社団法人　情報処理学会　研究報告
IPSJ SIG Technical Reports

2004－AVM－47（3）
2004／12／9

Multi-directional Picture を用いた動的光線空間情報圧縮

岡　慎也[†]　圓道　知博[†]　藤井　俊彰[†]　谷本　正幸[†]
[†]名古屋大学大学院　工学研究科電子情報学専攻谷本研究室
〒464-8603 名古屋市千種区不老町
E-mail: [†]oka@tanimoto.nuee.nagoya-u.ac.jp, [†]{yendo, fujii, tanimoto}@nuee.nagoya-u.ac.jp

あらまし　本研究において我々は動的光線空間の圧縮の手法について検討を行う.提案手法は全ての符号化対象画像を多方向のフレーム間予測可能なピクチャとして符号化するものである.提案法はイントラブロックの数を少なくすることによって符号量の削減を行う.また多方向のフレーム間予測によって予測誤差を減らすことより符号化効率の改善をはかるものである.

Dynamic Ray-Space Coding using Multi-directional Picture

Shinya OKA[†], Tomohiro ENDO[†], Toshiaki FUJII[†], Masayuki TANIMOTO[†]
[†]Department of Information Electronics, Graduate School of Engineering, Nagoya University
Furo-cho, Chikusa-ku, Nagoya, 464-8603, Japan
E-mail: [†]oka@tanimoto.nuee.nagoya-u.ac.jp, [†]{yendo, fujii, tanimoto}@nuee.nagoya-u.ac.jp

Abstract This study presents a new coding scheme for dynamic ray-space in which the proposed coder codes dynamic ray-space by using multidirectional inter-image predictions in both temporal and spatial domains. The coder codes pictures with few residual errors by using multi-directional predictions and codes at a low bit-rates by decreasing the number of intra blocks. These two advantages improve coding efficiency.

Keywords Dynamic ray-space, Multi-directional picture, MPEG encoder

1. Introduction

We proposed Free Viewpoint Television (FTV) [1], which allows a user to watch a scene from arbitrary viewpoints. To represent photo-realistic, 3-dimensional space, we used a ray-space method [2]. Synthesizing an arbitrary viewpoint image of a scene requires many cameras, and the necessary image data is comprised of the moving picture data multiplied by the number of cameras. Since the amount of data is enormous, an efficient coding method is needed.

Until now, lots of research on the compression of multi-view images and ray-space has been done. On the other hand, research on the compression of multi-view images that considers temporal domain redundancy is scarce. In this paper, we propose a new compression method that considers not only spatial domains but also temporal domains for multi-view videos.

2. Dynamic Ray-Space

For the representation of photo-realistic, 3-dimensional space, we use a ray-space method [2] that enables FTV to display a scene from arbitrary viewpoints.

Ray-Space methods assume that rays emitted from an object go straight without attenuation. Thus, the rays captured by cameras placed at different positions are regarded as the same ray. As shown in Fig. 1, by using images captured by aligned cameras, the images captured by virtual front or backside cameras can be generated by collecting the rays of images captured by actual cameras because the rays captured by an actual camera and the rays of a virtual camera are the same.

For synthesizing an arbitrary view image, a much larger number of images is needed. When generating a virtual camera view, collecting the rays from the directions of a virtual camera is needed. The number of images for collecting the rays that generate the virtual view is more than one.

In addition, considering changes in the temporal domain, we must treat a vast number of images because each time the images captured by cameras are changing.

3. Dynamic Ray-Space Coding

Dynamic ray-space has an enormous amount of data. Since cameras capture the same objects and scenes in a spatial domain, each image is similar to the image captured by a neighboring camera. In a temporal domain, there is little difference between images captured at current and previous times; therefore the essential information for representing dynamic ray-space is not so large.

As a compression scheme that effectively uses such correlation between images, a hybrid video coder that uses a MPEG encoder (MPEG-1, MPEG-2, and other video coding standard) is well-known. This encoder uses inter-image predictions for removing redundancy between images. MPEG encoders also effectively address the issue of coding dynamic ray-space,.

3.1 MPEG encoders

MPEG encoder codes motion pictures by using intra and inter-image correlation. In Fig. 2, an MPEG encoder codes 3 pictures. First, it codes pictures at t=0 by applying an intra-image coding scheme, e.g., DCT. Next, the MPEG encoder codes picture t=2 by using intra coding and inter-image prediction. By using inter-image predictions, similar parts of already coded pictures (t=0 to picture t=2) are copied, so it doesn't have to send intra coding blocks

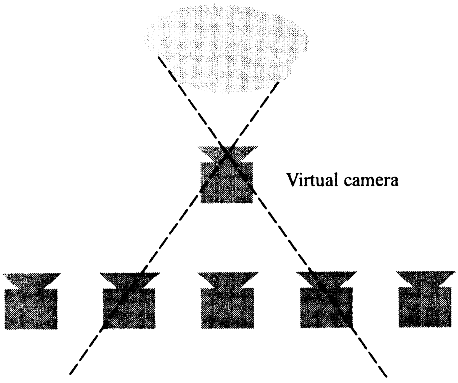

Fig. 1 Synthesizing an arbitrary viewpoint view.

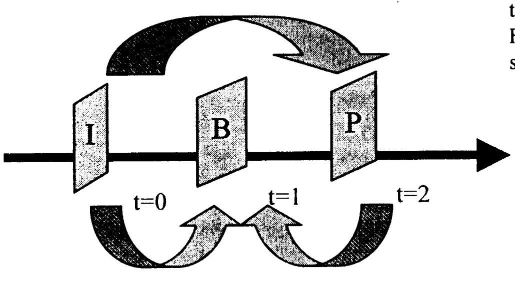

Fig. 2 Picture type of MPEG encoder

of input images; therefore inter-image prediction can significantly reduce bits. Finally the MPEG encoder codes picture t=1 by also using inter-image prediction which uses reference images t=0 and t=2 and intra coding. By using two reference images, the coder can choose and copy the better parts of the image and reconstruct it with fewer residual errors.

We usually call picture t=0 the intra picture (I-picture), picture t=2 the predictive picture (P-picture), and picture t=1 the bi-directional picture (B-picture).

3.2 Coding mode selection

When coding the pictures using a MPEG coder, the user desires to maximize the coder's efficiency by coding at lower bits with few residual errors. Lagrange multipliers [3] are a known solution.

Let us define set C whose elements denote coding modes, and R_i and D_i denote the bit rate and distortion of the coding mode. The cost function of Lagrange multiplier J is defined as (1):

$$J_i = D_i + \lambda R_i (i \in C) \quad (1)$$

By choosing coding modes of minimum cost function J, the MPEG encoder's performance is maximized by blocks.

3.3 Pre-experiment

Because a MPEG encoder can also remove redundancy in temporal and spatial domains of dynamic ray-space, in pre-experiments we applied MPEG encoders to temporal or spatial domains to investigate which domain is more correlated in dynamic ray-space. If one domain is more correlated than others, a MPEG encoder can code dynamic ray-space with less residual error and fewer bits.

In this experiment, we used two types of test sequences captured in our laboratory: "xmas" and "aquarium." "Xmas" was captured in 3 mm camera intervals at a distance of 300 mm. The object was moved roughly in a temporal domain whose resolution was 640 × 480. "Aquarium" was captured with sparse camera intervals, per 3 degrees in a circle with a radius of 350 mm; the movement of objects in the temporal domain was smooth, and the image resolution was 320×240. Figs. 3 and 4 show images of "xmas" and "aquarium."

We applied the MPEG encoder to these two sequences in temporal or spatial domains. We set the coding dynamic ray-space unit at a 15×15 temporal and spatial image matrix. The coding results applied the MPEG encoder (the number of the B-picture is 2) to the "xmas" sequence, and the selected coding modes of each encoder are shown in Figs. 5 and 6; the results of the "aquarium" sequence are shown in Figs. 7 and 8.

Fig. 4 Test sequence: "aquarium"

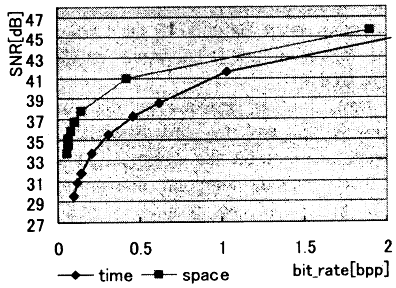

Fig. 5 "Xmas"coding results.

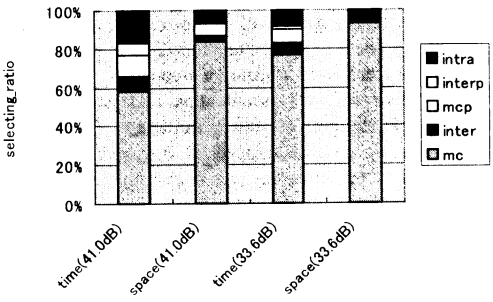

Fig. 6 "Xmas" selection coding mode ratios.

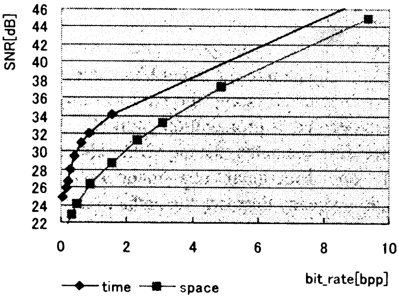

Fig. 7 "Aquarium"coding results.

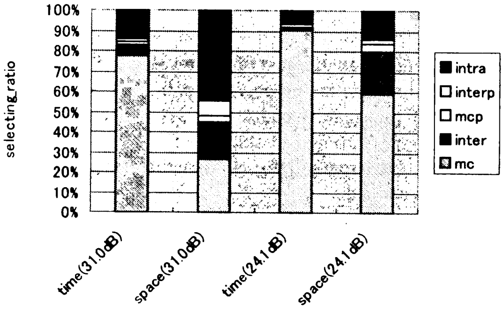

Fig. 8 "Aquarium"selection coding mode ratios.

In Figures 5, 6, 7, and 8, time is the coding result after applying the MPEG encoder in a temporal domain, while space is the result after applying it in a spatial domain. In Figs. 6 and 8, intra denotes the intra coding mode selection ratio of the MPEG encoder, mc means the selection ratio of the motion compensation inter-image prediction mode, mcp is the mc mode that codes the residual error, inter is a bilinear interpolation that uses two motion compensated blocks, and interp is the inter mode with residual error coding.

As shown by Figs. 5 and 7, coding efficiency depends on the direction in which the MPEG encoder is applied. When applying the MPEG encoder to "xmas," the coding efficiency of the space coder is better than the time coder because inter-image prediction in spatial domain works better because the "xmas" sequence is dense in a spatial domain; but the "aquarium" sequence is sparse. On the other hand, when applying the MPEG encoder to "aquarium," the coding efficiency of the time coder is better than the space coder because inter-image prediction in a temporal domain works better because the "aquarium" sequence is highly correlated in a temporal domain; but the "xmas" sequence is hardly correlated. Figs. 6 and 8 show that the MPEG encoder works better when choosing a smaller number of intra, mcp, and interp modes, and more mc and inter modes. Since inter-image prediction in highly correlated domains can predict images with few residual errors; by applying inter-image prediction in highly correlated domains, the number of selection coding residual error modes can be reduced, and the number of the selection mc and inter can be increased. As a result, the coder can reduce the number of bits.

From the above results, to design a more effective encoder for coding dynamic ray-space, we should concentrate on the inter-image prediction direction.

4. Proposed Coder

To overcome the above problem, we propose an encoder that enables all coding pictures to utilize inter-image prediction in both temporal and spatial domains. By utilizing inter-image prediction in both domains, coding pictures can be predicted with few residual errors without depending on the correlation between neighboring images in temporal and spatial domains. The proposed coder has to send fewer intra blocks; therefore it can save bits.

4.1 Multi-directional Picture

In the literature [4], we proposed a new picture type for coding dynamic ray-space named Multidirectional picture (M-picture), shown in Fig. 9, which utilizes inter-image prediction in both temporal and spatial domains. This picture has the following 21 coding modes classified into 5 categories:

1. Intra (1 mode).
2. Inter-image prediction.
(This mode uses a single frame out of four in temporal and spatial domains: forward, backward, rightward, leftward; 4 modes).
3. Inter-image predictions and coding residual error (4 modes).
4. Bilinear interpolation.
(This mode uses and combines two frames out of four in temporal and spatial domain, 6 modes).
5. Bilinear interpolations and coding residual error (6 modes).

The introduction of this picture improved coding efficiency because multidirectional inter-image prediction enables the coder to generate pictures with few residuals. In these results, the proposed coder codes all pictures as M-pictures. By using M-pictures, the proposed coder works better independently of the application of the inter-image prediction direction and codes pictures with fewer residual errors and bits.

When coding all input images as M-picture, the blocks coded as intra blocks and the direction of the inter-image prediction is important. Criteria should maximize the performance of the proposed coder to choose the best modes for coding efficiency.

4.2 Effective mode selection

Criteria should maximize the coding efficiency of the proposed coder. In many combinations of coding modes, the combination conducted the criteria to enable the coder to code pictures at lower bits with few residual errors. The total of the bits and the distortion of the coded picture should be minimized and conformed by the proposed criteria of their sum coded and conformed by the other criteria.

Lagrange multiplier cost function J is a function of R and D, the minimum sum of R and D which means that the total sum of J is minimized; therefore the coder achieved maximum efficiency.

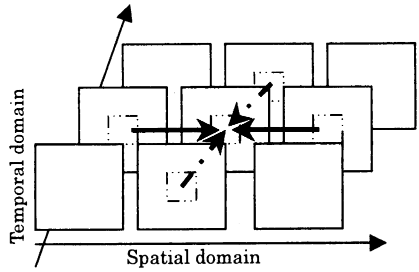

Fig. 9 Multidirectional picture

For realizing a proposed coder that conforms the criteria to minimize cost function J, we created the following algorithm.

The proposed coder focuses on the propagation of pixels from one intra block pixel by inter-image prediction as shown in Fig. 10. Figs. 7 and 8 show that a high performing coder selects fewer intra block coding modes. The larger the pixel area of propagation, the larger the number of generated blocks that use inter-image prediction; then the input images are coded by a small number of intra blocks and a large number of blocks using inter-image prediction. By reducing the number of intra blocks, the coder can code input images at low bits.

The proposed coder considers the inter-image prediction direction. If it is good, the generated block predicted by the inter-image prediction contains few residual errors, and so the distortion of the predicted image becomes smaller.

By considering both the area of pixel propagation by inter-image prediction and the direction of the inter-image prediction, the proposed coder minimizes the cost function of J and reduces the bits and residual error.

4.3 Coding Algorithm of Proposed Coder

When actually coding all the pictures, the algorithm requires the below steps.

Step 1 is a preparation process to decide the coding modes, which decide the paths from a macroblock to the macroblocks of neighbor input images by inter-image prediction.

To decide the pixel propagation path for each macroblock, block matching is applied to the neighbor input images in temporal and spatial domains and then motion vectors of each direction is recoded and the macroblock addresses are recoded to referred macroblocks, as shown in Fig. 11.

For all macroblocks step 1 is repeatedly applied.

Step 2 calculates the evaluation function value based on Lagrange multiplier cost function J and the area of the propagated pixels. The macroblock with minimum evaluation value is coded as an intra macroblock, and the direction of the paths to other macroblocks is decided as the direction of the inter-image prediction.

First, the algorithm starts one macroblock coded as an intra macroblock, and then by using macroblock addresses, which are recoded in step 1, the propagation path is decided. When deciding the path, the algorithm calculates J per unit area for each path as Fig. 12, and then the path of minimum J per area is decided as the propagation path.

Next, the algorithm considers the paths from the intra macroblock and the propagated block from the intra macroblock. The path to the macroblock address recoded to these two macroblocks and the path of minimum J per unit area are decided as propagation paths (Fig. 13).

When deciding the paths, the algorithm allows blocks to select only one directional inter-image prediction. If a block's path is already decided, the paths to the block from different directions are ignored.

With deciding the path to other macroblocks, the algorithm calculates J per unit area using (2).

$$J_{MB} = \sum_{k=0}^{n-1} J_k \Big/ \sum_{k=0}^{n-1} A_k \quad (2).$$

In (2), J_k and A_k are the cost function value and the area of blocks included in the propagation path, and n means the number of propagation paths. This J_{MB} value per unit is recoded to the starting macroblock.

Until no paths remain, for all macroblocks, step 2 is repeatedly applied.

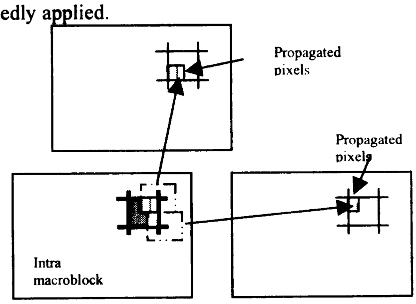

Fig. 10 Propagation of intra pixels.

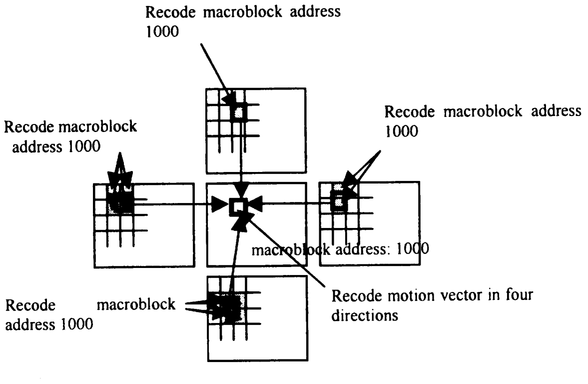

Fig. 11 Preparation of deciding paths of propagation pixels.

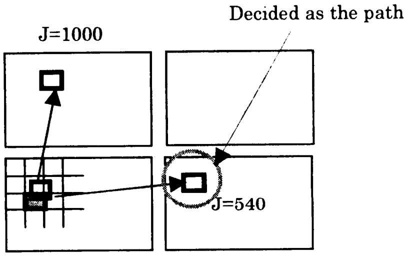

Fig. 12 First step of propagation path.

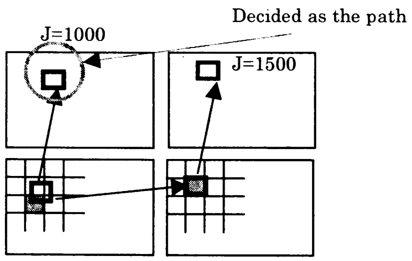

Fig. 13: Second step of deciding the propagation path process.

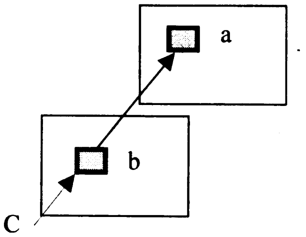

Fig.15: The condition preventing making inter-image prediction loop

Step 3:

Here the algorithm considers many combinations of step 2. The proposed coder repeats step 2 until it decides all macroblocks to be coded as intra or inter-image macroblocks.

After the macroblock coded as intra or inter-image predicted block is decided, the macroblock mode cannot be changed. These macroblocks are constrained by already decided modes; therefore, the total cost function J depends on the order of deciding which macroblock is coded as intra blocks. For calculating the minimum sum of the cost function, we must consider many combinations of step 2.

In addition, this algorithm must decide if the direction of the inter-image prediction can reconstruct the image. If the decided direction of the inter-image prediction is decided, then decoder cannot reconstruct the image. As in Fig. 14, an inter-image prediction in the wrong direction makes a loop. In Fig. 15, suppose that block a is predicted from block b. Defining set C includes blocks that trace the inter-image prediction of block b, and the condition for preventing the decided direction from making a loop is (3)

$$a \notin C \quad (3)$$

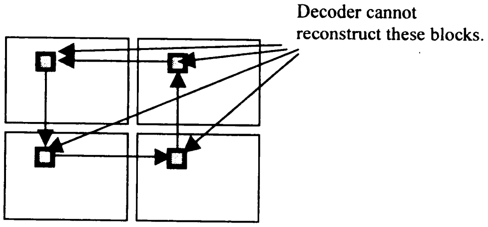

Fig.14: Making loop of inter-image predictions

When deciding the propagation path with only condition (3), the selected direction of the inter-image prediction can reconstruct the image.

Step 4 decides the actual coding mode to all macroblocks. By step 3, if a macroblock is chosen to be coded as an intra macroblock, it is just coded as an intra mode. The macroblock, chosen to be coded as an inter-image predicted macroblock, is considered by coding modes, motion compensation inter-image predictions, and coding the residuals. Bilinear interpolation combines motion compensated blocks predicted in other directions from already decided ones and codes residual bilinear interpolation. When considering bilinear interpolation modes, coders should pay attention to condition (3). Then the minimum cost function mode is selected as an actual coding mode.

Step 4 is applied to all macroblocks.

4.4 Huffman Coding

As mentioned above, our proposed coder has 21 coding modes; macroblock type coding needs more bits. In Fig. 15, the outside images of the coding dynamic ray-space unit cannot use inter-image predictions in all four directions. Therefore, by applying different Huffman codes to the picture at different positions, the coding efficiency of the proposed coder can be slightly improved.

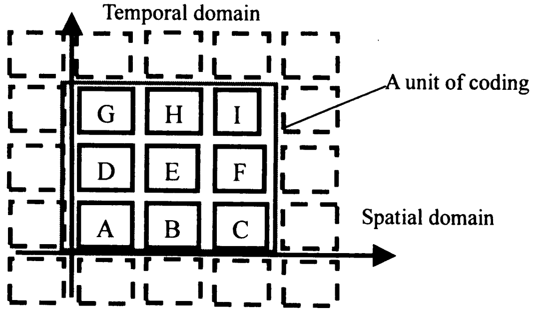

Fig.16 Number of anchor images

We investigated the probability of the inter-image prediction direction results by applying the "xmas" and "aquarium" test sequences and then made different Huffman tables for each picture at different positions.

4.5 Decoding Process of Proposed Coder

When decoding images, because we don't know where intra blocks exist, we must decode all bits stream coded units of coding dynamic ray-space. After decoding intra blocks, motion vectors, and residual information, we can decode the desired images with such information; when the image block is coded as intra mode, the block can be immediately reconstructed. But for reconstructing other coding mode blocks, the decoder should trace the motion vector direction.

5. Experiment

To investigate the effectiveness of the proposed coder, we applied it to the "xmas" and "aquarium" test sequences.

For comparison, we used the following 3 types of coder and proposed coders.

Time: The encoder applies the MPEG encoder in temporal domain on each camera. First we applied the MPEG encoder in spatial domains to first time images of sequences, and then this encoder used the already coded frames as reference frames for inter-image prediction. Then the encoder encoded the other images of sequence applying the MPEG encoder in temporal domains.

Space: We applied the MPEG encoder images captured by one camera in a temporal domain, and then this encoder uses the already coded frames as reference frames for inter-frame prediction.

Time and space: First encoder coded top, bottom, far left and far right images of the coding unit of dynamic ray-space using MPEG encoder, and then by using these already coded images, the rest of pictures are coded as M-pictures as shown in Fig. 17.

Proposed: This coder is a proposed coder that codes all pictures as M-pictures using the algorithm mentioned in section 4.3.

The results applying these 4 coders to "xmas" and "aquarium" are shown in Figs. 18 and 19; the proposed coder works better at all bit rates. The reason of improved coding efficiency of proposed coder is that many directional inter-image predictions can predicts high quality image with few residual, and bit stream includes fewer intra blocks, thus the amount of bits are reduced.

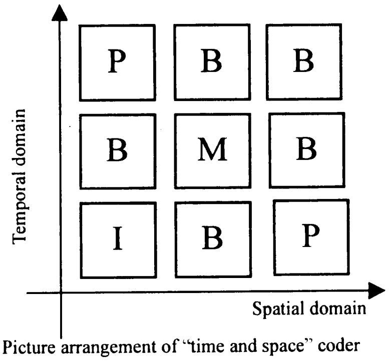

Fig17: Picture arrangement of "time and space" coder

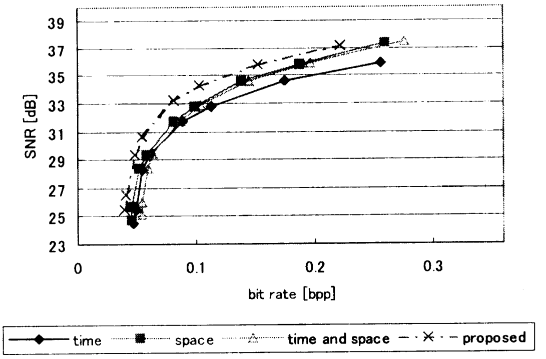

Fig.18:"Xmas" coding results

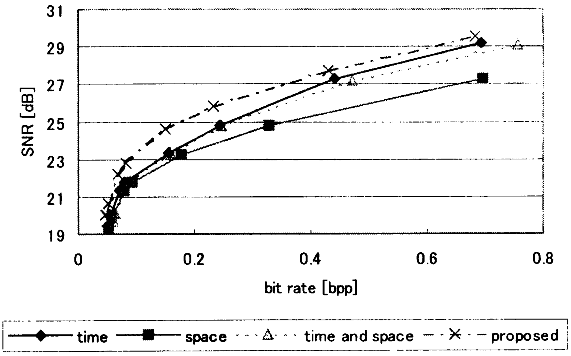

Fig.19:"Aquarium" coding results

6. Conclusions

We studied coders for dynamic ray-space. The results of pre-experiments showed that the direction of applying inter-image prediction significantly effected on the performance of the coder. To overcome this problem, we proposed a coder, which allowed all coding pictures to use many directional inter-image predictions. The proposed coder has two advantages: macroblocks can be predicted with few residual errors by using multidirectional inter-image prediction; therefore, the distortion of the reconstructed image is smaller. And the number of intra blocks is reduced by allowing all coding pictures to use inter-image predictions; therefore, the bits are smaller.

To realize the proposed coder we used Lagrange multiplier cost function per unit area of pixels as an evaluation function value. To evaluate the coding efficiency of the proposed coder, we used 4 types of coders. The experimental results showed that the proposed coder works better at all bit rates, especially at higher bit rate; the proposed coder works better than other coders.

As future work, the proposed coder is very time-consuming, we want to create a faster algorithm.

References

[1] T. Fujii and M. Tanimoto, "Free-viewpoint Television based on the Ray-Space representation," SPIE ITCom 2002, pp.175-189, August 2002.

[2] T. Fujii, T. Kimoto, and M. Tanimoto, "A New Flexible Acquisition System of Ray-Space Data for Arbitrary Objects," IEEE Trans on Circuits and Systems for Video Technology, vol.10, no.2, pp.218-224 Mar. 2000.

[3] G. Sullivan and T. Wiegand, "Rate-distortion optimization for video compression," IEEE Signal Processing Mag., pp 74-90 Nov. 1997.

[4] S.Oka, P.N. Bangchang, T.Fujii, and M.Tanimoto, "Dynamic Ray-Space Coding for FTV," (in Japanese) Proc. 3D Image Conference June, 2004, pp 139-142, Tokyo, Japan.