FRACTAL CODING OF A MULTI-VIEW 3-D IMAGE

Takeshi Naemura and Hiroshi Harashima

Department of Electrical Engineering, University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113, Japan
Email : naemura@harashima.t.u-tokyo.ac.jp

ABSTRACT

This paper deals with two important problems for the 3-D image communication; (i) how to synthesize intermediate-viewing-angle images of a Multi-View 3-D Image, and (ii) how to compress the tremendous amount of data required for the stereoscopic visual effect. In order to synthesize a 3-D image having a larger number of views from a 3-D image having fewer views, we first propose a new concept of view interpolation technique; The resolution of Viewpoint-axis of a Multi-View 3-D Image is enhanced for the purpose of increasing the number of viewpoints. Then we present a new data compression method which is scalable to any resolution, i.e. scalable to any number of views. For this purpose, the fractal coding scheme with the feature of resolution independence is examined, and is shown to be applicable to the 3-D image communication.

1. INTRODUCTION

Three-dimensional (3-D) television is being considered as a next-generation medium that can revolutionize information systems [1]. The creation of the 3-D image communication and broadcasting systems will require the development of various technologies. This paper deals with two important problems as follows:

(i) VIEW INTERPOLATION
how to synthesize intermediate-viewing-angle images of a Multi-View 3-D Image.

(ii) DATA COMPRESSION
how to compress the tremendous amount of data required for the stereoscopic visual effect

A Multi-View 3-D Image consists of several 2-D images of an object recorded from slightly different directions (Fig.1). If the number of viewpoints (cameras) is large enough, the observer can see continuously right side of the object when he moves rightward, and left side of it when he moves leftward. As the number increases, however, the amount of data expands and the input system becomes more complex.

The view interpolation technique which synthesizes images observed from intermediate viewing angle between cameras will contribute to reduce the complexity of the input system, and it is also useful for the flexibility of the 3-D image communication.

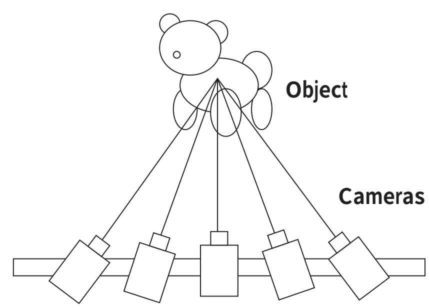

Figure 1: Camera configuration to take a Multi-View 3-D Image

Most of previous works on view interpolation of a Multi-View 3-D Image has been based on the result of structure recovery[2]. This approach, however, cannot be easily applied to arbitrary objects because of the ill-posed inverse problem.

On the other hand, the concept of disparity compensation has been employed for the 3-D image data compression[3]. We consider that the data compression method of a Multi-View 3-D Image should have the scalability to any number of viewpoints, that is, arbitrary number of views should be directly extracted from the compressed data. For this purpose, extension of disparity compensation is also examined[4].

In this paper, a new concept of view interpolation which is free from the difficulty in structure recovery is proposed, and it is extended to a new data compression method which has the scalability to any number of views.

2. RESOLUTION ENHANCEMEHT FOR VIEW INTERPOLATION

A Multi-View 3-D Image has unique structure that reflects the real world as illustrated in Fig.2. The Multi-View 3-D Image on the left side of Fig.2 is sorted by viewpoint locations. When the vertical coordinate [Y] is fixed, we can generate a 2-D image **VX-plane** whose axes are horizontal [X] and viewpoint [V] only.

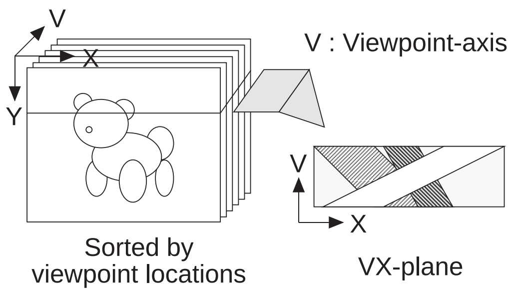

Figure 2: VX-plane of a Multi-View 3-D Image

As the resolution of Viewpoint-axis is equal to the number of viewpoints, the number can be virtually increased by enhancing the resolution of Viewpoint-axis. In this paper, we assume that the view interpolation is achieved by the resolution enhancement in VX-planes without recovering the structure [5].

3. FRACTAL CODING

Application of the fractal theory to iterated transformations has been studied and examined for 2-D image coding [6][7]. Fractal coding schemes exploit the self-similarity within the original image to obtain a fractal approximation. The fractal approximation as a reconstructed image has a unique feature : the scalability to any resolution. To take advantage of this feature, the self-similarity modeling is applied to VX-planes for the purpose of interpolation, in this paper.

3.1. 2-D Self-Similarity Modeling

In modeling process (Fig.3), range block $\mathbf{R}_{ij}$ is a partition of the whole image. For each $\mathbf{R}_{ij}$, a matching domain block $\mathbf{D}_{kl}$ is searched within the same image so that $\mathbf{R}_{ij} \approx T_{ij}\mathbf{D}_{kl}$. Transformation T_{ij} is a composition of a spatial contraction φ, a contrast scaling α_{ij} ($|\alpha_{ij}| < 1$), and an isometry ι_{ij}, of the form:

$$T_{ij}\mathbf{D}_{kl} = \iota_{ij} \ \{\alpha_{ij} \ \varphi \ (\mathbf{D}_{kl} - \mu_{Dkl}) + \mu_{Rij}\}.$$

where μ_{Dkl} and μ_{Rij} are the mean values of brightness level of $\mathbf{D}_{kl}$ and $\mathbf{R}_{ij}$, respectively. In this paper, the root-mean-square distortion is used to determine the similarity between a range and a domain block.

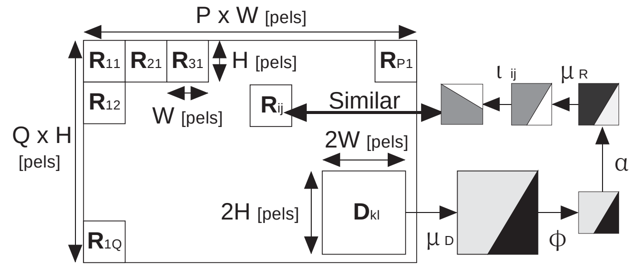

Figure 3: Modeling Process : For each range block $\mathbf{R}_{ij}$, a set of a domain block $\mathbf{D}_{kl}$ and transformation T_{ij} which mimizes the root-mean-square distortion between $\mathbf{R}_{ij}$ and $T_{ij}\mathbf{D}_{kl}$ is searched.

Starting from an arbitrary image, each range block is computed from the corresponding matching domain block (Fig.4). Computing all the range blocks once is called an iteration. When $m = n = 1$ in Fig.4, the reconstructed image will be very similar to the original after several iterations.

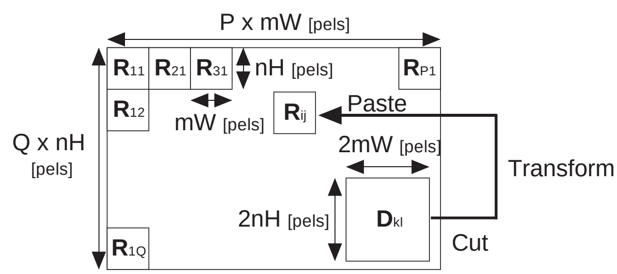

Figure 4: Synthesizing Process : Each range block is computed from the corresponding matching domain block. By changing the size of each block ($n \neq 1$ or $m \neq 1$), the resolution of the image is easily enhanced.

When $m > 1$ or $n > 1$, detailed structure will be synthesized in the higher resolution image. Thus, the resolution of Viewpoint-axis can be enhanced by exploiting the self-similarity of VX-planes.

The range blocks should not overlap each other, for the purpose of data compression. However, this usually causes distortion between neighboring blocks in higher resolution version of the image. Such block distortion in VX-planes would degrade the quality of interpolative (extrapolative) images and be considered intolerable. In order to suppress the block distortion, the range blocks are now allowed to overlap each other. During the synthesizing procedure, the values of overlapped pixels are computed as the average of the outputs of the different adjacent blocks leading to these pixels.

3.2. 3-D Self-Similarity Modeling

The basic idea of 2-D self-similarity modeling is easily applied to 3-D images by extending 2-D range and domain blocks to 3-D cubes or parallelepipeds.

Fig.5 shows how a 17-View 3-D CIF Image ($352 \times 288 \times 17$) is divided into range blocks ($8 \times 8 \times 5$).

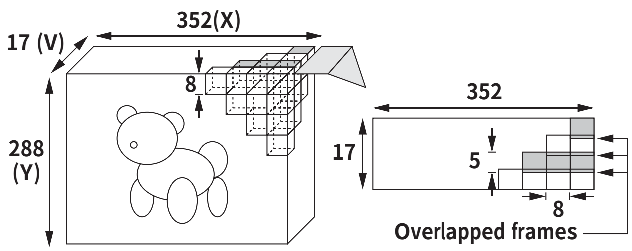

Figure 5: Partitioning of a Multi-View 3-D Image for the 3-D Self-Similarity Modeling : Range blocks are allowed to overlap.

When the size of a range block is $rx \times ry \times rz$, that of a domain block is $2rx \times 2ry \times 2rz$, in this paper. The pixel values of a contracted domain block are the average values of eight neighboring pixels in the domain block.

Contrast Scaling α is obtained as follows [5][7]:

$$\alpha = \frac{C_{rd}(0,0)}{\sigma_d^2}$$

where C_{rd} denotes the covariance of $(\mathbf{R}-\mu_R)$ and $(\mathbf{D}-\mu_D)$, σ_d the variance of $\mathbf{D}$.

There are several kinds of 3-D isometry transformations, for instance, orthogonal reflections about XY, YV or VX-plane, rotations around X, Y or V-axis or combination of these transformations. For the purpose of data compression, just four transformations illustrated in Fig.6 are selected, in this paper.

Fig.7 shows the domain block search path in XY-plane which has the form of a spiral and starts with the domain block in which the target range block is included and centered[7]. Each domain block allocated on the search path has index number. Domain blocks are also allocated along the Viewpoint-axis. In this paper, the matching domain block is searched more exhaustively near the range block, and more sparsely as the index number increases (Fig.7).

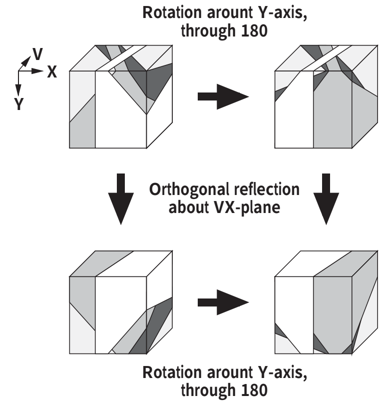

Figure 6: 3-D isometry transformations ι : Just four transformations are selected out of several kinds of 3-D isometries for the purpose of data compression.

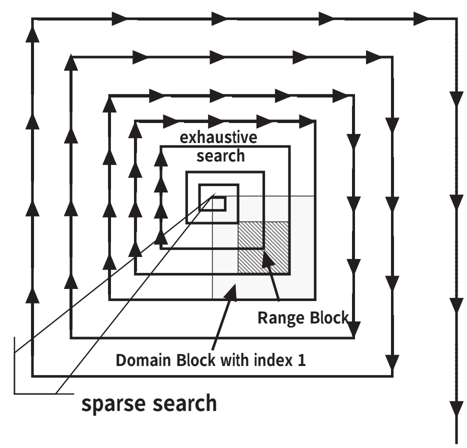

Figure 7: Domain block search path in XY-plane : Each domain block allocated on this path has index number. A matching domain block is searched among these domain blocks.

3.3. 3-D Fractal Coding

The transformation parameters of the self-similarity modeling is useful for view interpolation, but they seem to be still redundant.

In this paper, contrast scaling parameters are quantized to 16 levels ($\pm1.00, \pm0.87, \pm0.73, \pm0.60, \pm0.46, \pm0.33, \pm0.19, \pm0.06$). To take advantage of the spiral search described above, the indices of matching domain blocks on the search paths are encoded, instead of (X, Y, V) coordinates.

Furthermore, adaptive partitioning and classification of range blocks are examined. The range block sizes are $8\times8\times5$, $8\times4\times5$, $8\times4\times3$ and $4\times4\times3$ pels. The decision about splitting the larger range blocks

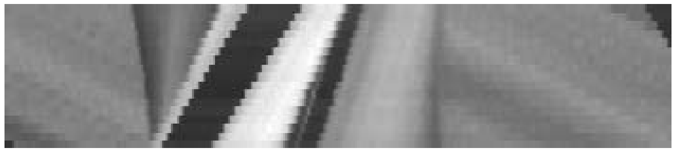

Figure 9: VX-plane of the original 17-view 3-D image

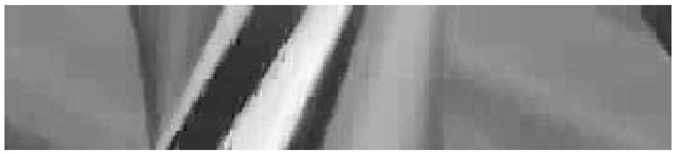

Figure 10: VX-plane of the synthesized 51-view 3-D image (high resolution version of Fig.9)

depends on the mean-square-error(MSE) between the range block and the matching domain block (Fig.8).

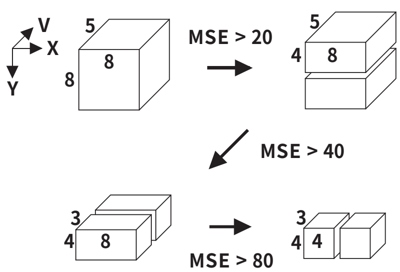

Figure 8: Division of range blocks

All range blocks are classified by their variance σ^2 of brightness into the classes Shade ($\sigma^2 < 8$), Midrange ($8 \leq \sigma^2 < 16(4 \times 4 \times 3), 32(8 \times 4 \times 3), 64(8 \times 4 \times 5), 128(8 \times 8 \times 5)$) and Edge (others). Range blocks with Shade property are approximated by their mean. For range blocks with Midrange property, matching domain blocks are searched among 32 (2^5) blocks without using isometries ι; for other range blocks, 256 (2^8) blocks using isometries. Information in bits needed to represent a block transformation of each class is presented in Table1.

Finally, each parameter (size, class, mean value μ_R, index, contrast scaling α and isometries ι) is entropy coded separetely.

4. EXPERIMENTAL RESULTS

A 17-View Color 3-D CIF Image ($352 \times 288 \times 17$, 16 [bits/pel]) of a toy dog is used for this experiment.

class	size	class	μ_R	index	α	ι
Shade	2	2	8	–	–	–
Midrange	2	2	8	5	4	–
Edge	2	2	8	8	4	2

Table 1: Information in bits for the representation of block transformations of each block type : Information for block size, class and mean value μ_R is needed for any class of range blocks. Index denotes the location of the matching domain block. α denotes contrast scaling factors, and ι isometries.

	parameters [bits]	PSNR [dB]
Y	186402	37.52
U	25871	41.21
V	24497	41.89
Total	236770	

Table 2: Experimental result before entropy coding

The result of 3-D self-similarity modeling before entrotу coding is presented in Table2. There are 5235 Shade blocks, 892 Midrange blocks and 2630 Shade blocks for Y-component. Using Table 1, the amount of information for Y-component is calculated as follows :

$$5235 \times 26 + 892 \times 21 + 2630 \times 12 = 186402 \ [bits]$$

By entropy coding, the amount of data is compressed from 236770 [bits] to 188856 [bits], i.e. this new method has resulted in encoding of the 17-View

Figure 11: Interpolated image of the synthesized 51-View 3-D Image

Color 3-D CIF Image yielding 188856 [bits] at 37.52 [dB] PSNR. This means that a 17-View Color 3-D CIF Image is reconstructed at a bit rate of 0.1095 [bits/pel] with a PSNR of 37.52 [dB].

Fig.9 and 10 shows VX-planes which contains the left eye and nose of the toy dog. Fig.10 is synthesized by enhancing the resolution of Viewpoint-axis in VX-plane of a 17-View 3-D Image (Fig.9). In order to synthesize a 51-View 3-D Image, the resolution of Viewpoint-axis is enhanced to 51 in each VX-plane. Fig.11 shows an interpolated image of the synthesized 51-View 3-D Image. In this case, the apparent bit rate is 0.1095 / 3 = 0.0365 [bits/pel]. And by this method, the complexity of an input system with more than 51 cameras seems to be reduced to that of an input system with 17 cameras.

5. CONCLUSION

For the purpose of data compression and view interpolation of a Multi-View 3-D Image, the fractal coding scheme is examined. This new method has resulted in encoding of a 17-View Color 3-D CIF Image yielding a bit rate of 0.1095 [bits/pel] with an PSNR of 37.52 dB and succeeded in synthesizing 51-View Image directly from the compressed data, apparently yielding a bit rate of 0.0365 [bits/pel]. These results show the potential applicability of the method to the next generation 3-D image communication system.

6. REFERENCES

[1] M.Hatori and I.Yuyama:"Progress Towards Three-Dimensional Television", *TAO First International Symposium on Three Dimensional Image Communication Technologies*, S4-1, 1993.

[2] T.Fujii, J.Hamasaki, and M.Pusch : "Data compression of an autostereoscopic 3D image", *The international symposium on three dimensional image technology and arts*, Seiken symposium, Tokyo, 1992.

[3] M.E.Lukacs:"Predictive Coding of Multi-Viewpoint Image Sets", *ICASSP '86*, pp. 521-524, 1986.

[4] T.Fujii and H. Harashima : "Data Compression and Interpolation of Multi-View Image Set", *IEICE Special Issue on 3D Image Processing*, 1994.

[5] T.Naemura and H.Harashima : "Self-Similarity Modeling for Interpolation and Extrapolation of Multi-Viewpoint Image Sets", *ICASSP '94*, **V**, pp.369 – 372, 1994.

[6] A.E.Jacquin : "Image Coding Based on a Fractal Theory of Iterated Contractive Image Transformations", *IEEE Trans. Image Process.*, **1**, 1, pp. 18-30, 1992.

[7] K.U.Barthel, T.Voyé and P.Noll : "Improved Fractal Image Coding", *PCS'93*, 1.4, 1993.