A Method for Surface Reconstruction and Synthesizing Intermediate Images for Multi-viewpoint 3-D Displays

Mahito Fujii, Takayuki Ito, Sei Miyake

Science and Technical Research Laboratories
NHK(Japan Broadcasting Corporation)
1-10-11, Kinuta, Setagaya-ku, Tokyo 157, Japan
Tel: +81-3-5494-2369, Fax: +81-3-5494-2371, Email: fujii@strl.nhk.or.jp

ABSTRACT

In this paper, a method for 3-D surface reconstruction with two real cameras is presented. The method, which combines the extraction of binocular disparity and its interpolation, can be applied to the synthesis of images from virtual viewpoints. The synthesized virtual images are as natural as the real images even when we observe the images as stereoscopic images. The method opens up many applications, such as synthesizing input images for multi-viewpoint 3-D displays, enhancing the depth impression in 2-D images and so on.

We also have developed a video-rate stereo machine able to obtain binocular disparity in 1/30 sec with two cameras. We show the performance of the machine.

I. INTRODUCTION

Recently, many 3-D displays have been developed. A type of 3-D display which gives us 3-D images from many viewpoints has been developed. In particular, multi-viewpoint autostereoscopic displays[1] are very attractive because observed 3-D images change according to the movement of the viewing position and they give the observers very natural sensation. However, when many real cameras are used to get many input images simultaneously, it is difficult to adjust the various parameters of the cameras, such as convergence, zoom, and color. One idea for avoiding this troublesome matter is to synthesize images from intermediate viewpoints between two real cameras.

In order to synthesize intermediate images, we have developed a method for recovering the 3-D surface of objects from 2-D images taken by two cameras. Once the 3-D surfaces of objects have been recovered in a computer, it is fairly easy to synthesize images from virtual viewpoints by means of computer graphic techniques.

The method is composed of two processes. In the first process, binocular disparity at edges and lines are extracted by a neural network model. The model makes use of the uniqueness and smoothness constraints to eliminate false matchings. The second process is interpolation of binocular disparity at places where binocular disparities can not be extracted in the first process. The smoothness constraint is also utilized in this process.

In this paper, we first introduce the method and then show that it can be used to synthesize images from several virtual viewpoints. Lastly, we show a machine which can obtain binocular disparity from stereoscopic images at 64 x 64 pixels in 1/30 sec.

II. EXTRACTION OF BINOCULAR DISPARITY

We have developed a neural network model which extracts binocular disparities from a pair of 2-D images[2]. The binocular disparity is the positional difference between the pixels in the two images which have the same physical origin. The difference comes from projections of the object taken from different viewpoints. The most difficult step in extracting binocular disparities is solving the correspondence problem[3]. The step consists of discovering pixels of the same physical origin in the two images. The human brain naturally solves this problem. Evidence for this ability is that we can perceive depth when viewing a random-dot stereogram[4]. Therefore, the following findings in psychology and physiology are helpful.

(1) Orientation of features (for example, edges or

lines) have a strong influence on human depth perception in stereopsis[5].

(2) Neurons with sensitivity to binocular disparity have been found by physiological studies (for example, [6], [7]).

(3) The mechanism of human stereopsis has a disparity gradient limit[8].

(4) The physical world gives some general constraints in projecting the 3-D real world on the retinas. Compatibility, uniqueness and smoothness act as constraints[3].

The neural network model is constructed based on these findings.

The main part of the neural network model consists of many cells which have small receptive fields in the left and right retinas. The receptive field means the areas on the two retinas by which the cell is affected. Each of the cells compares a pair of input patterns with memorized patterns, such as edges with a specific disparity and a specific orientation. A cell outputs large responses when a pair of input patterns has a disparity and an orientation similar to the memorized one(Figure 1). Thus, in our model, the distributed responses of cells represent the 3-D surface shape of an object.

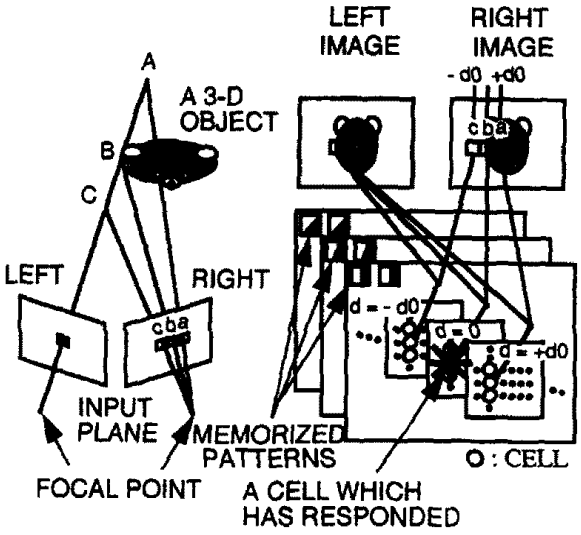


Figure 1. A scheme for extracting binocular disparity in the neural network model.

There are, however, some cells which extract false matchings. In order to depress the cells extracting false matchings, the model make use of the uniqueness and smoothness constraints.

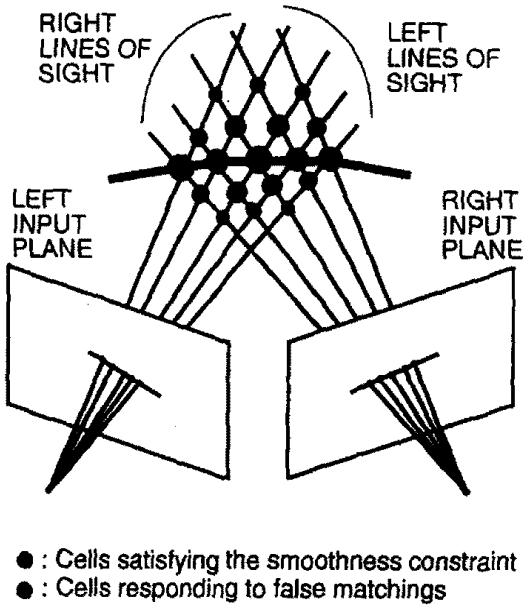


Figure 2. A schematic diagram of the uniqueness and smoothness constraints.

The smoothness constraint is based on the fact that binocular disparity varies smoothly almost everywhere. The uniqueness constraint comes from the fact that we generally see a point on only one surface for each line of sight.

Figure 2 shows a schematic diagram explaining how the uniqueness and smoothness constraints work in our model. Each cell is assigned for detecting a 3-D position corresponding to a point of intersection between the left and the right lines of sight. Excitatory connection exists among neighboring cells satisfying the smoothness constraint. On the other hand, inhibitory connections exists among cells on each line of sight. The inhibitory interaction comes from the uniqueness constraint which means that only one cell responds to correct binocular disparity on each line of sight. As a result of these interactions, the model is able to extract reliable binocular disparities at edges and lines in input images except for a very few false matchings. The model gives a binocular disparity $d_0(x, y)$ and a disparity reliability $u(x, y)$ at each position (x, y). Figure 3 shows an example of binocular disparities extracted by the model when a toy car is used as an object.

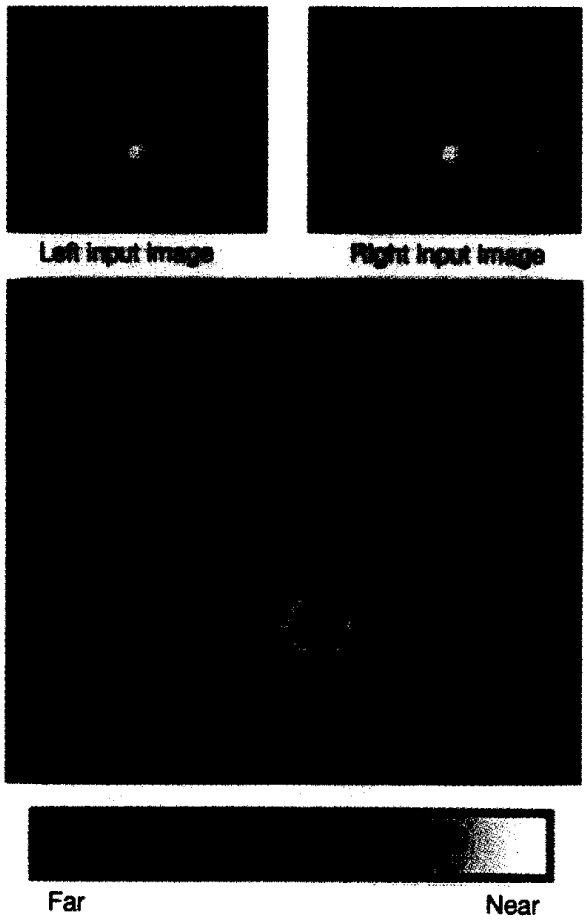


Figure 3. An example of binocular disparities extracted by the model (with inputs images).

III. 3-D SURFACE RECONSTRUCTION

We consider binocular disparities as approximated depth because the camera parameters in geometrical terms are unknown.

In our model, binocular disparities can be extracted only at edges or lines in images as one can see from Figure 3. Hence binocular disparities which the model can not extract must be estimated or be interpolated based on the extracted binocular disparities.

The interpolation can be realized by the standard regularization framework[9]. In this framework, a 3-D surface which minimizes a cost function is obtained. Here, the cost function is defined as a weighted sum of the data fitting terms and smoothness terms at all positions as follows.

$$e = \sum_{x,y=2}^{N,N} [\alpha(x,y) \times (d_0(x,y) - d_f(x,y))^2 + \beta \times \{ (d_f(x,y) - d_f(x-1,y))^2 + (d_f(x,y) - d_f(x,y-1))^2 \}] \quad (1)$$

$$\text{find } \{d_f(x,y)\} \text{ ;''''} e \rightarrow \min \quad (2)$$

A set $\{d_f(x,y)\}$ which minimizes Eq. (1) is the depth recovered at all positions in the left input image. Here, the constant β is a weighting of the smoothness term to the data fitting term.

The weight $\alpha(x,y)$ of the data fitting term at each position is varied according to the sum of disparity reliabilities $u(x,y)$ in the area V.

$$\alpha(x,y) = \begin{cases} C & \text{if } u(x,y) \neq 0.0 \\ 0.0 & \text{otherwise} \end{cases} \quad (3)$$

$$C = \sum_V c(x-x', y-y', d_f - d'_f) \times u(x',y') \quad (4)$$

$$c(\xi,\eta,\zeta) = \frac{\exp\{-\frac{\xi^2+\eta^2+\zeta^2}{2\sigma^2}\}}{\sigma^3\sqrt{(2\pi)^3}} \quad (5)$$

Here, d_f and d'_f mean $d_f(x,y)$ and $d_f(x',y')$, respectively.

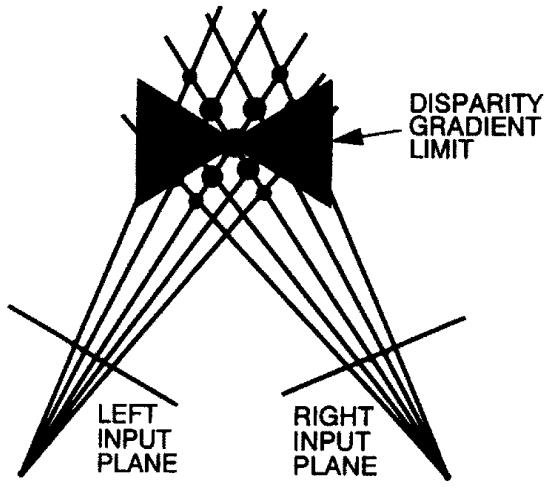


Figure 4. 2-D cross section of disparity gradient limit (top view).

The area V corresponds to the disparity gradient limit.

$$V = \langle''''(x,y)''''|''''L_x, L_y, L_d''''\rangle \quad (6)$$

$$L_x: |x - x'| \leq b \tag{7}$$

$$L_y: |y - y'| \leq b \tag{8}$$

$$L_d: \frac{|d'_0 - d_0|}{\sqrt{\left(\frac{(x' + (x' + d'_0))}{2} - x\right)^2 + (y' - y)^2}} \leq a \tag{9}$$

Here, d_0 and d'_0 mean $d_0(x, y)$ and $d_0(x', y')$, respectively.

In order to find $\{d_f(x, y)\}$ which minimizes Eq. (1), the steepest descent method is used as follows.

$$d_f^{T+1}(x, y) = d_f^T(x, y) - \gamma \Delta d_f^T(x, y) \tag{10}$$

$$\Delta d_f^T(x, y) = \alpha(x, y) \times (d_0(x, y) - d_f^T(x, y)) + \beta \times \{4 d_f^T(x, y) - d_f^{T+1}(x - 1, y) - d_f^T(x + 1, y) - d_f^{T+1}(x, y - 1) - d_f^T(x, y + 1)\} \tag{11}$$

Here, T denotes the number of iterations to find a (relative) minimum of Eq. (1). The method drastically reduces the influence of false matchings and can reconstruct the smooth 3-D surface of the object.

We have also developed an effective method when an object has an almost convex surface. After Eq. (10) is calculated for all (x, y) in each iteration, the following process for all (x, y) is also done.

$$\hat{d}_f^{T+1}(x, y) = \max\{d_f^{T+1}(x, y), v_1, v_2\} \tag{12}$$

$$v_1 = \frac{d_f^{T+1}(x - 1, y) + d_f^{T+1}(x + 1, y)}{2} \tag{13}$$

$$v_1 = \frac{d_f^{T+1}(x, y - 1) + d_f^{T+1}(x, y + 1)}{2} \tag{14}$$

Once the 3-D surfaces of objects are recovered in a computer, then it is easy to synthesize images from virtual viewpoints by means of computer graphic techniques.

Figure 5 shows examples of synthesized images of a 3-D surface observed from four different virtual viewpoints. In this case, the left image is used for surface texture.

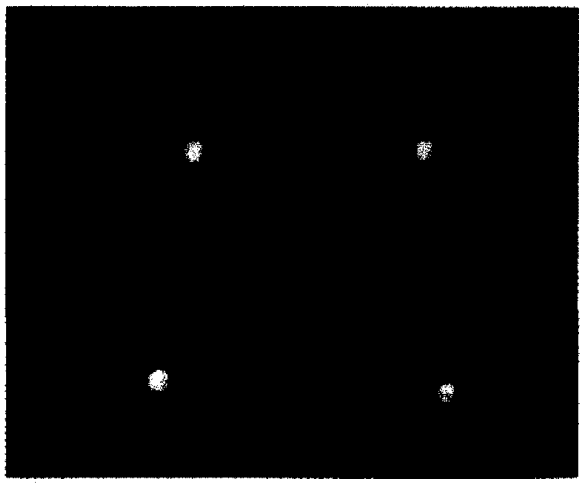
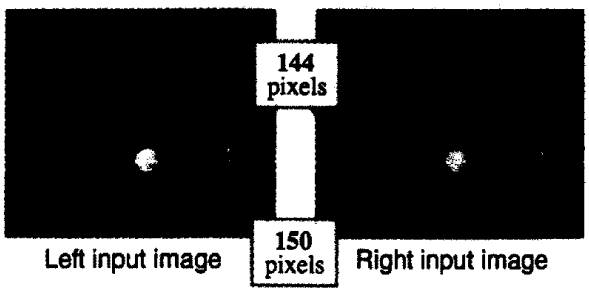


Figure 5. An example of synthesized images.



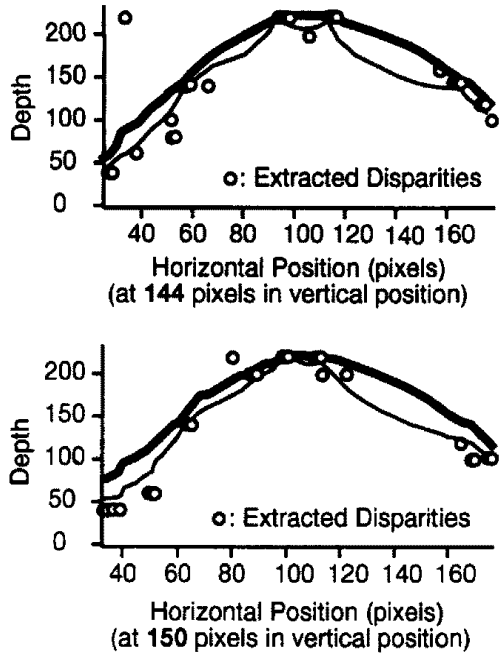
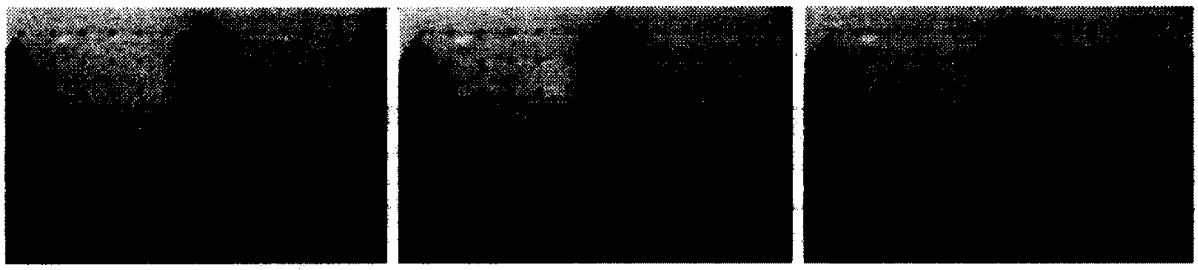


Figure 6. Cross sections of a reconstructed surface along two horizontal lines (thick lines: with convexity assumption, thin lines: without convexity assumption).



Form a left viewing position | The left input image | From a right viewing position

Figure 7. An example of synthesized images.

Figure 6 shows effectiveness of utilizing convexity assumption about an object. As one can see, the method shows good performance in recovering the convexity of a toy car because this toy car has a surface which is almost convex.

In order to avoid falling into a local minima of the cost function, the multiresolution analysis technique is also used. The result at coarser resolution gives good initial values for the iterative calculation of the finer one in the interpolating process.

Another example is shown in Figure 7. The left and the right images in this figure are synthesized by our method. In this case, the method based on convexity assumption is not used. The middle image is an original left input image. When you see the left and the middle images, or the middle and the right images as a pair of stereoscopic images, you can perceive depth.

As you can see from the result, the method can be applied for a multi-viewpoint 3-D display system because the method is able to produce intermediate images for as many viewpoints as the 3-D display requires.

The original images were supplied by the Calibrated Imaging Laboratory located at the Carnegie Mellon University in the U.S.A.

IV. A VIDEO-RATE STEREO MACHINE

The method mentioned above shows good performance in extracting binocular disparity. It takes, however, a lot of calculation time to get binocular disparities in a whole image. That is why a video-rate stereo machine has been developed to extract binocular disparity from stereoscopic images. The machine, which is only a prototype, has the following abilities.

The maximum performance of the machine is; 1) high frame rate of 30 frame/sec; 2) a depth map of 64 x 64 pixels; 3) a disparity search range of up to 512 pixels.

In order to obtain binocular disparity in 1/30 sec, the machine has several salient technical points. They are; 1) using block matching DSP(STI 3220) which is normally used for motion vector detection; 2) using 4 DSP in parallel; 3) using multiresolution image processing to reduce the search time needed for finding the best matching points. Figure 8 shows an example of binocular disparities extracted by the machine in one frame. The machine will be available for making cue signals for segmentation based on depth as well as synthesizing images from virtual viewpoints.

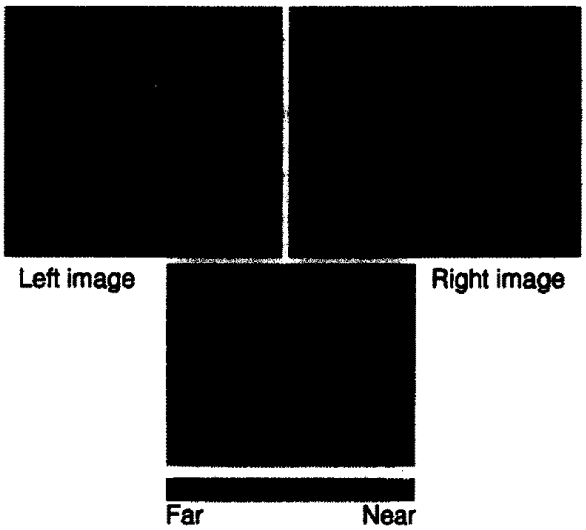


Figure 8. An example of binocular disparities extracted by the video-rate stereo machine with input images.

V. DISCUSSION

We have constructed the neural network model not only to be used for reconstructing 3-D surfaces but also to investigate the mechanism of human depth perception. The proposed model is based on stereopsis findings from psychology and physiology. By repeating the process of constructing a model and evaluating its output, we could point out which mechanism might be essential for human depth perception. If we knew the mechanism of human depth perception, we could transmit 3-D information much more efficiently or could adjust parameters of 3-D display to individual desire.

There are several other methods to get depth information, such as laser scanning and so on. In broadcasting, information on color and texture are very important. Therefore, these methods require us to take images of objects in addition to the depth measurement. Moreover, laser light may not be applied to human beings. These problems do not occur when the depth information is obtained from images taken by cameras. In particular, methods to extract depth information with only two cameras is very important because a lot of stereoscopic 3-D TV programs with the NTSC standard and the HDTV standard have already been stocked.

There are, however, some problems to be solved in our approach. An intermediate image corresponds to viewing position which is different from the real camera position. Generally speaking, if the viewing position changes, the brightness and color of an object change. In order to synthesize a correct intermediate image, we need information on the reflectance properties of surfaces and on light sources. Another problem occurs when we apply the method to scenes which include many objects. In such a case, discontinuity of depth should be detected because the proposed method makes use of the smoothness constraint. We need to detect the discontinuity of depth and switch off the smoothness constraint at the discontinuity points.

VI. CONCLUSION

In this paper, we showed that the proposed method can reconstruct a 3-D surface of an object from 2-D images taken by cameras from only two viewpoints. It was also shown that images from virtual viewpoints can be synthesized from the reconstructed 3-D surface of the object. Moreover, we showed that the developed video-rate stereo machine is able to obtain binocular disparity in 1/30 sec with two cameras although the size of the image is 64 x 64 pixels.

The method opens up many applications including the synthesis of input images for multi-viewpoint 3-D displays.

The authors wish to thank to Mr. Taiji Nishizawa, Mr. Takehiro Izumi, and Dr. Toshio Motoki of NHK Science and Technical Research Laboratories for their encouragement and support in this research. We also wish to thank to Mr. Toshio Nakagawa and the researchers of the Human Vision Research Division for their help.

REFERENCES

1. H.Isono et al., "50-inch autostereoscopic full-color 3-D TV display system," San Jose, California: Proc. SPIE, vol.1669, 1992
2. M.Fujii, "A neural network model for binocular disparity extraction," NHK Laboratories Note, no 378, 1990
3. D.Marr, "Vision", W.H.Freeman and Company, San Francisco, 1982
4. B.Julesz, "Foundations of cyclopean perception", The University of Chicago Press, Chicago, 1970.
5. M.S.Livingstone and D.H.Hubel, "Psychophysical evidence for separate channels for the perception of form, color, movement, and depth," J.Neuroscience, vol.7, no.11, pp.3416-3468, 1987
6. G.F.Poggio and T.Poggio, "The analysis of stereopsis", Ann, Rev, Neurosci., vol.7, pp.379-412, 1984
7. R.Maske, S.Yamane and P.O.Bishop, "End-Stopped cells and binocular depth discrimination in the striate cortex of cats", Proc. R. Soc. Lond. B, vol.229, pp.257-276, 1986
8. P.Burt and B.Julesz, "A disparity gradient limit for binocular fusion," Science, vol.208, no.9, pp.615-617, 1980
9. T.Poggio et al., "Computational Vision and Regularization Theory", Nature, vol.317, no.26, pp.314-319, 1985