 7, respectively. In the DEF method in these figures, almost all blocky noise was decreased, however, blocky noise was seen in several parts of the images. In the proposed method, blocky noise was decreased and the reconstructed images were improved. However, the edge area of the reconstructed images was notched. This is because the high frequency signals were filtered with 5/3 filters, and this reduced the PSNR for the proposed method.

B. Objective Evaluation

The experimental results are shown in Fig. 8. The red, blue and green lines are the results of the proposed method, the DEF method and MPEG decoding, respectively. In Fig. 8, the PSNR of the proposed method is smaller than those of the DEF method and MPEG decoding at all the bit rates. It is noted that the PSNR is not an absolute evaluation but a relative one of difference between the original image and the reconstructed image, therefore this is not a serious problem.

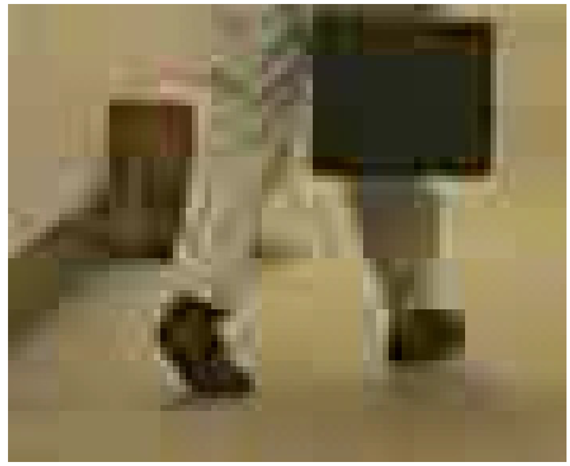
The GBIM of the proposed method is smaller than those of the other method at all bit rates, so the quantity of blocky noise in the proposed method is smaller than those in the other method in the objective evaluation. Also, as the bit rate is set lower and lower, the GBIM of the MPEG decoding and the DEF method increases. Therefore, the quantity of blocky noise in those methods increases. However, the GBIM of the proposed method does not change at all bit rates, so the quantity of blocky noise in this method is invariable.

(a) Original image
(Avg.PSNR : 27.13[dB], Avg.GBIM : 3.02)

(b) DEF method
(Avg.PSNR : 27.19[dB], Avg.GBIM : 2.25)

(c) Proposed method
(Avg.PSNR : 26.99[dB], Avg.GBIM : 1.89)

Fig. 4. Part of an original MPEG-II image and reconstructed images (Foreman, 15th frame, 403 [kbits/s])



(a) Original image
(Avg.PSNR : 28.76[dB], Avg.GBIM : 2.96)

(b) DEF method
(Avg.PSNR : 28.87[dB], Avg.GBIM : 2.56)

(c) Proposed method
(Avg.PSNR : 28.29[dB], Avg.GBIM : 2.09)

Fig. 5. Part of an original MPEG-II image and reconstructed images (Hall monitor, 45th frame, 403 [kbits/s])

(a) Original image
(Avg.PSNR : 21.58[dB], Avg.GBIM : 3.08)

(b) DEF method
(Avg.PSNR : 21.65[dB], Avg.GBIM : 2.54)

(c) Proposed method
(Avg.PSNR : 21.39[dB], Avg.GBIM : 2.05)

Fig. 6. Part of an original MPEG-II image and reconstructed images (Football, 15th frame, 604 [kbits/s])

(a) Original image
(Avg.PSNR : 22.40[dB], Avg.GBIM : 2.67)

(b) DEF method
(Avg.PSNR : 22.43[dB], Avg.GBIM : 2.37)

(c) Proposed method
(Avg.PSNR : 22.11[dB], Avg.GBIM : 1.99)

Fig. 7. Part of an original MPEG-II image and reconstructed images (Stefan, 15th frame, 623 [kbits/s])

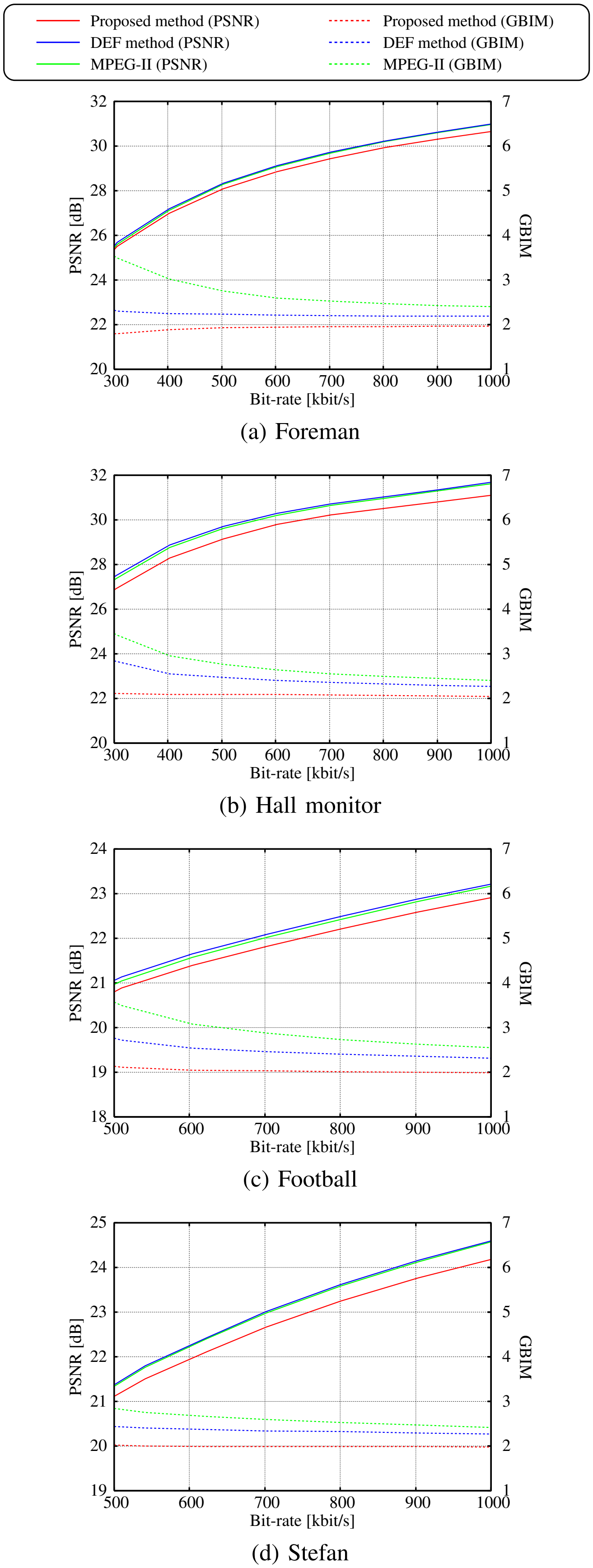


Fig. 8. Relationship between bit rate v.s. Avg. PSNR and Avg. GBIM

IV. Conclusion

In this study, we proposed a decoding method of the DCT coefficients using inverse wavelet transform as a way to reduce blocky noise for moving pictures, and we showed its behavior using C language simulation to check the usefulness of the proposed method.

By considering the frequency components, it was possible to save the edge deletion of filtering so that fine images were obtained. As a result, the examples demonstrated that a better quality of reconstructed images was obtained.

For further research, we are going to correct the notched edge area as a way to improve the PSNR and the reconstructed images for the proposed method. And we are going to study the algorithm to determine the masking thresholds. And we are going to improve the reconstructed image quality by interpolating the non-existent high frequency components which are made using the properties of the wavelet transform coefficients.

References

[1] Wallace, G.K., "The JPEG still picture compression standard", IEEE Trans. on Consumer Electronics, Vol. 38, Issue 1, pp. 18–34, Feb. 1992.
[2] ISO/IEC 11172-2, "Information technology – Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s – Part 2: Video", International Standard, 1993.
[3] A.S.Lewis and G.Knowles, "Image Compression Using the 2-D Wavelet Transform", IEEE Trans. on Image Processing, 1(2), pp. 244–250, 1992.
[4] ISO/IEC 15444–1, "Information technology – JPEG 2000 image coding system – Part 1: Core coding system", International Standard, Dec. 2000.
[5] Y. Yang, N. P. Galatsanos and A. K. Katsaggelos, "Regularized reconstruction to reduce blocking artifacts of block discrete cosine transform compressed images", IEEE Trans. on Circuits and Systems for Video Technology, vol. 3, no. 6, pp. 421–432, Dec. 1993.
[6] Y. Yang, N. P. Galatsanos and A. K. Katsaggelos, "Projection-based spatially adaptive reconstruction of block-transform compressed image", IEEE Trans. on Image Processing, vol. 4, pp. 896–908, Jul. 1995.
[7] H. S. Malvar and D. H. Staelin, "The LOT: Transform coding without blocking eects", IEEE Trans. on Acoustics, Speech and Signal Processing, vol. 37, no. 4, pp. 553–559, Apr. 1989.
[8] Chengjie Tu and Trac D. Tran, "Context based entropy coding of block transform coefficients for image compression", IEEE Trans. on Image Processing, vol. 11, no. 11, pp. 1271 1283, Nov. 2002.
[9] Draft ITU-T Recommendation H.263 Version 2, "Video coding for low bit rate communication", 1998.
[10] H.R.Wu, M. Yuen, "A generalized block-edge impairment metric for video coding", IEEE Signal Processing Letters, Vol.4, No.11, pp.317–320, Nov. 1997.