Interactive Visual Simulation in a Quasi-three-dimensional World based on the Structuralization of Images

Makoto Kato and Hideo Noyama
Systems Development Laboratory, Hitachi, ltd.
1099 Ozenji, Asao-ku, Kawasaki, Kanagawa 215, JAPAN
E-mail: kato@sdl.hitachi.co.jp, noyama@sdl.hitachi.co.jp

ABSTRACT

To use photographs to construct a virtual world in order to perform interactive visual simulation, we proposed a structuralization concept that makes interactive image collage and viewpoint change around objects possible. Most of the virtual reality systems use computer graphics to visualize their virtual worlds, but, it would be very convenient to instead use photographs or video sequences, because they can be acquired more easily and their quality is higher. Objects in photographs cannot be moved, however, and it is difficult to change the point from which they appear to be viewed. We therefore use a "structuralization of images," in which physically significant regions (corresponding to objects) are segmented in detail and are classified according to their optical transmission and reflection properties. Depth information is added to the segmented regions, which can be treated and seen as a quasi-three-dimensional world in photographs. We have also developed a structuralized morphing technique that makes it possible to change viewpoint around an object in photographs.

1. Introduction

Although most virtual reality systems [Cruz-Neira et al., 1993], [Kitamura et al., 1994] use computer graphics to visualize their virtual worlds, there have also been considerable efforts made to use photographs when constructing virtual worlds [Nakamae et al., 1986], [Hirose & Yokoyama, 1992], [Hirose et al., 1994]. This is because the generation of computer graphics requires laborious processes for modeling and rendering, whereas photographs can be taken easily and are superior in quality. Two disadvantages of using photographs, however, are that objects in photographs cannot be moved and that it is difficult to change viewpoints.

Photo retouch tools on personal computers are widely used in synthesizing new images by overlapping, and elaborate collage images are created by iterating the processes of masking, overlapping, scale adjusting, etc. Such tools themselves, however, are not suitable for the use in visual presentation or interactive visual simulation. They are instead usually used in preparing materials displayed in visual presentation.

A photograph-based interactive visual simulation system that can easily move objects, change their colors, and superpose them in a scene as is done in virtual reality applications where computer graphics is applied would be very convenient, and to construct such an advanced visual simulation system, we developed a concept of image structuralization. Instead of the simple image region masking performed in photo retouch tools, interactive system we propose selects physically significant regions, (which correspond to objects) segments them in detail and classifies them according to their optical transmission and reflection properties. Depth information is added to the segmented regions, so the structuralized images can be seen in photographs as a quasi-three-dimensional world.

We also developed, as an expansion of the usual morphing technique [Beier & Neely, 1992], [Wolberg, 1990], a structuralized morphing technique making it possible to change viewpoint around an object in photographs. We are utilizing these functions to develop virtual reality systems that are based on photographs and videos and that can be used in combination with computer graphics.

The structuralization of image is discussed in the next section and visual presentation utilizing image collage in quasi-three-dimensional worlds is explained in section 3. After discussing the structuralized morphing technique and its application to viewpoint change in photographs in section 4, we conclude by discussing the relationships between this structuralization concept and the computer graphics used in virtual reality systems.

2. Image Structuralization

To structuralize an image, we must segment the image into physically significant regions, add attributes to each region, and define the three-dimensional arrangement of the regions.

2.1 Image segmentation

Because of a digital image is an array of pixels originally and has no structures in it, physically significant regions have to be specified when we perform visual simulation. A woman's image, for example, can be segmented into a face, cloths, arms, legs, and a background as shown in **Fig. 1**. This segmentation makes visual simulation possible—such as changing the color of specified parts of a object in the image and moving a specified object to overlap another background image—because the processed pixels are determined according to which region they are in.

We have developed tools utilizing color information [Kato, Miyaoka & Noumi, 1990] and edge information [Kato, Morita, Miyaoka & Noumi, 1990] to help segment the physically significant regions and thereby speed the authoring processes prior to the visual simulation, but these tools are not discussed here.

Fig. 1 Image segmentaion.

2.2 Adding attributes

The optical properties of each of the segmented regions have to be specified by adding attributes. In the simplest case, one specifies whether it is opaque or semitransparent (including completely transparent). Our system considers two kinds of semitransparency, one is called adding semitransparency and the other called multiplying semitransparency.

A typical example of adding semitransparency is when the new pixel values for a window glass are calculated by weighted averaging when the glass is made to overlap background images.

When a pixel $A(i,j)$ in an adding semitransparent object is overlapped onto a pixel $B(k,l)$ in a background image, the pixel $C(m,n)$ in the newly synthesized image $C(m,n)$ becomes

$$C(m,n) = (1-p)A(i,j) + pB(k,l), \quad (1)$$

where p is the transmission coefficient.

Similarly, shadows on the ground are categorized as multiplying semitransparent objects. When B_0 is a parameter which expresses the average pixel value in the background image around the shadow to be created, the formula for the new pixel becomes

$$C(m,n) = A(i,j)B(k,l)/B_0 \quad (2)$$

In more elaborate treatments, attributes contain surface color information and surface reflection properties. In the dichromatic reflection model [Klinker et al., 1988], a pixel value in a region with a uniform dielectric surface is considered to be a sum of surface reflection and body reflection. For smooth surfaces, the former corresponds to specular reflection and the latter corresponds to diffuse reflection. Let $\mathbf{Cs}$ be the surface reflection color vector and $\mathbf{Cb}$ be the body reflection color vector. Then the observed color vector for pixel $C(i,j)$ is expressed as

$$\mathbf{C}(i,j) = Ms(i,j)\mathbf{Cs} + Mb(i,j)\mathbf{Cb} + \mathbf{Ca}, \quad (3)$$

where $Ms(i,j)$ and $Mb(i,j)$ are respectively the surface reflection scale factor and the body reflection scale factor and where $\mathbf{Ca}$ is the ambient environmental color contribution vector. $\mathbf{Cs}$, $\mathbf{Cb}$, and $\mathbf{Ca}$ are obtained by analyzing on the color distribution in the RGB-color space of the region. Then scale factors $Ms(i,j)$ and $Mb(i,j)$ are calculated for each pixel in the region. Thus if some other color vector is substituted for the body reflection color vector, the new pixel value expresses the new color of the object under the same lighting condition. And if some other color vector is substituted for the surface color vector, the new pixel value expresses the original color under a new lighting condition. In the structuralization process of our system, $Ms(i,j)$ and $Mb(i,j)$ for each pixel in the region, together with three color vectors are extracted (in some normalized form), and can be used to as a attributes of the specified region [Takizawa et al., 1991].

2.3 Defining three-dimensional arrangement

In the last step of the structuralization, three dimensional information about each of the objects is specified [Noyama et al., 1992]. The quasi-three-dimensional world we define here consists of planar objects on a planar ground (**Fig. 2**). Such a world is often a good model for simple visual simulation and is described in detail in the next section. If objects to be overlapped are considered to be on the planar ground of the background image, the vertical positions of the lower bounds of the objects in the background image correspond directly to the depth in the quasi-three-dimensional world. The scales of objects are also decided when

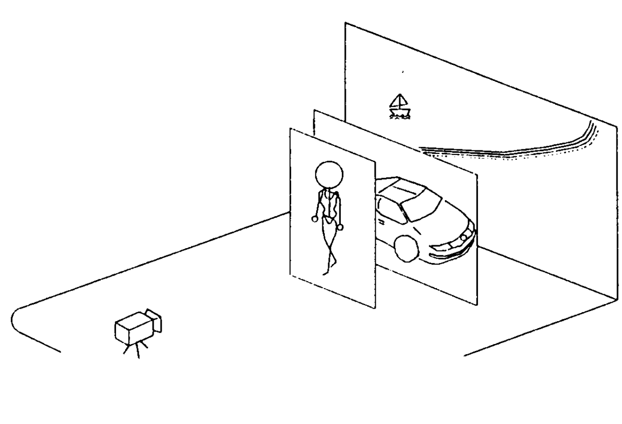


Fig. 2 Quasi-three-dimensional world.

the three-dimensional information is specified.

An object is fitted into the world by specifying its length at two positions. As shown in **Fig. 3**, for example, the length of an object at the distance of y_1 from the bottom of the image is specified to be l_1 and the length of the same object at distance y_2 is specified to be l_2. Then the length l of the object at the distance of y from the bottom of the image should be calculated by the following formula.

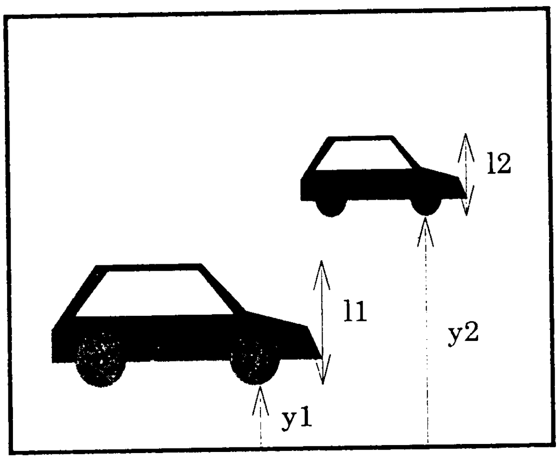
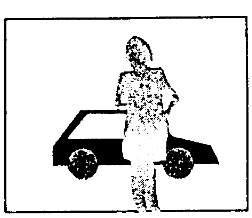


Fig.3 Vertical posion and depth

$$l(y) = \frac{l_2 - l_1}{y_2 - y_1}(y - y_1) + l_1. \tag{4}$$

The structuralization of images is done through these three processes, and the results are kept in a form of chart to be used in the visual simulation program (**Fig. 4**). When we regard the process of structuralization as authoring, interactive visual simulation corresponds to presentation. Structuralized images can then be seen as a presentation content (**Fig. 5**).



region	ID	attribute	depth	optical parameter	more
	0	background	–	–	–
	1	opaque	2	–	color attribute file name
	2	opaque	2	–	–
	3	semitransparent	2	50 %	–
	4	opaque	1	–	–

Fig. 4 Image structuralization.

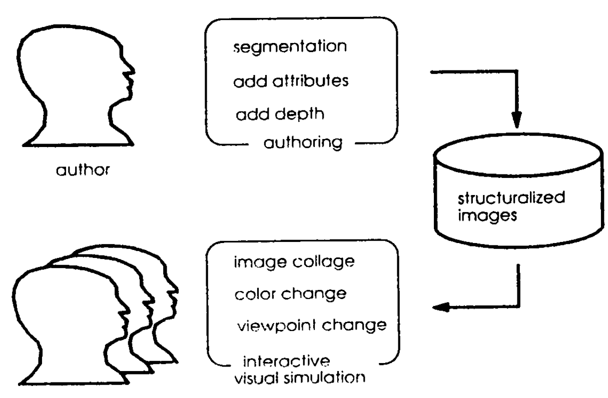


Fig. 5 Structuralized images as a presentaion title.

3. Interactive Visual Simulation and other Special Effects

Visual simulation in the quasi-three-dimensional world can be performed on a personal computer without any special additional hardware. And the structuralized images make various kinds of interactive visual simulation possible. This section shows a few examples of visual simulation and describes other special effects based on structuralization.

3.1 Object superposition onto the background image with the automatic scale adjustment and depth calculation

One often wants to superpose some object onto other background image. Such superposition is very useful for in architectural simulation, layout simulation, and clothing simulation. An object is usually cut out from one image (A) and superposed onto another background image (B), and both of the images are structuralized in the authoring process. There may be objects in the image B, too, and position of the plane of the ground in image B is known. By only pointing to the superposing position in the image B, the size of the object in the image A is calculated automatically. The resulting superposed image of a woman is shown in **Fig. 6**. By pointing to another position in the background image, we get another superposition (**Fig. 7**). When parts of an object in image A are calculated to be hidden by objects in image B, those parts are not written into the newly synthesized image.

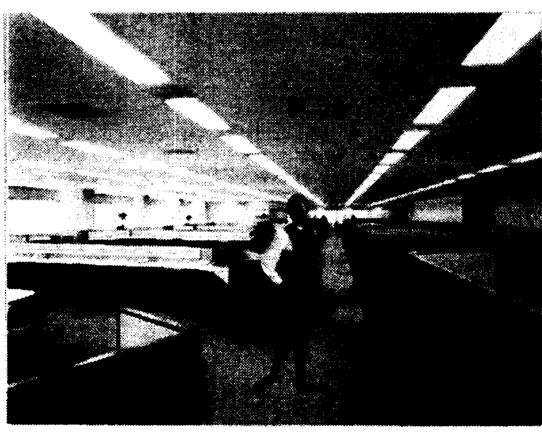


Fig. 6 Image superposition: A woman is superposed onto the images of a office.

Fig.7 Image superposiotion.

3.2 Superposition of computer graphics onto the structuralized image

Figure 8 shows the computer graphics image of a roadside sign for parking superposed in a visual simulation. Because the background image is structuralized and a quasi-three-dimensional planar world is constructed, by only pointing to the superposing position, you can send message to the graphic module to render the sign in the correct size and pose.

3.3 Color change and other special effects

The color of an object that has attributes for color information and surface reflection properties can be changed simply by only selecting a color from the color menu in the display. The colors of the objects in superposed images can also be changed. And if an object has attributes giving information about its deformation pattern, a selected texture image can be mapped onto the segmented region. The kind of cloth, for example, can be changed. A more advanced texture mapping technique let us do texture mapping while preserving the detailed shade and shadow existing in the background images [Morita et al., 1992].

Fig. 8 A computer graphics image of a roadside sign for a parking is superposed into the structuralized background image.

If we compare this kind of simulation with systems that use computer graphics, we find that the only major functions it cannot perform are to change viewpoint and to change the pose of an object. Even these functions, however, are possible when we combine this simulation with the structuralized morphing technique described in the following section.

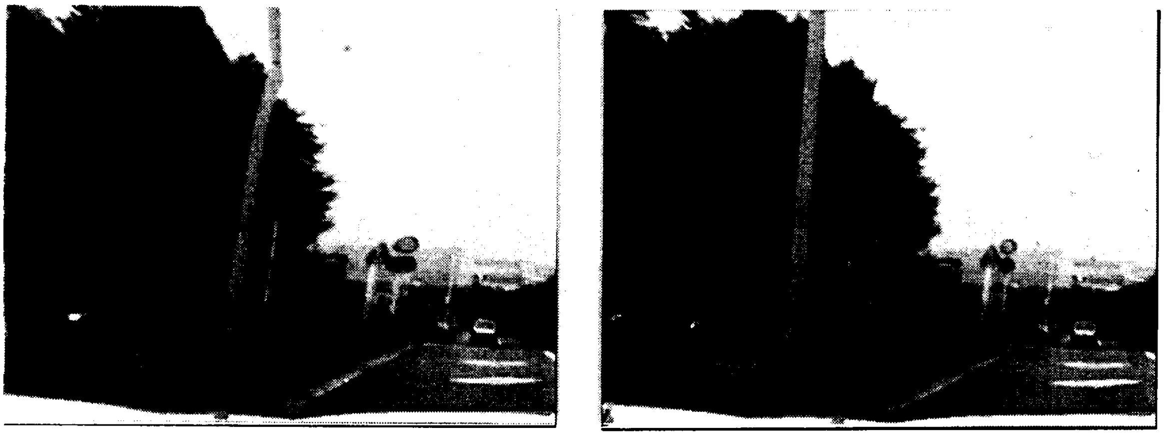
4. Viewpoint Change by Structuralized Morphing

4.1 Morphing as a means for viewpoint change

Morphing between two or more images is a very useful visual technique [Beier & Neely, 1992], [Wolberg, 1990] and is usually performed between two different objects, such as a car and a panther, or two different persons. But it can also be performed between the same object in different poses, between the same scenes from different viewpoints, and between the same scenes or objects at different times. Morphing can thus be a software engine that can express object rotation, viewpoint change, and time evolution. This section describes only viewpoint change simulation by the image morphing.

First, let us briefly review the process of the ordinary morphing, which consists of two tasks: warping two images so that they have the same shape, and cross dissolving. The main task is the warping, and most morphing software defines this warping function by specifying the corresponding points, lines, or polygons between two images. Cross dissolving is a weighted summation of two images after warping. The weight factor is usually a nonlinear monotonic function of how the synthesized image is similar to one of the two input images.

If we simply apply this procedure to the image pair taken from the different viewpoints, we find large distortion around the specified objects (**Fig. 9**). That's because we treated the whole image at the same time. When we change our viewpoint around a specified object, we see that the object appears to move against the background. This observation implies that we should treat each object separately when using morphing as a means of viewpoint change. This is a natural extension of the structuralization of images and is described in the next subsection.



(a) (b)

Fig. 9 Distortion caused by morphing: (a) simple morphing, (b) structuralized morphing.

4.2 Basic procedure of structuralized morphing

Structuralized morphing consists of four steps. The first is the image segmentation, which is the same as in the structuralization of images. The second is the supplimentation of the image area under the object. In the next two steps, morphing is performed separately for each object region, and each object is overlapped according to its depth. The supplimentation step is needed because in the original image, there is no information about the area under object that is going to be moved.

4.3 Experimental results

Figure 10 shows four images taken from different view points: lower left, lower right, upper left, and upper right. In this arrangement of input images, you can easily understand that the point of view can be changed two dimensionally both horizontally and vertically. Warp functions are specified for each of the neighboring pairs and **Fig. 11** show the interpolated image. The image arrangement is of course not limited to this pattern. You can arrange more than two images horizontally or vertically, and images can be arranged three-dimensionally. Thus you can change viewpoint from far to near as well as shifting it horizontally and vertically. In this example, only the tree in the front of the image is segmented and treated separately. If we segment more objects in the image, the quality of the interpolated images improves and the scene looks more natural.

5. Discussions

Fig. 10 Four original imgaes for view point change

This paper described a photograph-based visual simulation program and the concept of structuralization of images, and it described the viewpoint change techniques based on the structuralized morphing. This system makes interactive simulation easy and can be applied to virtual reality systems. Although the lowest levels of this simulation program use algorithms similar to those of photo retouch tools, this program differs from photo retouch tools in that it works interactively on the basis of previously authored structuralized images. Not only does this program avoid the laborious modeling and rendering processes needed for computer graphics-based simulation, it also enables high-quality images to be exploited without losing the freedom of scene construction in computer.

Structuralization of images is, in a sense, a process in which attributed data such as used in color change simulation ("intrinsic image" in the terminology of Klinker et al. [1988]) is obtained from photographs, brought to a stage midway in the rendering processes of computer graphics, and then redirected back to produce a rendered image. Structuralization thus becomes complete if we can acquire exact three-dimensional information from the photographs [Mori et al., 1994]. We believe it will be a key technique in advanced virtual reality systems based on photographs [Park et al, 1995], [Fujii & Harashima, 1995].

(a) (b)

Fig. 11 Interpolated view, (a) in the center of the bottom, (b) in the center of the four images.

Other work is, from the stand point of using photographs, related to our approach. Chen and Williams [1993] used morphing as a means of viewpoint changing, but they required range data. Furthermore, they used simple morphing on whole image, rather than on a structuralized image. Bolles, Bakers, and Marimont [1987] proposed epipolar plane image theory and used it in changing viewpoints, but they needed a lot of images to do this. Our structuralized morphing, on the other hand, needs only a few images and does not need range data.

It is perhaps useful to briefly consider the relation between this morphing and the stereoscopic three-dimensional measurement. When we make a correspondence to calculate the warping function between two images of the same scene from different viewpoints, the process is quite similar to the calculation of disparity in binocular stereo. And although morphing in two dimension do not use three-dimensional information explicitly, these warping process based on correspondence use it implicitly.

REFERENCES

Beier, T. & Neely, S. [1992] "Feature-based image metamorphosis," *SIGGRAPH'92 Proceedings*, pp. 35-42.

Bolles, R. C., Baker, H. H. & Marimont, D. H. [1987] "Epipolar-plane image analysis: An approach to determining structure from motion," *International Journal of Computer Vision*, vol. 1, pp. 7-55.

Chen, S. E. & Williams, L. [1993] "View interpolation for image synthesis," *SIGGRAPH'93 Proceedings*, pp. 279 - 288.

Cruz-Neira, C., Sandin, D. J. & DeFanti, T. A. [1993] "Surround-screen projection-based virtual reality: the design and implementation of the CAVE," *SIGGRAPH'93 Proceedings*, pp. 135 - 142.

Fujii, T. & Harashima, H. [1995] "Data compression and interpolation of multi-view image set," *IEICE transactions on information and systems*, vol. E77-D, no. 9, pp. 987 - 995.

Hirose, M. & Yokoyama, K. [1992] "VR application for transmission of synthetic sensation," *The second international conference on artificial reality and tele-existence*, pp. 145 - 154.

Hirose, M., Takahashi, K., Koshizuka, T. & Watanabe, Y. [1994] "A study on image editing technology for synthetic sensation," *The fourth international conference on artificial reality and tele-existence*, Tokyo, Japan, Nihon Keizai Shinbun, Inc., pp. 63 - 70.

Kato, M., Miyaoka, S. & Noumi, M. [1990] "Interactive image segmentation method based on the color information," *Proceedings of the 1990 IECIE Spring conference*, p. 7-263 (in Japanese).

Kato, M., Morita, T., Miyaoka, S., & Noumi, M. [1990] "Application of Hopfield network to human-guided contour extraction," *MVA'90 IAPR workshop on machine vision applications*, Tokyo, Japan, pp. 413 - 416.

Kitamura, Y., Takemura, H. & Kishino, F. [1994] "Coarse-to-fine collision detection for real-time applications in virtual workspace," *The fourth international conference on artificial reality and tele-existence*, Tokyo, Japan, Nihon Keizai Shinbun, Inc., pp. 147 - 157.

Klinker, G. J., Shafer, S. A. & Kanade, T. [1988] "The measurement of highlights in color images," *International Journal of Computer Vision*, vol. 2, pp. 7 - 22.

Mori, Y., Kato, M., Miyaoka, S. & Noumi, M. [1994] "A proposal of interactive 3D information extraction," *Proceedings of the 48th Annual Convention IPS Japan*, vol. 2, pp. 35 - 36, (in Japanese).

Morita, T., Kato, M. & Miyaoka, S. [1992] "Texture mapping technique considering shadow and shade," *Proceedings of the 1992 IECIE Spring conference*, p. 7-206 (in Japanese).

Nakamae, E., Harada, K., Ishizaki, T. & Nishita, T. [1986] "A montage method: the overlaying of the computer generated images onto a background photograph," *SIGGRAPH'86 Proceedings*, pp. 207 - 214.

Noyama, H., Kato, M., Miyaoka, S. & Noumi, M. [1992] "Image composition method by introducing the pseudo-3D-Model," *Proceedings of the 1992 IECIE Autumn conference*, p. 6-237 (in Japanese).

Takizawa, Y., Kato, M, & Miyaoka, S. [1991] "Color change simulation system." *Proceedings of the 1991 IECIE Autumn conference*, p. 6-182, (in Japanese).

Park, J., Yagi, N. & Enami, K. [1995] "Image synthesis based on estimation of camera parameters from image sequence," *IEICE transactions on information and systems*, vol. E77-D, no. 9, pp. 973 - 986.

Wolberg, G. [1990] *Digital Image Warping*, (IEEE Computer Society Press).