# 3D CUBE VIDEO CODING USING PHASE-BASED MOTION ESTIMATION AND EZW-IP CODER

*Kiyohiro Uemura*, Tomoyoshi Oguri*, Masaaki Ikehara*, and Truong Nguyen***

*Faculty of Science and Technology, Keio University
3-14-1 Hiyosi Kouhoku-ku Yokohama,Kanagawa
223-8522, Japan
Phone: +81-45-563-1141, Fax: +81-45-563-7453
Email: kuemura@tkhm.elec.keio.ac.jp
ikehara@tkhm.elec.keio.ac.jp

**Department of Electrical and Computer Engineering,
University of California San Diego
9500 Gilman Drive La Jolla, California 92093-0407, USA
Phone: 858-822-5554, Fax: 858-822-3425
Email: nguyent@ece.ucsd.edu

**ABSTRACT**

In this paper, we propose a new three dimensional video coding scheme based on a Phase-based motion estimation, and new three dimensional EZW-IP coder. In the conventional video coding methods like MPEG/H.26x, hybrid coding scheme using motion estimation and DCT transform is adapted. But actually, these methods use only the temporal correlation of several frames. We pay attention here, and aimed to use much more temporal correlation to reduce the redundancy. The motions between successive frames are compensated effectively by phase-based motion compensation(MC) method. Moreover, it is not necessary to prepare the reference image for each input image frame by frame in this proposal method. 8 predicted-error frames are collected, and one CUBE is generated. And it is encoded by three-dimensional EZW-IP method. Our coding scheme provides some desirable features including exact bit-rate control, quality and resolution scalability, progressive transmission, and good compression performance.

## 1. INTRODUCTION

Transform coding has played an important role in various areas of signal processing and communication. It has been widely applied to data compression.

The conventional hybrid video coding techniques like the MPEG and H.26x comprise motion estimation (ME) to reduce the temporal redundancy. Although ME is done in both directions for the prediction (eq. Using B-pictures), they are not enough to reduce the temporal correlation of video information. Recently, 3-D video coding system has much attention due to its efficieny[2]-[4]. However, these method may be complex and the improvement may not be high.

In this paper, we propose a new 3D cube video coding using phase-based ME and 3D EZW-IP coder. At first, the first input frame which is I-picture is wavelet-transformed and encoded by WDR[1]. With this reference frame, several frames are motion compensated by phase-based ME. Collecting eight predicted-error frames, CUBE is constructed. Finally the CUBEs are transformed by 3D-DCT and the transformed CUBEs are encoded by 3D EZW-IP which can encode simultaneously eight or more predicted error frames. In the proposed method, we use only one reference frame per one CUBE so that we can achieve the simple algorithm, while the conventional ME needs a reference image frame by frame, which have to be decoded before prediction. With this method, we can derive the high performance video coder compared to the conventional method.

## 2. PROPOSED ALGORITHM

At first, we show the block diagram of our proposed video coding method in Fig.1.

In order to perform estimation in the transform domain, we adapt the FFT for the transformation. The first input image is I-picture, and it is intra coded by WDR. Fundamentally, the following frames are set to P-pictures. The decoded I-picture is set to the next CUBE's reference frame at first. Each P-pictures are motion estimated and compensated by our phase correlation ME/MC method. This method is kind of block base function. Based on each block divided within the frame, motion compensation is performed by operating a phase.

The point that this proposed method uses a time correlation enough might be understood from this algorithm.

Error(Difference) signal for 8 frames and motion vectors are detected using only one reference frame. Here, 3-dimensional DCT that is one of the important points of this proposed method is applied to difference signal. And it is coded by 3D-CUBE EZW-IP and the next reference frame is generated simply from the last frame of the CUBE.

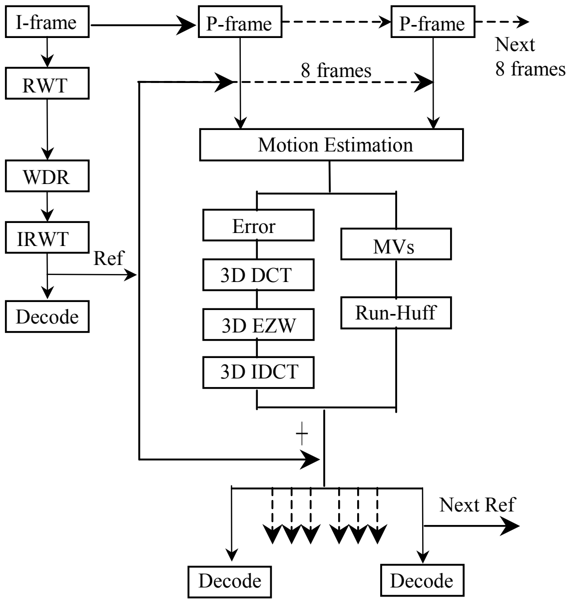



Fig.1 The block diagram of proposed algorithm

## 3. WAVELET DIFFERENCE REDUCTION (WDR)

We use the WDR[1] for intra coding. Generally speaking, SPIHT and WDR were known as an effective encoding method for the still picture. It is not hard to see that WDR doesn't have greater computational complexity than SPIHT.

For one thing, WDR does not need to search through quadtrees as SPIHT does. Thus WDR was adopted here.

It is scheduled to adjust an effective further encoding method in the future. However, since the proposed 3D encoding method is very efficient and shows high performance, the intra coding performance actually has only a small influence to the whole system.

## 4. PHASE CORRELATION MOTION ESTIMATION

In this proposed method, we use the motion estimation and compensation. In many conventional video coding systems, block matching (BM) ME is used widely.

BM is a block base ME method. A target block in a given frame searches the best matching block in the previous frame by calculating the sum of absolute difference (SAD), and the motion vector is obtained.

Phase correlation is a frequency domain motion measurement method that makes use of the shift property of the Fourier transform - a shift in the spatial domain is equivalent to a phase shift in the frequency domain. Using the rotation and scale properties of the Fourier transform, if we use log-polar axes we can find the rotation and scale as a shift in the frequency domain invariant to any translation. Phase correlation is based on the evaluation of the phase of the Cross Power Spectrum between two images.

Given two images $I_1$ and $I_2$ that differ only a displacement ($d_x$ , $d_y$) ie. $I_2(x,y) = I_1(x-d_x, y-d_y)$

Using the shift theorem of the Fourier transform, we can write

$$e^{(w_x d_x + w_y d_y)} = \frac{I_1(w_x, w_y) I_2^*(w_x, w_y)}{\left| I_1(w_x, w_y) I_2^*(w_x, w_y) \right|} \qquad (1)$$

The simple expression of whole flow is ;

1. Windowing both images due to repeating nature of the frequency spectrum
2. Calculate the Fourier Transform
3. Filter out the DC component (removes any remaining global illumination) and any high frequency noise; as it is convenient to do so by simply multiplying the spectral components
4. Calculate normalized cross power spectrum (the "Phase Correlation formula" above)
5. Inverse Fourier Transform
6. Find peak on correlation surface (we used two peaks)
7. Interpolate the surface for sub pixel accuracy

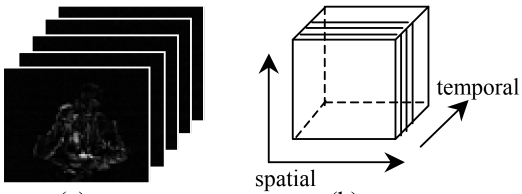
Since motion prediction of the proposed method is performed per block, compared with prediction of MPEG performed per pixel, there is less calculation cost. Furthermore, rather than motion vector per pixel, motion vector per block has obviously few bits required for description.

Moreover, as one of the greatest point of our proposed method, we should have only one reference frame to the several input images (equally one ref. image for one CUBE) so that there is no calculation cost for preparing the reference images frame by frame. In other word, it is not necessary to transform a required reference frame into spatial expression through the filter bank of composition when predicting (This operation is done in the conventional video coding systems), and calculative cost can be cut down.

Actually the reference frame for next CUBE is made when the previous CUBE is decoded. That is, the last frame of previous decoded CUBE is used as the next reference frame.

## 5. THREE-DIMENSIONAL DCT

After the phase-based motion estimation between the reference frame and current frame, the predicted frames will be stored into buffer. Then the Error frames are made. The Error frames are the difference picture between the predicted frames and current frames. (So, a portion with a big motion comes out in many cases as a target outline.) In our proposed method, eight error frames make one CUBE (Fig.2 (a)).




(a) (b)

Fig.2 Image of the CUBE (a):making of error frames CUBE and (b):3-dimensional DCT

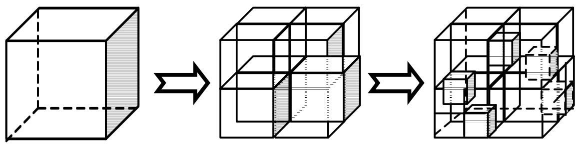


Fig.3 CUBE division

Actually this CUBE which consists of 8 Error frames is encoded by the proposed 3D CUBE coding scheme. The CUBE size is varied easily, but from some experimental results, we set 8 here.

Next, we apply the three-dimensional DCT to the CUBE. This transform has the general two-dimensional (spatial) DCT and an extra temporal transform as shown in Fig.2 (b).

This operation is very important in order to gather the energy in the CUBE. For the simulation, we used the 8× 8 × 8 size 3D DCT. Actually this stage makes the following 3-dimensional EZW-IP coder more effective.

## 6. THREE-DIMENSIONAL EZW-IP CODER

In the conventional video coding algorithm, the same coding approach is applied to both intra and inter coding. However, there are different characteristics between I-picture that is an ordinary picture and P-picture that is an error picture. Thus, it is obviously resembled that the suitable coding methods should be adapted respectively.

The 3D EZW-IP coding method makes the same time coding of several frames (eq. one CUBE) possible. We modified the general 2D EZW-IP coding for 3-dimensional version. In our strategy, this 3D EZW-IP coding is adapted to the CUBE after 3D DCT.

At first, we calculate the maximum coefficient within the target CUBE. Then we divide the CUBE to the bit-planes. In the bit-plane domain, the important signal exist on each plane, the CUBE will be divided into 8 sub-cubes as shown in Fig.3. And this division will be continued to each pixel unit, if the divided CUBE includes 1.

The division form is set to the initial condition for the following layers. And the scale will be down one by one.

If it says more concretely, we use the LIS(List of Insignificant Pixels), LSP(List of Significant Pixels), and LIB(List of Insignificant Bands) in order to collect the each characteristic signal.

### 6.1. Profits of 3D EZW-IP Coder

The following advantages are expectable if subband domain motion estimation(phase-base ME) and 3D EZW-IP are combined.

(a) Calculation of motion presumption is easy.
(b) Exact bit-rate control is possible.
(c) The restored video has the quality and resolution scalability.
(d) Progressive transmission is attained.
(e) There is no block noise.

## 7. EXPERIMENTAL RESULTS

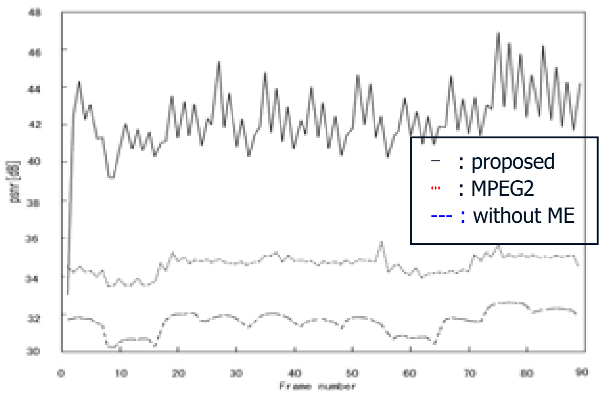
In our simulations, the type of standard video sequence ‘salesman’ (No.1-90) and ‘foreman’ (No.211-300) are used. These are CIF-sized (288x352, 90frames) videos with 30 frames/s. Throughout this paper, our attention is paid at the compaction of whole coding bits. Visual quality of the decoded video sequence (or an image) is the fidelity or the closeness between the decoded frames and the original frames (or image) when perceived by a viewer. In this paper, we use the peak signal to noise ratio (PSNR),which is obtained by

$$PSNR = 10 \log \left( \frac{255^2}{mse} \right) [dB],$$

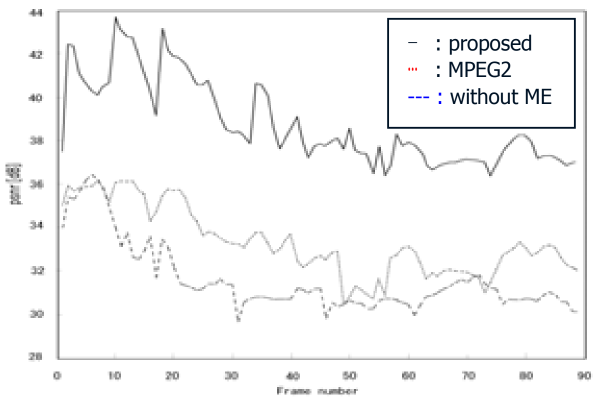
where mse is the mean squared error of original and decrypted. As comparison, the MPEG2-TM5 coding is performed. In the MPEG2 coding, we coded whole 90 frames at 475k(bps) for salesman and 1M(bps) for foreman. Among all frames, only the first frame is I-frame and the others are P-frames. B-frame is not used. The mean PSNR over whole 90 frames are shown in Table I.

Table I : Averaged PSNR over 90 frames [dB]

| | | proposed | MPEG2 | Without ME |
|---|---|---|---|---|
| salesman | 475k (bps) | 42.2331 | 34.5952 | 31.5736 |
| foreman | 1 M(bps) | 38.7882 | 31.5898 | 33.3327 |



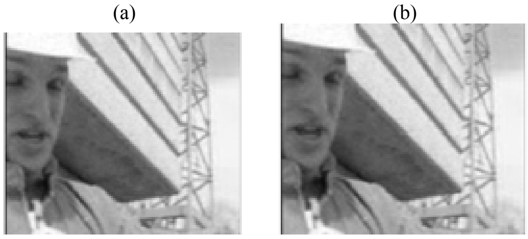

(a)




(b)

Fig.4 Performance comparison between the proposed method, MPEG2 and non ME method. (a)salesman:475k bps (b)foreman 1M bps

(a) (b)



(c) (d)

Fig.5 Visual performance comparison between the 74th decoded 'Foreman' frames
(a)original (b)propose (c)MPEG2 (d)without ME

And the PSNR performance comparison over 90 frames is shown in Fig.4. The data indicate that the proposed method achieves very high performance against MPEG2.

In fact, our proposed method shows the high PSNR performance.

It should be observed that proposed method has an up-and-down sharp change. This depends on the bit control. Because a constant bit was given to each P-picture in the proposed method, it became such a result. Therefore, we will make an effective bit allocation as the next step of this work. As the visual performance, the 74th decoded frames of four methods are shown in Fig.5.

The Motion Vectors were coded by Run length – Huffman coding method. This is a kind of famous lossless coding scheme. The required bits for MVs of each frame were around 1k bits.

## 8. CONCLUSION

We proposed a new three-dimensional video coding method using phase correlation motion estimation and 3D EZW-IP coder. Our experimental results showed the very high performance to MPEG2. In our strategy, firstly the phase-based ME was applied to the incoming signal. Then we made the CUBEs of error frames and the 3D DCT and the 3D EZW-IP coder was adapted. We need only one reference frame for each CUBE. The motion compensation in our proposed phase-based method showed better performance compared with the conventional block-based motion compensation in a spatial domain, so that our 3D EZW-IP coder achieved quite the very good performance. Moreover, since it is a motion estimation scheme of block unit, few bits are required for description of motion vectors. As the infrastructure maintenance of a broadband network progresses on more, it can be said that our proposed video coding method is an effective and useful system.

## 9. REFERENCES

[1] Mainak Biswas and Truong Nguyen, "A Novel Motion Estimation Algorithm Using Phase Plane Correlation for Frame Rate Conversion"

[2] Beong-Jo Kim and William A. Pearlman "An Embedded Wavelet Video Coder Using Three-Dimensional Set Partitioning in Hierarchical Trees (SPIHT)"

[3] Beong-Jo Kim and Zixiang Xiong 'Low Bit-rate Scalable Video Coding with 3-D Set Partitioning in Hierarchical Trees (3-D SPIHT)', IEEE Transactions on Circuits and Systems for Video Technology, Vol.10, No.8, Dec 2000

[4] W. Asano and M. Ikehara 'The Video Coder with Filtebank and Subband Motion Estimation'