IEICE TRANS. INF. & SYST., VOL. E82–D, NO. 3 MARCH 1999

INVITED PAPER *Surveys on Image Processing Technologies - Algorithms, Sensors, and Applications -*

Compression and Representation of 3-D Images

Takeshi NAEMURA[†], **Masahide KANEKO**[††], *and* **Hiroshi HARASHIMA**[†], *Members*

SUMMARY This paper surveys the results of various studies on 3-D image coding. Themes are focused on efficient compression and display-independent representation of 3-D images. Most of the works on 3-D image coding have been concentrated on the compression methods tuned for each of the 3-D image formats (stereo pairs, multi-view images, volumetric images, holograms and so on). For the compression of stereo images, several techniques concerned with the concept of disparity compensation have been developed. For the compression of multi-view images, the concepts of disparity compensation and epipolar plane image (EPI) are the efficient ways of exploiting redundancies between multiple views. These techniques, however, heavily depend on the limited camera configurations. In order to consider many other multi-view configurations and other types of 3-D images comprehensively, more general platform for the 3-D image representation is introduced, aiming to outgrow the framework of 3-D "image" communication and to open up a novel field of technology, which should be called the "spatial" communication. Especially, the light ray based method has a wide range of application, including efficient transmission of the physical world, as well as integration of the virtual and physical worlds.

key words: 3-D image coding, stereo images, multi-view images, panoramic images, volumetric images, holograms, display-independent representation, light rays, spatial communication

1. Introduction

The three-dimensional (3-D) imaging technology is a very promising approach expected to outgrow the conventional framework of visual communication. Several kinds of 3-D displays have been developed to enhance the reality of visual experience by providing the natural depth sensation, and they have made rapid progress. The realization of 3-D image communication, however, requires the development of various other support of technologies; 3-D image capture, 3-D image representation, 3-D image transmission, 3-D image handling and so on. The recent status of 3-D TV researches about human factors and systems is reported in [1].

This paper surveys the results of major studies in this field of technology, and focuses on the following technical interests.

- compression of 3-D images, aiming to transmit a huge amount of data efficiently
- display-independent representation of 3-D images, aiming to enhance the flexibility of communication

The display-independent feature is one of the promising approaches to outgrow the conventional framework of "image" communication and to open up a novel field of technology, which should be called the "spatial" communication.

Sections 2 to 4 of this paper survey various coding methods specialized for each kind of 3-D images; stereo pairs, multi-view images, panoramic images, volumetric images and holograms. Then, Sect. 5 reviews another 3-D image coding paradigm of display-independent representation.

2. Stereo Image Communication

The stereo image communication utilizes a stereo pair, comprising two images of the same scene captured from different perspectives. Since stereoscopic systems can easily provide the depth sensation, they have been widely used for various kinds of applications; education, medicine, entertainment, virtual reality and so on.

2.1 Disparity Estimation

One of the most important ideas in the study of 3-D image processing is that of "disparity". The disparity estimation is concerned with the analysis of the geometric differences between two or more views of the same scene. It is well-known, however, that the disparity estimation is an ill-posed problem. Thus, the smoothness constraint on the disparity field is generally introduced; for instance the relaxation labeling [2] and the dynamic programming [3].

As for the stereoscopic video, the relationship between motion and disparity fields can be utilized to realize more effective estimation of both vector fields. Recognizing this fact, the stereo-motion fusion has been investigated [4], [5]. The most simple way of insuring coherent disparity and motion analysis is to predict one vector field by a linear combination of the others. For the sequential stereo pairs, the following equation holds for the same 3-D point projected on the four 2-D images (See Fig.1) :

$$\boldsymbol{d}_{t+1} - \boldsymbol{d}_t = \boldsymbol{m}_l - \boldsymbol{m}_r \tag{1}$$

Manuscript received Jan 16, 1999.

[†]The authors are with the Department of Information and Communication Engineering, School of Engineering, The University of Tokyo, Tokyo, 113-8656 Japan.

[††]The author is with Course in Electronic Engineering, Graduate School of Electro-Communications, The University of Electro-Communications, Chofu-shi, 182-8585 Japan.

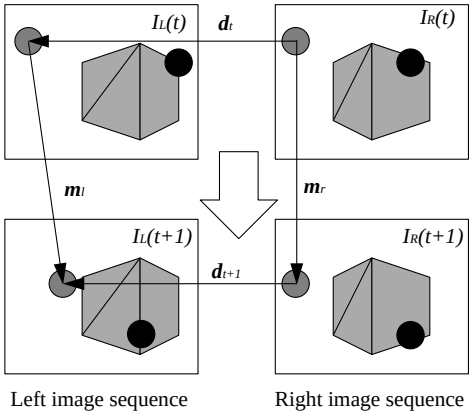


Fig. 1 Coherent disparity and motion analysis.

where $\boldsymbol{d}_{t+1}$ and $\boldsymbol{d}_t$ stand for disparities at time $t+1$ and t, respectively, and $\boldsymbol{m}_l$ and $\boldsymbol{m}_r$ are left and right motion vectors [6]. In other words, the disparity information greatly simplifies the motion estimation [7]. On the other hand, stereo matchings between current frames can be predicted from previous ones and the result of motion estimation [8]. Many approaches have been proposed for the motion-stereo fusion research, e.g. a multi-resolutional dynamic programming [9], a simultaneous motion-disparity estimation and segmentation [10] and a maximum *a posteriori* estimator [11].

The details of several techniques which have been investigated for this purpose are reviewed in [12], [13].

2.2 Basics of Stereo Image Coding

A stereo pair exhibits the spatial correlation between the left and right image. This correlation can be exploited by predictive coding schemes to transmit or store the stereo pair using little more amount of data than one of the pair [14]. Actually, most of the works on stereo image compression have been concentrated on the disparity compensated prediction (DC) which is similar to the motion compensated prediction (MC), widely used in the field of 2-D video coding. The algorithm can be described as follows :

1. The left image is coded independently of the right image, and transmitted to the receiver.
2. The right image is divided into some blocks. For each block, a similar block in the encoded left image is assigned (in other words, the disparity of each block is estimated). Then the disparity data is transmitted to the receiver.
3. The blocks in the right image are predicted by their similar blocks as shown in Fig.2. Finally, the prediction error (the residual image) is transmitted to the receiver.

Since the scheme of DC is similar to that of MC,

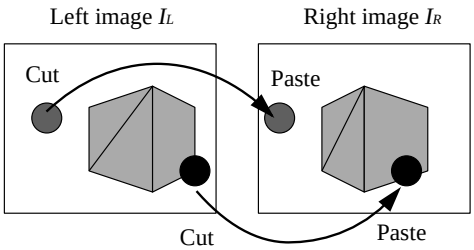


Fig. 2 Disparity compensation.

many products for MC, widely used in 2-D video coders, can be applied for DC. It is important, however, to know the subtle differences between DC and MC. The apparent differences are listed below.

- The disparity $\boldsymbol{d}$ consists of two components : a horizontal component d_x and a vertical component d_y. When the optical axes of stereo cameras are aligned in parallel, d_y is always zero and d_x is always positive, whereas in a 2-D video sequence, the image of an object can undergo vertical and horizontal displacements in both directions, as well as rotation.
- In a stereo pair, every object in the left image is displaced in the right image, while only a few objects change their locations from frame to frame in a typical video sequence.
- In a video sequence, the intensities of corresponding points in sequential frames can be assumed to be equal. This assumption, however, is often false in a stereo pair, because of the individual differences between the left and right cameras.
- In order to effectively compress the amount of data without subjectively intolerable distortions, it is possible to exploit the way, in which the brain processes a stereo pair.

2.3 Improvements in Stereo Image Compression

In the 1980s,

- the basic idea of DC was proposed [14],
- the adaptive combination of DC and MC was evaluated [15],
- the DC approach and the 3-D DCT (Discrete Cosine Transform) approach were compared [16], and
- the statistical characteristics of stereo pairs and residual images were investigated [17].

Since then, several improvements have been carried out for this technical field [18]. They are briefly reviewed in the following.

2.3.1 MPEG Compatible Coding

It is possible to utilize the temporal scalability feature of MPEG-2, when the left sequence is coded on the

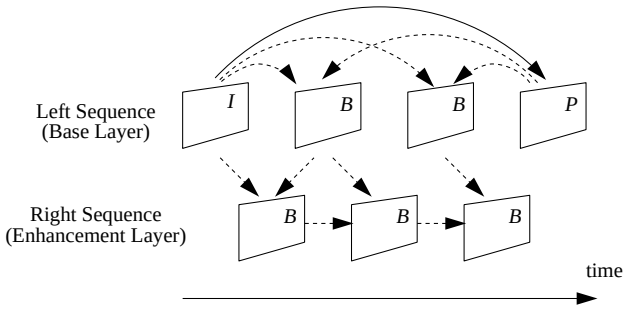


Fig. 3 Temporal scalability coding of stereo video.

lower layer and the right one is transmitted as the enhancement layer. The left bit-stream offers the basic non-stereoscopic signal, and results in the full stereoscopic video when it is combined with the right one. For example, all the enhancement frames can be configured to be B-pictures (Bidirectionally predictive-coded picture) as shown in Fig.3. Thus, the disparity vectors can be transmitted as prediction vectors which are embedded in the basic MC framework of MPEG-2 [19], [20].

In the RACE DISTIMA project, MPEG-2 compatible hardware demonstrators for data compression and transmission of stereoscopic video were developed [21].

2.3.2 Human Visual System

Psychophysical experiments showed that a stereo pair with one high resolution image and another lower resolution image is sufficient to provide the depth sensation. From this point of view, mixed-resolution coding methods for a stereo pair were proposed [22], [23]. The property of the spatio-temporal contrast sensitivity of human visual model [24], and the visual pattern image coding (VPIC), which codes images using a predefined set of visual patterns significant for the human visual system [25], can be also exploited. It was reported, however, that the subjective quality of a stereo image sequence fell approximately midway between the quality of the left and right images [26]. It is still important to investigate the characteristics of the human visual system, in order to achieve efficient compression of 3-D image data.

2.3.3 Suppressing Residuals

Outgrowing the conventional framework of DC, more effective prediction methods were proposed to suppress the residuals.

In order to suppress the differences in the gain of the left and right cameras, the disparity-compensated transform-domain predictive coding (DCTDP) was proposed [22]. Let R be the transform (e.g. DCT coefficients) of the right image subblock and L be the transform of the matching subblock from the encoded left image. R is predicted by a linear predictor of the form

$$\hat{R} = A \otimes L + B \tag{2}$$

where the array multiplication operation $\otimes$ is performed by multiplying the (i, j)th element of the first matrix by (i, j)th element of the second matrix. The matrices A and B are chosen based on the statistics obtained from a training set so that each coefficient is predicted using a minimum variance predictor.

On the other hand, the subspace projection technique (SPT) was proposed to combine DC with the residual coding [27]. The idea is to code left images independently and to code right images using a locally adaptive transform T, which is applied to each $m \times m$ block, $\boldsymbol{b}_r$ of the right image. The transform T is chosen to be a projection operator that projects each $m \times m$ block $\boldsymbol{b}_r$ onto a subspace $\boldsymbol{S}$ of $R^{m \times m}$. This subspace is spanned by a set of N vectors. The first k vectors in the spanning set are block-dependent vectors and can be obtained from left images using disparity compensation. The remaining $N - k$ vectors are fixed, $\boldsymbol{F} = \{\boldsymbol{f}_i\}_1^{N-k}$, and are selected so that they approximate the low frequency components of the block by forming a polynomial approximation to the vector $\boldsymbol{b}_r$. The sequential orthogonal subspace updating method (SOSU) [28] can be classified into this category.

2.3.4 Toward Multi-view Image Communication

There are some other contributions for the stereo image compression;

- a 3-D MC method [29],
- an NTSC-compatible representation [30], and
- a dependent bit allocation framework [31].

On the other hand, the object-based and model-based methods [32] are the recent trends in the stereo image coding. In general, perceptually better image quality can be achieved by the segmentation-based prediction than by the popular block-based prediction. The object-based coding methods play an important role in encoding the segmented region corresponding to an object in a 3-D space. Especially, the model-based methods, which describe a scene in a structural way, are well-suited for the 3-D image coding. It is very important that these methods are useful for interpolating intermediate views between the left and right views (in other words, synthesizing multi-view images). From this point of view, it can be said that the studies on the stereo image coding are getting to involve the multi-view image coding. Further explanations of this new field of technology are presented in the following section.

3. Multi-View Image Communication

A set of multi-view images comprises several images of the same scene acquired from different perspectives.

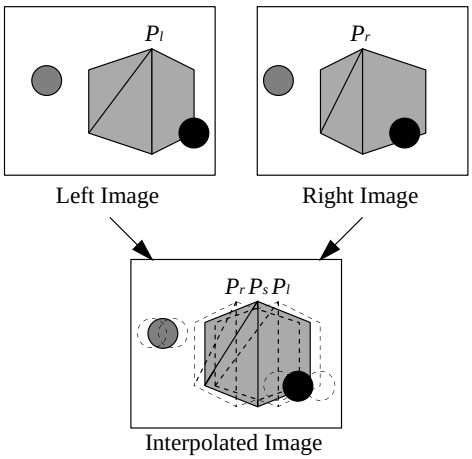


Fig. 4 View interpolation.

While the depth perception from a stereo pair is derived from the binocular disparity, that from multi-view images is obtained from both of the binocular disparity and the motion parallax. Thus, multi-view images can enhance the 3-D visual effect by providing the look-around capability.

As the number of viewpoints (cameras) 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 to enhance the 3-D visual effects (See Fig.4).

In Fig.4, an intermediate point P_s is interpolated according to the equation $\vec{P_r P_s} = s\vec{P_r P_l}$, where $\vec{P_r P_l}$ is the disparity and $s \in [0, \ldots, 1]$. Furthermore, this technique is also useful for the efficient representation of multi-view images, because a large number of views can be compressed into a smaller number of key views by which all other original views can be interpolated.

Consequently, there are two issues that should be addressed in a coding system of multi-view images :

- how to compress the huge amount of data required for the autostereoscopic visual effect, and
- how to interpolate intermediate views.

3.1 From Stereo Image Communication

It is straightforward to apply several techniques, developed for the stereo image coding or the 2-D video coding, to the multi-view image coding. Applications of H.261 coders [33], SPT (subspace projection technique) [34], and MPEG-2 for a 2-D array of images [35] were reported.

As for the view interpolation, several methods have been proposed, some of which are listed below;

- an optical parallel processing system for a stereo pair [36],
- a method for horizontally aligned cameras [37],
- a method for symmetric trinocular cameras on a 2-D plane [38], and
- an interactive method using range data (depth information) [39].

In any case, a non-smooth disparity field will degrade the perceptual quality of interpolated views. The block-based method may suffer from the blocking artifacts. Moreover, for stereo video sequences, the temporal consistency of disparity fields was shown to be applicable to suppress the annoying artifacts of interpolated images [40]. Thus, the estimation of a dense disparity field, in which each pixel has its own disparity information, is essential for the view interpolation. In such a context, the following methods have been investigated;

- dynamic programming methods [41], [42], and
- modified block matching algorithm [43], [44].

The methods of compressing a dense disparity field have been also proposed. They can be evaluated by the quality of interpolated images at the receiver. Some of them are listed below.

- MPEG based methods
  - The MPEG-2 standard with 3-D computer graphics techniques [45].
  - Incomplete 3-D layered representation using video object plane (VOP) of MPEG-4 standard [46].
- 3-D model-based methods
  - Simple convex-surface models [47]–[49].
  - A 3-D layered model [50], [51].
  - Joint 3-D motion/disparity compensation using a wire-frame model [52].
  - Globally rigid and locally deformable 3-D wire-frame models [53]–[55].
- Others
  - The chain map method, which combines several disparity maps into one with very low inherent redundancy [56].
  - Delaunay triangulation of a disparity field [57].
  - Interframe coding of disparity fields [58].
  - A rate-distortion framework for estimating disparity fields [59].
  - A nonlinear interpolator, which generates a full sized map that partially preserves original disparity edge information from block-based disparity data [60].

While this field of technology is relatively new, a hardware system for the real time stereoscopic video conferencing with viewpoint adaptation has been already realized [61].

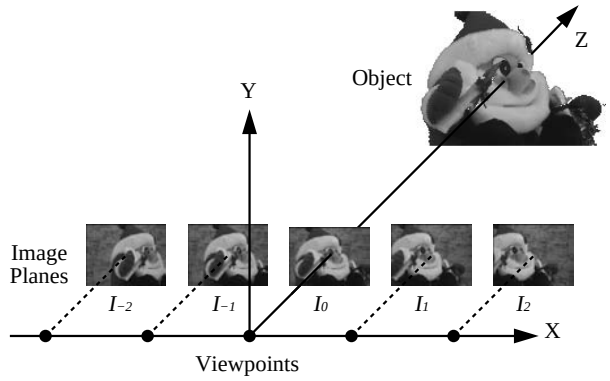


Fig. 5 An example of multi-view camera configurations (Images are presented by Univ. of Tsukuba).

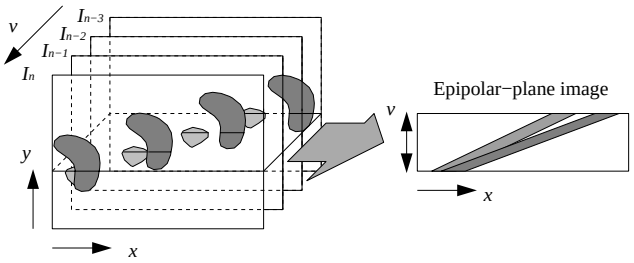


Fig. 6 Synthesis of epipolar-plane images.

3.2 Epipolar-plane Images and Its Application

Multi-view images have the unique structure that reflects appearances of the scene being viewed. A horizontal lineup of equi-distant cameras with parallel optical axis as shown in Fig.5 is considered in this section.

The concept of *epipolar plane image* (EPI) was proposed for building a 3-D representation of a static scene from a sequence of images [62], [63]. Figure 6 illustrates a simple example of synthesizing EPIs from multi-view images, captured by the input system illustrated in Fig.5. Sorting each view I_v by its viewpoint location v, the solid block representation of multi-view images is constructed as shown in the left side of Fig.6. A horizontal slice of this solid of data is an EPI. It is important that an EPI is essentially composed of straight lines whose slope indicates disparity and that each line contains all the information about corresponding points recorded in the multi-view images as shown in the right side of Fig.6.

In order to realize both of the data compression and the view interpolation of multi-view images, a structure recovery method, based on EPI analysis, has been investigated [64]. Especially, a triangular patch model has been examined and utilized due to its compatibility with the DC method, based on affine transformation [47]. In order to cope with occlusions, a method of segmenting the solid of data [65] was applied to synthesize the triangular-patch models of multi-view images [50], [51]. Figure 7 shows the 3-D layered model synthesized from multi-view images [51]. The shape and texture of the model is designed to minimize the differences between input images and their corresponding images reconstructed from the model.



Fig. 7 Synthesized 3-D layered model[51].

On the other hand, in order to avoid the difficulties in the structure recovery, a kind of wave-form based method has been investigated for both of the data compression and the view interpolation [66]. Since the resolution of v-axis in the solid block representation of multi-view images is equal to the number of viewpoints, the number can be virtually increased by enhancing the resolution of v-axis. Thus, view interpolation can be achieved by the resolution enhancement in epipolar-plane images without recovering the structure. In such a context, fractal based compression which is scalable to any resolution was applied for the multi-view compression [66].

4. Other 3-D Images

4.1 Panoramic Images

A panoramic image is an orientation-independent representation of a 3-D scene, which contains all the information needed to look around in 360 degrees. It can be created by stitching together overlapping photographs [67]. A series of pictures can't be stitched together without being warped. This warping process is independent of the scene components, but it depends on the camera configuration. This integration can be seen as a data compression method for several photographs of the same viewpoint.

4.2 Volumetric Images

Compression of volumetric image data is imperative for clinical picture archiving and communication systems, tele-medicine and telepresence networks. A fractal based approach [68] and a wavelet based approach [69] have been reported in this field of technology.

4.3 Holographic Images

For the compression of holographic data, it was noticed that, in the frequency domain, a hologram data can be seen as a pattern related to a perspective view of the scene objects. Consider the case where a hologram is segmented into elements, and they are transformed into the frequency domain. Since the hologram data, after segmentation and transformation, has the characteristics similar to a set of 2-D images, it is possible to apply the 2-D image compression techniques to the hologram data compression [70].

5. The Display-Independent Approach

Most of the works, surveyed in Sects. 2 to 4, are specialized for the limited combinations of input and output technologies; stereo pairs, multi-view images, panoramic images, volumetric images and holographic images. As for the multi-view images, several approaches have been proposed for each different camera configuration (the number of cameras, their positions, their directions, their focal lengths and so on). The next generation communication system, however, should not be restricted to the 3-D image format, but be flexible for various kinds of 3-D images. Not only image data captured by a camera, but also spatial data (such as stereo pairs, multi-view images, integral photographs, holograms, volumetric data, and geometric models of objects), should be utilized seamlessly in the next-generation visual communication system. For this purpose, it is important to establish a method of representing several kinds of spatial data in the same way, which is neutral for any input and output system.

Considering such requirements, the concept of the "Integrated 3-D Visual Communication" illustrated in Fig.8 was proposed [71]. Its key feature is the display-independent intermediate representation of visual data; all the differences between input and output systems will be absorbed by the intermediate representation, because it is neutral for any format of visual data. Thus, we can select our favorite way of observing the transmitted 3-D images, regardless of the input method. This function, as a new coding paradigm, will promote the progress of 3-D image communication.

5.1 Practical Methods for the Neutral Representation

To date, two practical methods for the neutral representation of 3-D visual data have been investigated : the structure-based approach [47], [50], [51] and the ray-based approach.

It is straightforward that the structure-based approach is a practical way of representing scene objects. For instance, VRML (Virtual Reality Modeling Language) is one of the standardized methods of representing geometric models in the virtual world. This approach is well-suited for handling, as well as displaying, objects in the virtual world. On the other hand, since the visual sensation is excited by rays of light incident on the retina, each of these rays can be seen as a primitive element of visual cues, and any kind of visual data can be decomposed into a set of light ray data. Conversely, arbitrary display-dependent visual data can be synthesized from the set of light ray data. The ray-based approach is useful for displaying objects in the physical world as well as in the virtual world.

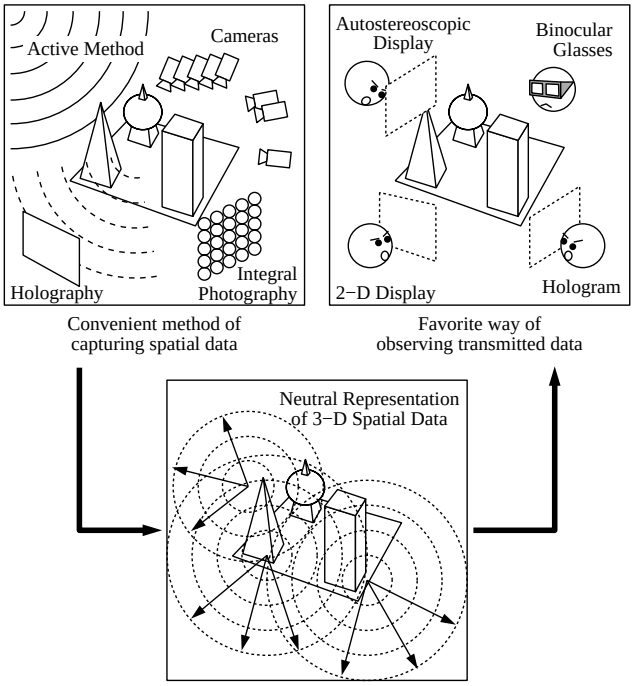


Fig. 8 Integrated 3-D visual communication.

In order to apply the structure-based method to the physical world, we must construct geometric models of every object in the world. It is, however, not easy to measure the accurate shapes, surface textures and reflection models of the physical objects. To solve this problem, a great deal of improvement in the field of structure recovery from images will be required. What is essential for the visual communication, however, is not the structural properties of scene objects, but the light rays emitted from them. The merit of the ray-based approach is that we can avoid the problems of structure recovery. From this point of view, the following sections will clarify the basic concept and applications of the ray-based approach.

5.2 Ray-Based Representation of Visual Data

The ray-based approach was introduced to the field of 3-D image communication including holographic displays in [72] after the proposal of structure-based representation of light rays [73]. It has a close connection with the concept of image-based rendering [74]–[78], which

attracts a great deal of attention in the field of computer graphics. Both of them aim to avoid the difficulties in constructing geometric models of objects in the physical world. Since an image can be regarded as a set of light rays, the ray-based approach involves the image-based one.

A set of light rays passing through any point in a 3-D space is called a pencil, which is distinguished by its position (X, Y, Z). Each of the light rays which constitute a pencil can be identified by its directions (θ, ϕ), representing its longitude and latitude angle of propagation. Therefore, by providing a five-dimensional data space $XYZ\theta\phi$, we can store the light ray data (power or wavelength) $f(X, Y, Z, \theta, \phi)$ at any point in any direction, separately. This data space is called the "Plenoptic Function"[79] or the "Ray-Space"[72].

A panoramic image, which records the incident light arriving from all directions at a point (X_a, Y_a, Z_a), can be represented as follows :

$$f'(\theta, \phi) = f(X, Y, Z, \theta, \phi)|_{(X,Y,Z)=(X_a,Y_a,Z_a)}. \quad (3)$$

This means that a panoramic image corresponds to a 2-D sub-space of the 5-D data space. From this point of view, $f(X, Y, Z, \theta, \phi)$ can be regarded as a set of panoramic images at any location in a 3-D space. In the same way, it can be said that a hologram plane $Z = g(X, Y)$ contains light ray data within a 4-D subspace

$$f'(X, Y, \theta, \phi) = f(X, Y, Z, \theta, \phi)|_{Z=g(X,Y)}. \quad (4)$$

In many cases, a 4-D data space is enough to represent 3-D images, though it is sure that there are some applications requiring a 5-D data space. Further investigations are given in [80]–[82].

5.3 Framework of Ray-Based System

In order to apply the ray-based method to the intermediate representation illustrated in Fig.8, we must consider the following three principal parts; (a) input phase, (b) transmission phase and (c) output phase.

In the input phase (a), input visual data is converted into light rays and mapped onto a 4-D data space. We can also synthesize light rays from geometric models in the virtual world by the ray-tracing technique. For 2-D image input, the camera position (X_a, Y_a, Z_a) and the relationship between image coordinates (x, y) and directions (θ, ϕ) of rays are required to be given or estimated. In many cases, interpolation of light rays will be required in this phase.

In the transmission phase (b), since the 4-D data space still contains a huge amount of data, an efficient compression method is strongly required. Vector quantization [77], fractal-based [66], DCT-based [83], and disparity-based [84] methods have been examined.

In the output phase (c), the light ray data appropriate for the display system at the receiver is extracted or interpolated from the transmitted data space. Some kinds of optical effects, such as fogging and shadowing, can also be added.



Fig. 9 CyberMirage.

5.4 Integration of Virtual Scenes and Real Objects

Finally, a kind of mixed reality applications, in which physical objects represented by light ray data are placed in the virtual world represented by geometric models, is outlined. Figure 9 shows an example of a cyber space called CyberMirage [85], [86]. Components inside a building are presented by polygonal models and dresses are presented by the light ray data. These objects are implemented so that the observer can move viewpoints and operate them in real time on SGI Indigo2 IMPACT or equivalent machines.

Representation of a virtual space using polygonal model data is based on a subset of VRML 1.0. Descriptors of VRML are expanded so that they can merge light ray data with VRML data. Light ray data are objects which can only be observed since they do not have explicit shape models. A scheme is, however, built to move or rotate objects represented by light ray data in the CyberMirage system. CyberMirage having the features mentioned above is useful for representing products realistically in the display of a virtual mall.

6. Conclusions

This paper gave a bird's eye view on the results of various studies on 3-D image coding, which is one of the most critical fields of technology for the purpose of construction of a 3-D image communication system. Themes of this paper were focused on the efficient compression and the display-independent representation of 3-D image data.

Most of the works on 3-D image coding have been concentrated on the compression methods. Especially,

the concept of disparity compensation have been developed for the compression of stereo images. In order to consider many other types of 3-D images comprehensively, however, more general and neutral platform for the representation of 3-D images is required. Thus, the concept of the neutral representation as a viable alternative of 3-D image coding was introduced aiming to outgrow the framework of 3-D "image" communication and to realize the "spatial" communication. Especially, the ray based approach to this concept has a wide range of application, including the efficient transmission of the real space, as well as integration of the virtual and real spaces.

The 3-D image coding technology has a close connection with many other technological fields: computer graphics (CG), virtual reality (VR), and computer vision (CV), for example. More comprehensive studies related to these technologies will contribute to further development and progress of the spatial communication.

References

[1] T. Motoki, H. Isono and I. Yuyama: "Present status of three-dimensional television research", Proc. IEEE, **83**, 7, pp. 1009 – 1021 (1995).

[2] S. T. Barnard and W. B. Thompson: "Disparity analysis of images", IEEE Trans. Pattern Anal. & Machine Intell., **PAMI-2**, 4, pp. 333 – 340 (1980).

[3] Y. Ohta and T. Kanade: "Stereo by intra- and inter- scanline search using dynamic programming", IEEE Trans. Pattern Anal. & Machine Intell., **PAMI-7**, 2, pp. 139 – 154 (1985).

[4] A. M. Waxman and S. S. Sinha: "Dynamic stereo : Passive ranging to moving objects from relative image flows", IEEE Trans. Pattern Anal. & Machine Intell., **PAMI-8**, 4, pp. 406 – 412 (1986).

[5] A. M. Waxman and J. H. Duncan: "Binocular image flows : Steps toward stereo-motion fusion", IEEE Trans. Pattern Anal. & Machine Intell., **PAMI-8**, 6, pp. 715 – 729 (1986).

[6] A. Tamtaoui and C. Labit: "Constrained disparity and motion estimator for 3DTV image sequence coding", Signal Process. : Image Commun., **4**, 1, pp. 45 – 54 (1991).

[7] Y. C. Kim and J. K. Aggarwal: "Determining object motion in a sequence of stereo images", IEEE J. Robotics and Automation, **RA-3**, 6, pp. 599 – 614 (1987).

[8] J. P. Gambotto: "Determining stereo correspondences and egomotion from a sequence of stereo images", Proc. IAPR 10th ICPR, pp. 259 – 262 (1990).

[9] J. Liu and R. Skerjanc: "Stereo and motion correspondence in a sequence of stereo images", Signal Process. : Image Commun., **5**, 4, pp. 305 – 318 (1993).

[10] Y. Altunbasak, A. M. Tekalp and G. Bozdagi: "Simultaneous motion-disparity estimation and segmentation from stereo", Proc. IEEE ICIP'94, Vol. III, pp. 73 – 77 (1994).

[11] S. Malassiotis and M. G. Strintzis: "Joint motion/disparity MAP estimation for stereo image sequences", IEE Proc. Vis. Image Signal Process., **143**, 2, pp. 101 – 108 (1996).

[12] U. R. Dhond and J. K. Aggarwal: "Structure from stereo - a review", IEEE Trans. Syst., Man, & Cybernetics, **19**, 6, pp. 1489 – 1510 (1989).

[13] T. S. Huang and A. N. Netravali: "Motion and structure from feature correspondences : A review", Proc. IEEE, **82**, 2, pp. 252 – 268 (1994).

[14] M. E. Lukacs: "Predictive coding of multi-viewpoint image sets", Proc. IEEE ICASSP 86, pp. 521 – 524 (1986).

[15] W. A. Schupp: "A study on the efficient coding of three-dimensional moving pictures", Master's thesis, Course of Elec. Eng., School of Eng., The Univ. of Tokyo (1988).

[16] I. Dinstein, G. Guy, J. Rabany, J. Tzelgov and A. Henik: "On the compression of stereo images : Preliminary results", Signal Process., **17**, 4, pp. 373 – 382 (1989).

[17] H. Yamaguchi, Y. Tatehira, K. Akiyama and Y. Kobayashi: "Stereoscopic images disparity for predictive coding", Proc. IEEE ICASSP'89, pp. 1976 – 1979 (1989).

[18] B. Choquet, F. Chassaing, J. Fournier, D. Pelé, A. Poussier and H. Sanson: "3D TV studies at CCETT", Proc. TAO 1st Int. Symp. 3D Image Commun. Technol., S-1-2 (1993).

[19] B. L. Tseng and D. Anastassiou: "Compatible video coding of stereoscopic sequences using MPEG-2's scalability and interlaced structure", Proc. Int. Workshop on HDTV'94, 6-B-5 (1994).

[20] A. Puri, R. V. Kollarits and B. G. Haskell: "Basics of stereoscopic video, new compression results with MPEG-2 and a proposal for MPEG-4", Signal Process. : Image Commun., **10**, pp. 201 – 234 (1997).

[21] M. Ziegler, L. Falkenhagen, R. Horst and D. Kalivas: "Evolution of stereoscopic and three-dimensional video", Signal Process. : Image Commun., **14**, pp. 173 – 194 (1998).

[22] M. G. Perkins: "Data compression of stereopairs", IEEE Trans. Commun., **40**, 4, pp. 684 – 696 (1992).

[23] S. Sethuraman, M. W. Siegel and A. G. Jordan: "A multiresolution framework for stereoscopic image sequence compression", Proc. IEEE ICIP'94, Vol. II, pp. 361 – 365 (1994).

[24] S. C. Pei and C. L. Lai: "An efficient coding algorithm for 3D video with spatio-temporal HVS model and binary correlator disparity estimator", Proc. Int. Workshop on HDTV'95, pp. 2B–9 – 2B–16 (1995).

[25] D. Craievich and A. C. Bovik: "Stereo image compression using VPIC", Proc. IEEE ICIP'96, Vol. II, pp. 879 – 882 (1996).

[26] L. B. Stelmach and W. J. Tam: "Stereoscopic image coding : Effect of disparate image-quality in left- and right-eye view", Signal Process : Image Commun., **14**, pp. 111 – 117 (1998).

[27] H. Aydinoğlu and I. M. H. Hayes: "Stereo image coding : A projection approach", IEEE Trans. Image Process., **7**, 4, pp. 506 – 516 (1998).

[28] S. H. Seo and M. R. A. Sadjadi: "2-d filter-based disparity compensation using sequential orthogonal subspace updating", Proc. IEEE ICIP'98, Vol. II, pp. 603 – 607 (1998).

[29] N. Grammalidis, S. Malassiotis, D. Tzovaras and M. G. Strintzis: "Stereo image sequence coding based on three-dimensional motion estimation and compensation", Signal Process. : Image Commun., **7**, 2, pp. 129 – 145 (1995).

[30] F. Labonté, C. T. Dinh and P. Cohen: "A NTSC-compatible compact representation for stereoscopic sequences", Proc. IEEE ICIP'96, Vol. II, pp. 883 – 886 (1996).

[31] W. Woo and A. Ortega: "Dependent quantization for stereo image coding", Proc. SPIE VCIP'98, Vol. 3309, pp. 902 – 913 (1998).

[32] K. Aizawa and T. S. Huang: "Model-based image coding : Advanced video coding techniques for very low bit-rate applications", Proc. IEEE, **83**, 2, pp. 259 – 271 (1995).

[33] T. Naemura, K. Yoshioka, T. Fujii, Y. Nakaya and H. Harashima: "Disparity compensated predictive coding of multi-view 3-D images", Proc. 1993 ITE Annual Convention, 24-4 (1993). (in Japanese).

[34] H. Aydinoğlu and I. M. H. Hayes: "Multi-view image cod-

ing using local orthogonal bases", Proc. SPIE VCIP'97, Vol. 3024, pp. 340 – 351 (1997).
[35] M. H. Kiu, X. S. Du, R. J. Moorhead, D. C. Banks and R. Machiraju: "Two dimensional sequence compression using mpeg", Proc. SPIE VCIP'98, Vol. 3309, pp. 914 – 921 (1998).
[36] K. Oshima and T. Okoshi: "Synthesis of an autostereoscopic 3-D image from binocular stereoscopic images", Appl. Opt., **18**, 4, pp. 469 – 476 (1979).
[37] Y. Sasaki and T. Furukawa: "A synthesis algorithm of arbitrary intermediate-viewing-angle images from several pairs of stereograms", Trans. IEICE, **J63-D**, 9, pp. 813 – 814 (1980). (in Japanese).
[38] R. Skerjanc and J. Liu: "A three camera approach for calculating disparity and synthesizing intermediate pictures", Signal Process. : Image Commun., **4**, 1, pp. 55 – 64 (1991).
[39] S. E. Chen and L. Williams: "View interpolation for image synthesis", Proc. SIGGRAPH'93, pp. 279 – 288 (1993).
[40] E. A. Hendriks and G. Marosi: "Recursive disparity estimation algorithm for real time stereoscopic video application", Proc. IEEE ICIP'96, Vol. II, pp. 891 – 894 (1996).
[41] B. L. Tseng and D. Anastassiou: "A theoretical study on an accurate reconstruction of multiview image based on the viterbi algorithm", Proc. IEEE ICIP'95, Vol. II, pp. 378 – 381 (1995).
[42] N.Grammalidis and M.G.Strintzis: "Disparity and occlusion estimation for multiview image sequences using dynamic programming", Proc. IEEE ICIP'96, Vol. II, pp. 337 – 340 (1996).
[43] M. E. Izquierdo: "Stereo matching for enhanced telepresence in three-dimensional videocommunications", IEEE Trans. Circuits & Syst. Video Technol., **7**, 4, pp. 629 – 643 (1997).
[44] M. E. Izquierdo: "Stereo image analysis for multi-viewpoint telepresence applications", Signal Process. : Image Commun., **11**, pp. 231 – 254 (1998).
[45] B. L. Tseng and D. Anastassiou: "Multiviewpoint video coding with MPEG-2 compatibility", IEEE Trans. Circuits & Syst. Video Technol., **6**, 4, pp. 414 – 419 (1996).
[46] J. R. Ohm and K. Müller: "Incomplete 3D representation of video objects for multiview applications", Proc. PCS'97, pp. 427 – 432 (1997).
[47] T. Fujii and H. Harashima: "Data compression and interpolation of multi-view image set", IEICE Trans. Inf. & Syst., **E77-D**, 9, pp. 987 – 995 (1994).
[48] D. V. Papadimitriou and T. J. Dennis: "Stereo in model-based image coding", Proc. PCS'94, pp. 296 – 299 (1994).
[49] J. R. Ohm and M. E. Izquierdo: "An object-based system for stereoscopic viewpoint synthesis", IEEE Trans. Circuits & Syst. Video Technol., **7**, 5, pp. 801 – 811 (1997).
[50] T. Naemura, M. Kaneko and H. Harashima: "3-D object based coding of multi-view images", Proc. PCS'96, pp. 459 – 464 (1996).
[51] T. Naemura, T. Yanagisawa, M. Kaneko and H. Harashima: "3-D layered representation of multiview images based on 3-D segmentation", J. Inst. Telev. Eng., **50**, 9, pp. 1335 – 1344 (1996). (in Japanese).
[52] D. Tzovaras, N. Grammalidis and M. G. Strintzis: "Object-based coding of stereo image sequences using joint 3-D motion/disparity compression", IEEE Trans. Circuits & Syst. Video Technol., **7**, 2, pp. 312 – 327 (1997).
[53] S. Malassiotis and M. G. Strintzis: "Coding of video-conference stereo image sequences using 3D models", Signal Process. : Image Commun., **9**, 2, pp. 125 – 135 (1997).
[54] S. Malassiotis and M. G. Strintzis: "Object-based coding of stereo image sequences using three-dimensional models", IEEE Trans. Circuits & Syst. Video Technol., **7**, 6, pp. 892 – 905 (1997).
[55] I. Kompatsiaris, D. Tzovaras and M. G. Strintzis: "Flexible 3D motion estimation and tracking for multiview image sequence coding", Signal Process. : Image Commun., **14**, pp. 95 – 110 (1998).
[56] A. Redert and E. Hendriks: "Disparity map coding for 3D teleconferencing applications", Proc. SPIE VCIP'97, Vol. 3024, pp. 369 – 379 (1997).
[57] H. Fan and K. N. Ngan: "Disparity map coding based on adaptive triangular surface modelling", Signal Process. : Image Commun., **14**, pp. 119 – 130 (1998).
[58] D. Tzovaras, N. Grammalidis and M. G. Strintzis: "Disparity field and depth map coding for multiview 3D image generation", Signal Process. : Image Commun., **11**, pp. 205 – 230 (1998).
[59] D. Tzovaras and M. G. Strintzis: "Motion and disparity estimation using rate distortion theory for very low bit rate and multiview image sequence coding", Proc. SPIE VCIP'97, Vol. 3024, pp. 352 – 359 (1997).
[60] C. L. Pagriari, M. M. Perez and T. J. Dennis: "Reconstruction of intermediate views from stereoscopic images using a rational filter", Proc. IEEE ICIP'96, Vol. II, pp. 627 – 631 (1998).
[61] J. R. Ohm, K. Grŭneberg, E. Hendriks, M. E. Izquierdo, D. Kalivas, M. Karl, D. Papadimatos and A. Redert: "A realtime hardware system for stereoscopic videoconferencing with viewpoint adaptation", Signal Process. : Image Commun., **14**, pp. 147 – 171 (1998).
[62] R. C. Bolles, H. H. Baker and D. H. Marimont: "Epipolar-plane image analysis : An approach to determining structure from motion", Int. J. Comput. Vision, **1**, pp. 7 – 55 (1987).
[63] H. H. Baker and R. C. Bolles: "Generalizing epipolar-plane image analysis on the spatiotemporal surface", Int. J. Comput. Vision, **3**, pp. 33 – 49 (1989).
[64] T.Fujii, J.Hamasaki and M.Pusch: "Data compression for an autostereoscopic 3D image", Proc. Int. Symp. on Three Dimensional Image Technol. & Arts, pp. 171 – 178 (1992).
[65] T. Naemura, M. Kaneko and H. Harashima: "3-D segmentation of multi-view images based on disparity estimation", Proc. SPIE VCIP'96, Vol. 2727, pp. 1173 – 1184 (1996).
[66] T. Naemura and H. Harashima: "Fractal coding of a multiview 3-D image", Proc. IEEE ICIP'94, Vol. III, pp. 107 – 111 (1994).
[67] S.E.Chen: "QuickTimeVR - an image-based approach to virtual environment navigation -", Proc. SIGGRAPH'95, pp. 29 – 38 (1995).
[68] W. O. Cochran, J. C. Hart and P. J. Flynn: "Fractal volume compression", IEEE Trans. Visualization and Computer Graphics, **2**, 4, pp. 313 – 321 (1996).
[69] J. Luo and C. W. Chen: "Coherently three-dimensional wavelet-based approach to volumetric image compression", J. Electronic Imaging, **7**, 3, pp. 474 – 485 (1998).
[70] H. Yoshikawa and J. Tamai: "Holographic image compression by motion picture coding", Proc. SPIE Practical Holography X, Vol. 2652 of 01 (1996).
[71] H. Harashima: "Three-dimensional image coding, future prospect", Proc. PCSJ'92, pp. 9 – 12 (1992). (in Japanese).
[72] T. Fujii: "A Basic Study on the Integrated 3-D Visual Communication", PhD thesis, Course of Elec. Eng., School of Eng., The Univ. of Tokyo (1994). (in Japanese).
[73] T. Fujii and H. Harashima: "Coding of an autostereoscopic 3-D image sequence", Proc. SPIE VCIP'94, Vol. 2308, pp. 930 – 941 (1994).
[74] T. Yanagisawa, T. Naemura, M. Kaneko and H. Harashima: "Handling of 3-D objects using ray space", J. Inst. Telev. Eng., **50**, 9, pp. 1345 – 1351 (1996). (in Japanese).

[75] A. Katayama, K. Tanaka, T. Oshino and H. Tamura: "A viewpoint independent stereoscopic display using interpolating of multi-viewpoint images", Proc. SPIE Stereoscopic displays and virtualreality systems II, Vol. 2409, pp. 11 – 20 (1995).
[76] L. McMillan and G. Bishop: "Plenoptic modeling : An image-based rendering system", Proc. SIGGRAPH'95, pp. 39 – 46 (1995).
[77] M. Levoy and P. Hanrahan: "Light field rendering", Proc. SIGGRAPH'96, pp. 31 – 42 (1996).
[78] S. Gortler, R. Grzeszczuk, R. Szeliski and M. Cohen: "The lumigraph", Proc. SIGGRAPH'96, pp. 43 – 54 (1996).
[79] E. Adelson and J. Bergen: "The plenoptic function and the elements of early vision", Computer Models of Visual Processing (Eds. by M. Landy and J. Movshon), MIT Press, chapter 1 (1991).
[80] T. Fujii, M. Kaneko and H. Harashima: "Representation of 3D spacial information by rays and its application", J. Inst. Telev. Eng., **50**, 9, pp. 1312 – 1318 (1996). (in Japanese).
[81] T. Naemura, M. Kaneko and H. Harashima: "Efficient representation of 3-D spatial data based on orthogonal projection of ray data samples", J. Inst. Image Inf. & Telev. Eng., **51**, 12, pp. 2082 – 2089 (1997). (in Japanese).
[82] T. Naemura, M. Kaneko and H. Harashima: "Orthographic approach to representing 3-D images and interpolating light rays for 3-D image communication and virtual environment", Signal Process. : Image Commun., **14**, pp. 21 – 37 (1998).
[83] T. Takano, T. Naemura, M. Kaneko and H. Harashima: "3-d space coding based on light ray data - local expansion of compressed light ray data -", J. Inst. Image Inf. & Telev. Eng., **52**, 9, pp. 1321 – 1327 (1998). (in Japanese).
[84] T. Fujii, T. Kimoto, and M. Tanimoto: "Data compression of 3-d spatial information based on ray-space coding", J. Inst. Image Inf. & Telev. Eng., **52**, 3, pp. 356 – 363 (1998). (in Japanese).
[85] A. Katayama, S. Uchiyama, H. Tamura, T. Naemura, M. Kaneko and H. Harashima: "A cyber-space creation by mixed ray space data with geometric models", Trans. IEICE, **J80-D-II**, 11, pp. 3048 – 3057 (1997). (in Japanese).
[86] S. Uchiyama, A. Katayama, A. Kumagai, H. Tamura, T. Naemura, M. Kaneko and H. Harashima: "Collaborative cybermirage: A shared cyberspace with mixed reality", Proc. VSMM'97, pp. 9 – 18 (1997).

Takeshi Naemura received the B.E., M.E. and Dr.E. degrees in Electronic Engineering from the University of Tokyo, Tokyo, Japan, in 1992, 1994 and 1997 respectively. He joined the Department of Information and Communication Engineering at the University of Tokyo as a Research Associate in 1997, where he is currently an Assistant Professor. His research interests include augmented spatial communication (next generation 3-D image communication and sound field communication), image-based rendering, mixed reality, data compression, and human interface. He is a member of the IEEE Computer Society, the Institute of Image Information and Television Engineers, the Virtual Reality Society of Japan, and 3D Forum.

Masahide Kaneko received the B.E. degree in Electronic Engineering, M.E. degree in Electrical Engineering and Dr.E. degree in Electronic Engineering from the University of Tokyo, Japan in 1976, 1978, and 1981, respectively. From April 1981 to March 1994, he was with the Research and Development Laboratories of Kokusai Denshin Denwa Co., Ltd. (KDD). From April 1994 to March 1997, he was an associate professor of Department of Information and Communication Engineering, the University of Tokyo. In April 1997, he was reinstated in the Research and Development Laboratories of KDD. In April 1998, he has joined the University of Electro-Communications as an associate professor of Course in Electronic Engineering. His research interests include low bit-rate video coding, model-based image coding, processing of facial image information, integrated 3-D visual communications and human communications engineering. He is a member of the IEEE Computer Society, the Institute of Image Information and Television Engineers, and Japan Academy of Facial Studies (J-face).

Hiroshi Harashima received the B.E., M.E. and Dr.E. degrees in electrical engineering from the University of Tokyo, Tokyo, Japan, in 1968, 1970 and 1973, respectively. He joined the Department of Electrical Engineering at the University of Tokyo as an Assistant Professor in 1973, where he is currently a Professor. In 1984, he was a visiting associate at Stanford University, Stanford, CA. His research interests include communication theory, image coding and processing, and human communications enginnering. He is the co-author of the books entitled "Digital Signal Processing," and "Information and Coding Theory." Dr. Harashima received the 1973 Yonezawa Memorial Awards, the 1979 and 1991 Achievement Awards, and the 1989 and 1991 Best Paper Awards from the Institute of Electronics, Information and Communication Engineers (IEICE), Japan. He is a member of the Japanese Electrical Engineering Society, the Institute of Image Information and Television Engineers, and Japan Academy of Facial Studies (J-face).