Automatic Generation of Stereogram by Using Stripe Backgrounds

Atsushi Yamashita†, Akihisa Nishimoto†, Toru Kaneko† and Kenjiro T. Miura†
† Department of Mechanical Engineering, Shizuoka University
3–5–1 Johoku, Hamamatsu–shi, Shoizuoka 432–8561, Japan
voice: [+81](53)478–1067; fax: [+81](53)478–1067
e·mail: `yamashita@ieee.org`
www: `http://sensor.eng.shizuoka.ac.jp/~yamasita`

Abstract

In this paper, we propose a new method that can generate stereograms (stereoscopic photographs) automatically by compositing foreground images and background images with chromakey using a two-tone stripe background. As to the image composition of foreground images and background images, a chromakey is usually used. However, the conventional chromakey uses the unicolored blue background, and has a problem that one's clothes are regarded as the background if their colors are same. Therefore, we utilize the adjacency condition between two-tone striped areas on the background, and extract the foreground regions whose colors are same with the background. In addition, the stereogram consists of images for a left eye and those for a right eye. The relationships between foreground objects and background objects in composite images must be consistent. The relationships of objects between the left image and the right one must be also consistent. Therefore, the foreground images and the background images are merged by considering the geometric constraints.

Keywords: Image composition, stereogram, chromakey, region extraction

1 Introduction

In this paper, we propose a new method that can generate stereograms (stereoscopic photographs) automatically by compositing foreground images and background images with chromakey using a two-tone stripe background.

The composition of images [1] has significant consequences to creative designs such as cinema films, magazine covers, promotion videos, and so on. This technique can combine images of actors or actresses in a studio and those of scenery that is taken in another place. Additionally, it becomes very important to generate stereograms automatically for virtual realties in recent years. The stereogram consists of images for a left eye and those for a right eye [2]. Although the images themselves give just a bi-dimensional (x, y) representation, we can get the third dimension (depth z) from left and right images. It is because there is a disparity between left and right images and men or women can feel the depth from images that have proper disparities.

Therefore, when we generate stereograms, we have to solve the following problems:

1. how to extract foreground objects.
2. how to composite foreground object images and background scenery images.

As to the former problem about the extraction of the foreground objects, there are a lot of studies about image segmentation [3, 4], e.g. pixel-based segmentation, area-based one, edge-based one, and physics-based one. For example, the snake [5] is a popular technique based on the edge detection. However, the automatic methods for the arbitrary background have not been developed to a practical level. Additionally, robust methods are needed especially for a live program on TV [6, 7].

Therefore, an image composition method called chromakey that can segment human objects from a uniform color background and superimpose them with another background electronically has been proposed. This technique has been used for many years in the TV and the film industry.

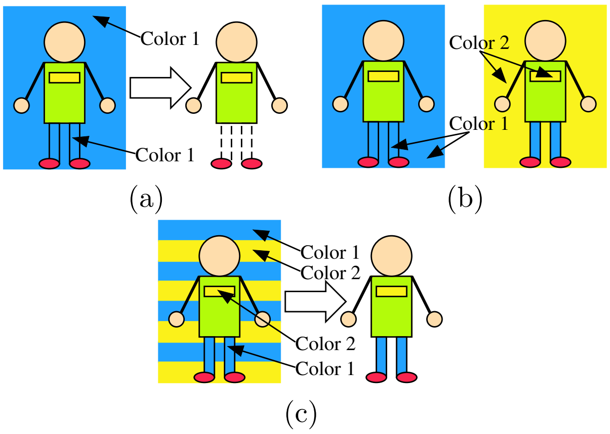

Figure 1: Region extraction with chromakey techniques using several types of the backgrounds. (a) Unicolor background. (b) Multi background [8]. (c) Striped background (Our proposed method).

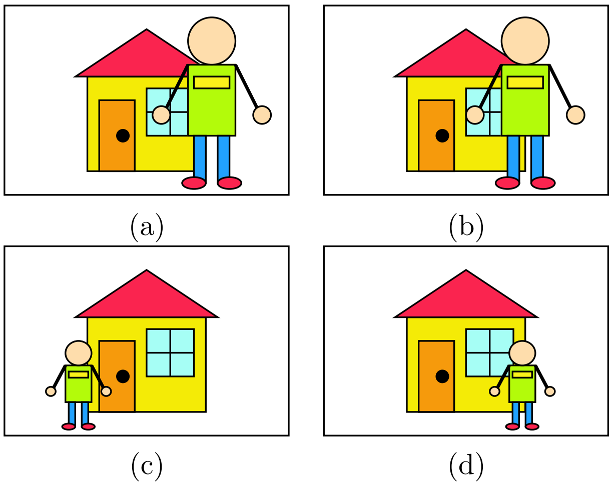

Figure 2: Image composition of foreground objects and background scenery. (a) Left image with size strangeness. (b) Right image with size strangeness. (c) Left image with disparity strangeness. (d) Right image with disparity strangeness.

The conventional chromakey technique uses background whose color is solid blue or green, and extracts the region in images whose colors are not same with the background. Although this can be successfully used, several problems associated with this method still need a solution.

One of the most critical and unsolved problems is that if one's clothes and the background have the same color, the former is regarded as the latter (Figure 1(a)). Smith and Blinn propose the blue screen matting method that allows the foreground object to be shot against two backing colors [8] (Figure 1(b)). This method can extract the foreground region whose colors are same with the background color. However, this multi-background technique cannot be used for live actors or moving objects because of the requirement for repeatability.

In this paper, we propose a new method that makes it possible to segment objects from a background precisely even if objects have the same color as the background, by using two-tone striped background (Figure 1(c)). In our method, the adjacency condition between two-tone striped areas on the background is utilized to extract the foreground regions whose colors are same with the background [9].

As to the latter problem about the image composition of foreground object images and background scenery images, the geometric constraints between them must be considered. The relationships between foreground objects and background objects in composite images must be consistent. When the size of the foreground object and that of the background object is not consistent, the composite images looks unnatural (Figure 2(a)(b)). The relationships of objects between the left image and the right one must be also consistent. When the disparity is incorrect, the composite left and right images looks strange (Figure 2(c)(d)).

Therefore, the geometric constraints are considered by utilizing the relationship between characteristic points between left images and right images, and between the foreground images and background images, respectively.

2 Stereogram Generation

The color $I(u,v)$ of a compositing image at a pixel (u,v) is defined as:

$$\begin{aligned} I(u,v) &= \alpha(u,v)F(u,v) \\ &\quad +(1-\alpha(u,v))B(u,v), \qquad (1) \end{aligned}$$

where $F(u,v)$ and $B(u,v)$ are the foreground and the background color, respectively, and $\alpha(u,v)$ is the alpha key value at a pixel (u,v) [1].

The color at a pixel (u,v) is same with the foreground when $\alpha(u,v)$ equals to 1, and that is same with the background when $\alpha(u, v)$ equals to 0. In other words, the region extraction with a chromakey technique is the problem of the alpha key extraction at each pixel (u, v).

The alpha key value is ordinary estimated by a color space approach , e.g. [10] and an image space approach, e.g. [5]. However, the foreground cannot be extracted by only using these two approaches when the colors of foreground objects are same with that of the background. The same is true for the two-tone striped backgrounds that we propose. Therefore, the boundaries where the color of the striped background changes are utilized for detecting the foreground region whose color is same as the background [9].

In our method, a baseline stereo camera that consists of a left camera and a right camera is used for taking the foreground left and right images and the background left and right images, respectively.

The procedure of our proposed method consists of four steps: 1) background color extraction, 2) striped region extraction, 3) foreground extraction, and 4) image composition.

2.1 Background Color Extraction

Candidate regions of the background are extracted by using a color space approach. Let C_1 and C_2 be the colors of the two-tone background in acquired images from the camera, and R_1 and R_2 be the region whose color is C_1 and C_2, respectively. R_i $(i = 1, 2)$ is defined as:

$$R_i = \{(u, v)|F(u, v) \in C_i\}, \quad (2)$$

where $F(u, v)$ is the color of an image at a pixel (u, v).

After the human operator decide C_1 and C_2 by the trial-and-error method, R_1 and R_2 are extracted, respectively (Figure 3(a)(b)).

2.2 Striped Region Extraction

The background mainly consists of the three regions: R_1, R_2, and R_3. R_3 is the intermediate region between R_1 and R_2 (Figure 3(b)). It is difficult to extract R_3 by a color space approach because the color of R_3 may be a composite of C_1 and C_2 and not either C_1 or C_2. Therefore, R_3 is extracted from adjacency conditions in our method.

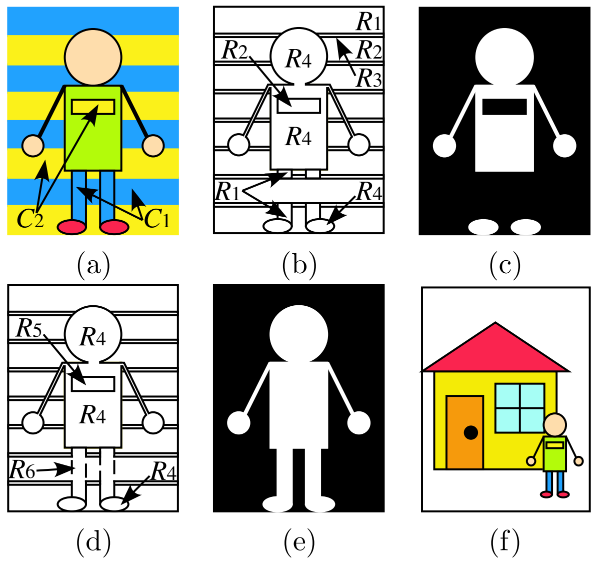

Figure 3: Overview of stereogram generation. (a) Original image. (b) Region segmentation without using the information of the striped area. (c) Foreground extraction without using the information of the striped area. (d) Region segmentation by using the information of the striped area. (e) Foreground extraction by using the information of the striped area. (f) Image composition of left and right images.

The difference between R_3 and the foreground is that R_3 contacts with both R_1 and R_2. In the case of horizontal stripe, the color of the upper and lower region of R_3 differs from each other. Therefore, R_3 is extracted by searching the area between R_1 and R_2 in the longitudinal direction.

After searching R_3, the foreground region R_4 is determined as follows:

$$R_4 = \{(u, v) \quad | \quad F(u, v) \notin (C_1 \cup C_2), \\ (u, v) \notin R_3\}. \quad (3)$$

In this procedure, all regions whose colors are same as the background are extracted as the background. An example of extraction error is shown in Figure 3(c). They must be corrected in the next step.

2.3 Foreground Extraction

The boundary between the foreground and the background is detected to recheck the foreground region whose color is same as the background, and the background region whose color is same as the

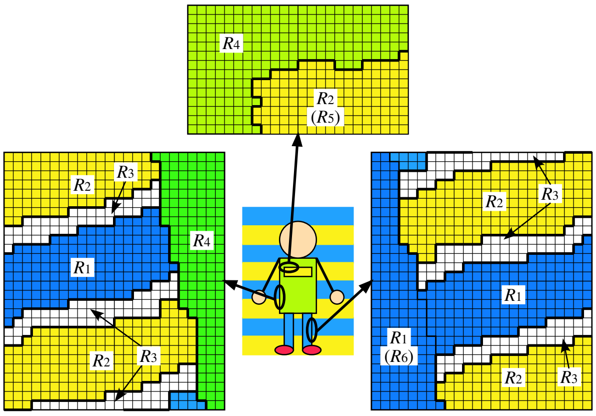

Figure 4: Boundary detection of foreground.

foreground. Let R_5 be the region whose color is same as the background and is inside the foreground, and R_6 be on the surface of the foreground (Figure 3(d)). To detect R_5 and R_6, the adjacency conditions with R_3 are utilized (Figure 4).

The difference between $R_{1,2}$ and R_5 is whether it contacts with R_3 or not, while the color of $R_{1,2}$ and R_5 is same. This is because there exists no two-tone striped region whose colors are same as the background inside the foreground objects. Therefore, the regions that do not contact with R_3 among R_1 and R_2 can be judged as R_5.

To determine R_6 region, endpoints of R_3 are utilized. R_3 regions become horizontal lines if the stripes and the scanlines of the camera are perfectly horizontal. However, this does not always occur. Therefore, the approximate line of each $R_{3,j}$ region is calculated. After that, the endpoints of the approximate lines are detected. The neighbor areas of the endpoints that are adjacent to R_6 have same color as C_1 or C_2, although the foreground next to the endpoints has different color from C_1 or C_2. After judging the endpoints next to R_1 or R_2, these points are connected smoothly with interpolation lines. To detect the boundary of R_6, each pixel which is on the interpolation line is regarded as R_4. As the result of these procedure, the region that is surrounded by the interpolation lines and R_4 becomes the same region as R_5 which do not contact with R_3. Therefore, R_6 can be obtained by the same way of detecting R_5.

On the other hand, the endpoints that do not contact with R_1 and R_2 are not connected with interpolation lines, and the boundaries of R_4 remain precisely. In this way, the distinction between R_4 region and R_6 region can be easily executed.

After that, the image is divided into six regions (Figure 3(d)) and the extraction procedure finishes. Accordingly, the foreground region of the image R_{fg} is obtained as follows (Figure 3(e)):

$$R_{fg} = \{(u,v)|(u,v) \in (R_4 \cup R_5 \cup R_6)\}. \qquad (4)$$

2.4 Image Composition

The conditions when the images are taken by the baseline stereo camera are estimated from the images to generate stereograms naturally. The optical axes of two cameras are parallel. Therefore, the 3-D coordinate of corresponding points between two images in the camera coordinate (X,Y,Z) is expressed as follows:

$$X = \frac{b(x_{i,l}+x_{i,r})}{2d}, \qquad (5)$$

$$Y = \frac{b(y_{i,l}+y_{i,r})}{2d}, \qquad (6)$$

$$Z = \frac{bf}{d_i}, \qquad (7)$$

where $(x_{i,l}, y_{i,l})$ and $(x_{i,r}, y_{i,r})$ is the coordinate value of the characteristic point i on the left and right images, respectively, f is the image distance, b is the length of baseline (the distance between two cameras), and $d = x_{i,l} - x_{i,r}$ is the disparity of the corresponding point i.

When compositing natural images that have proper disparities, the parameters f and b must be estimated. However, it is very difficult to estimate these parameters with no information of images. Therefore, the human operator gives information about the distance between two characteristic points in left and right images in our method. Two characteristic points are selected in consideration of the ease for the estimation from the general knowledge, e.g, body heights of human, the height of house, and so on.

f and b can be calculated in the following equation when more than three corresponding points are selected and the distance between these points are estimated by the human operator.

$$b^2\Big\{\Big(\frac{x_{i,l}+x_{i,r}}{2d_i} - \frac{x_{j,l}+x_{j,r}}{2d_j}\Big)^2 + \Big(\frac{y_{i,l}+y_{i,r}}{2d_i} - \frac{y_{j,l}+y_{j,r}}{2d_j}\Big)^2 + f^2\Big(\frac{1}{d_i} - \frac{1}{d_j}\Big)^2\Big\} = L_{i,j}^2, \qquad (8)$$

where $L_{i,j}$ is the estimated distance between two points i and j in the world coordinate.

After estimating b and f, the 3-D reconstruction of the scenery in images can be executed. Therefore, the foreground objects and the background scenery can be merged naturally by considering the geometric constraints between left and right images, and the foreground and background images, respectively.

3 Experiments

In our experiment, the colors of the background were set blue and yellow. In Figure 5, a sheet of blue paper that is same paper as the background is rolled around the left arm of the foreground person and a sheet of yellow paper is put on his shirt to examine the case of the same color clothes (Figure 5(a)). The result of the background color extraction and the detection of the endpoints of R_3 are shown in Figure 5(b) and (c), respectively. The extracted foreground is shown in Figure 5(d). This result shows that the foreground regions whose colors are same with the background are extracted without fail. The result of the image composition of the extracted foreground and another background is shown in Figure 5(e).

Figure 6 shows the results of automatically generated stereograms. Figure 6(a)(b) show original left and right background images. After estimating the parameters when the background images are taken, horizontal planes of floor can be calculated (Figure 6(c)(d)). The results of image composition (Figure 6(e)(f)) look natural. The effectiveness of our method has been verified by comparing the images in which a man actually stood on the same place as the composite images (Figure 6(g)(h)). The composite left and right images are very similar to the real left and right ones, respectively.

From these experimental results, the effectiveness of our proposed method has been verified.

4 Conclusions

In this paper, we propose a new method that can generate stereograms automatically with a chromakey technique using two-tone striped backgrounds. We utilize the adjacency condition between two-tone striped areas on the background, and extract the foreground regions whose colors are same with the background.

The geometric constraints are considered by utilizing the relationship between characteristic points between left images and right images, and between the foreground images and background images, respectively.

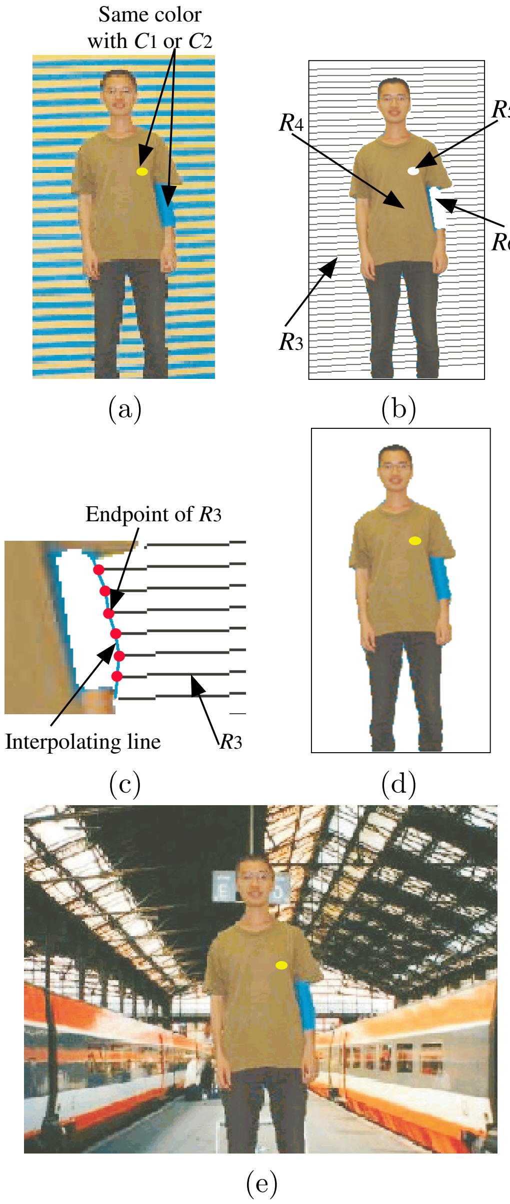

Figure 5: Result of region extraction II. (a) Original image. (b) Background color extraction. (c) Detection of R_6. (d) Foreground extraction. (e) Image composition.

Experimental results show that the foreground region is constantly extracted from the background despite of the colors of the foreground objects, and the stereo pair images after composition are very natural that have proper disparities.

As the future work, we have to improve the processing speed to treat with moving picture in real-time.

Acknowledgement

This work was partly funded from Hoso-Bunka Foundation, Inc., Japan.

References

[1] Thomas Porter and Tom Duff: "Compositing Digital Images," *Computer Graphics* (*Proceedings of SIGGRAPH84*), Vol.18, No.3, pp.253–259, (1984).

[2] Richard I. Land and Ivan E. Sutherland: "Real-Time, Color, Stereo, Computer Displays," *Applied Optics*, Vol.8, No.3, pp.721–723, (1969).

[3] King-Sun Fu and Jack Kin Yee Mui: "A Survey on Image Segmentation," *Pattern Recognition*, Vol.13, pp.3–16, (1981).

[4] Wladyslaw Skarbek and Andreas Koschan: "Colour Image Segmentation - A Survey," *Technical Report* 94–32, *Technical University of Berlin*, *Department of Computer Science*, (1994).

[5] Michael Kass, Andrew Witkin and Demetri Terzopoulos: "SNAKES: Active Contour Models," *International Journal of Computer Vision*, Vol.1, pp.321–331, (1988).

[6] Simon Gibbs, Constantin Arapis, Christian Breiteneder, Vali Lalioti, Sina Mostafawy and Josef Speier: "Virtual Studios: An Overview," *IEEE Multimedia*, Vol.5, No.1, pp.18–35, (1998).

[7] Andrew Wojdala: "Challenges of Virtual Set Technology," *IEEE Multimedia*, Vol.5, No.1, pp.50–57, (1998).

[8] Alvy Ray Smith and James F. Blinn: "Blue Screen Matting," *Computer Graphics* (*Proceedings of SIGGRAPH96*), pp.259–268, (1996).

[9] Atsushi Yamashita, Toru Kaneko, Shinya Matsushita and Kenjiro T. Miura: "Region Extraction with Chromakey Using Striped Backgrounds," *Proceedings of IAPR Workshop on Machine Vision Applications (MVA2002)*, pp.455–458, (2002).

[10] Yasushi Mishima: "A Software Chromakeyer Using Polyhedric Slice," *Proceedings of NICOGRAPH1992*, pp.44–52, (1992).

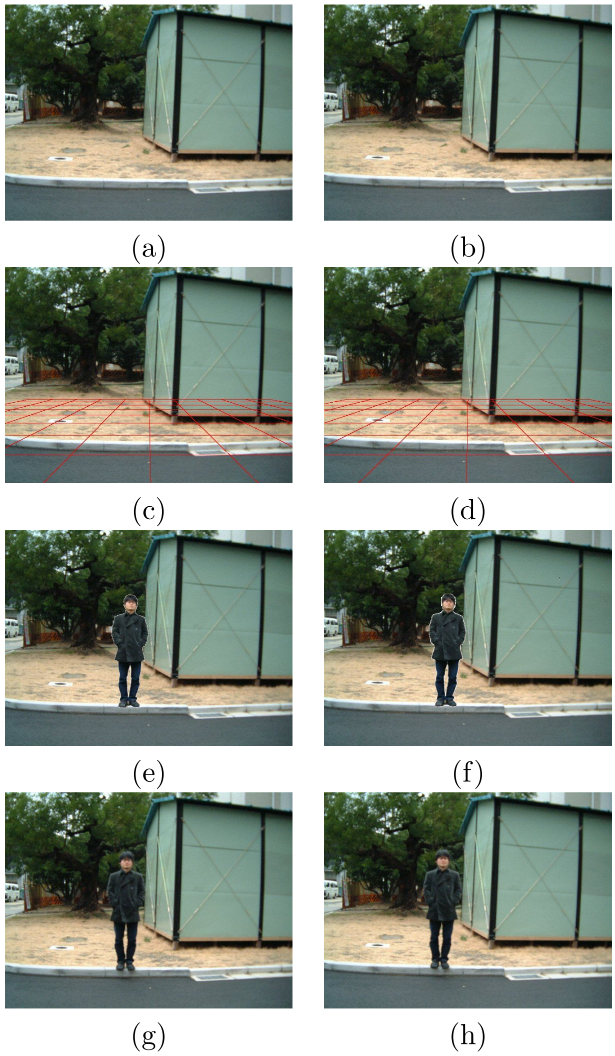

Figure 6: Result of stereogram. (a) Left background image. (b) Right background image. (c) Calculated horizontal plane of floor on the left image. (d) Calculated horizontal plane of floor on the right image. (e) Result of left image composition. (f) Result of right image composition. (g) Actual left image. (h) Actual right image.