Target Object Image Extraction from 3D Space using Stereo Cameras

Chae-Gon Yoo ,Chang-Sung Jung, and Chi-Jung Hwang
Tel. +82-42-866-0391, Fax.: +82-42-866-0457
Department of Computer Science
Chungnam National Univ
Kung-dong, Daejeon, Korea
e-mail : cgyoo@mail.ddc.ac.kr, csjung@ipl.cnu.ac.kr, cjhwang@ipl.cnu.ac.kr

Abstract: Stereo matching technique is used in many practical fields like satellite image analysis and computer vision. In this paper, we suggest a method to extract a target object image from a complicated background. For example, human face image can be extracted from random background. This method can be applied to computer vision such as security system, dressing simulation by use of extracted human face, 3D modeling, and security system. Many researches about stereo matching have been performed. Conventional approaches can be categorized into area-based and feature-based method. In this paper, we start from area-based method and apply area tracking using scanning window. Coarse depth information is used for area merging process using area searching data. Finally, we produce a target object image.

1. Introduction

Stereo matching techniques are widely used in computer vision area. In area-based research, area comparison methods have been proposed, and, in feature-based method, features like edge and corner are used as a reference point between stereo pair images. Both approaches still have broblems like poor texture data, repetitive pattern, noises, size of window and occlusion. Due to these problem, image extraction of complicated object in 3-D space has difficulties, while images of simple objects such as building and tower are relatively easy to extract because of their regularity of shape.

In this paper, we perform area-based pair image matching to acquire coarse depth distribution and area merging in each image to find object area which we are interested in. Similarity curves are produced using variable size window. Similarity curves have information about boundary property of pixels of object image. The proposed method can be applied to computer vision such as security system, dressing simulation by use of extracted human face, 3D modeling, and security system not using blue screen background.

In chapter 2, related previous works will be reviewed, in chapter 3, the proposed method will be described, and, in chapter 4, experimental result will be showed. We will conclude and comment future research.

2. Previous Researches

Most conventional approaches are categorized into feature-based and area-based methods. In area-based method , occlusion detection is tried to find disparity space[1]. Stereo aerial images are segmented into two different kinds of components, digital elevation model and digital terrain model in some research[2]. A single mobile camera has been used to perform 3D reconstruction of unknown static scenes and very precise reconstruction was processed at nearly video rate[3]. Photometric and geometric invariants are used to recognize objects[4]. A novel analytically-based face recognition system is presented[5]. 3D modeling and view point synthesis[6].

3. Extraction Processing

Front part of whole processing uses coarse area matching using scanning window. Intermediate depth of coarsely matched areas is calculated. Intermediate depth information is used for merging area based on specified depth threshold. Next process is focusing on determining correct boundary of the object area selected at previous step. To find pixels near to boundary of object, we calculate similarity curve for each pixel or pixel group by varying the size of comparing window. The pixels regarded as boundary part is post-processed with selected area. Finally, a pure object image can be clipped from background. Even though the background may be complicated random scene, the proposed method is not affected by the property of it. Figure 3-1. describes the system flow.

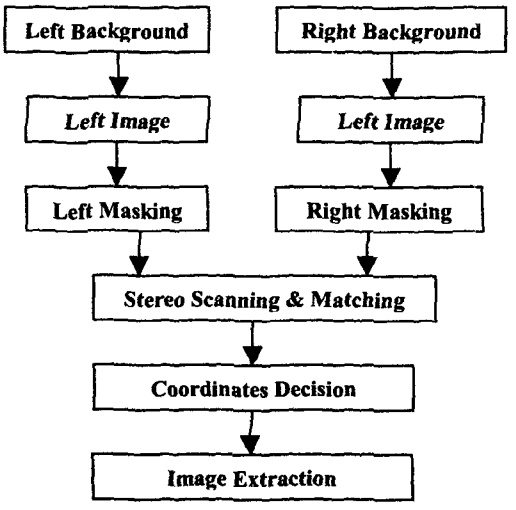
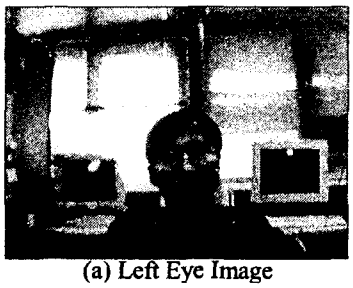


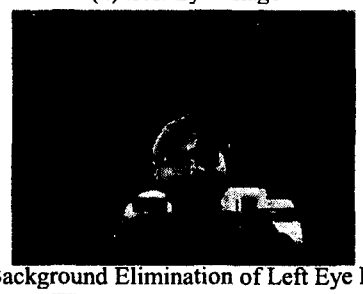
Figure 3.1 System Flow

3.1 Background Coloring

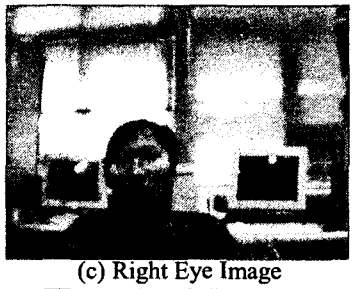
In this paper, we assume that background images of left and right images are acquired in advnace. We fill background field of left image with black color and that of right image with white color by use of background mask subtraction. This process is displayed in fiture 3.2.



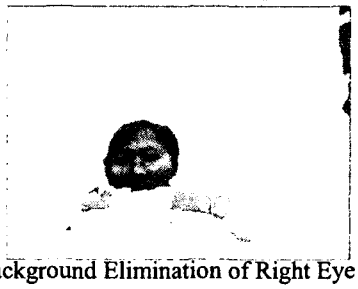
(a) Left Eye Image



(b) Background Elimination of Left Eye Image



(c) Right Eye Image



(d) Background Elimination of Right Eye Image

Figure 3.2 Background Coloring

This process produces larger error, when the background area is included in scanning window. This means goals of this step are background removing and efficient stereo matching. Even though images (b) and (d) of figure 3.2 show some parts of background are not filled accurately, these results are very important data on stereo matching step and those incompletely colored parts will be eliminated in the following steps. This process can be expressed as the followings.

$$\sum_{i=0}^{n}\sum_{j=0}^{n} \text{if}\,(\,L_{i,j} - B_{i,j} = 0\,)\quad L_{i,j} = Black \quad (1)$$

$$\sum_{i=0}^{n}\sum_{j=0}^{n} \text{if}\,(\,R_{i,j} - B_{i,j} = 0\,)\quad R_{i,j} = White$$

where L is a left eye image,
R is a Right eye image,
B is a Background image, and
i and j are indexes.

3.2 Similarity-Curve

Now, as we know rough area which has an object image we want to extract, we need to determine boundary pixels accurately. The property of pixel at inside of object is that it has high similarity to the corresponding pixel of the other pair image with a relatively large comparing window. On the contrary, the similarity of the pixels on boundary of object decreases abruptly from when the size of comparing window reaches a threshold. Similarity curve value of pixel x,y at window i is determined by formula (2).

$$C_{xyi} = \sqrt{\int_{x-i}^{x+i}\int_{y-i}^{y+i} (li(m,n) - ri(m,n+w))^2\,dm\,dn} \quad (2)$$

where
x, y is coordinates of pixel,
i is window size,
li is intensity of pixel in left eye image,
ri is intensity of pixel in right eye image,
and w is offset of x of target object in both eye images.

w of formula (2) may be assumed preliminarily or acquired at the area based matching step. In the latter case, w should be calculated at the lower level of pyramid

3.3 Object Image Extraction

At this step, the target object image is torn from it background. We can use focal length of cameras, base line between two cameras as input data of the proposed system. Scanning windows in both images try to decide the similar areas of the two images. The selected area which is acquired at area-based matching step is obvious target object area and the pixels which have relatively higher curve similarity are also considered as object area. By this criteria, we can separate the interested object from the complicated random background not using blue screen. Fiture 3.3 shows the applicatoin of the proposed system.

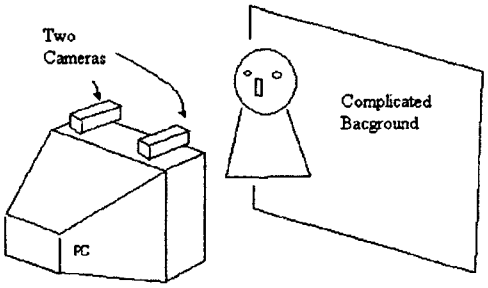


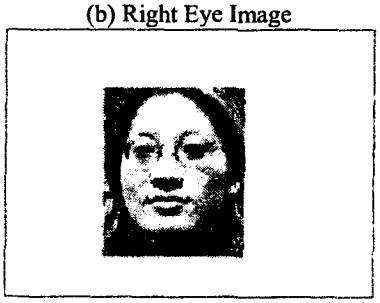
Figure 3.3 Stereo Image Acquisition

4. Experimental Results

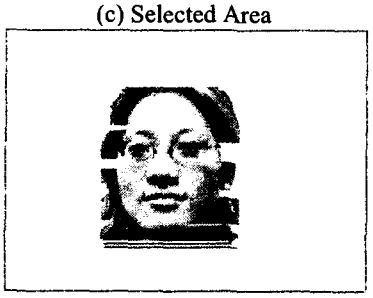
We used Pentium IV 1.4G Mhz system with 256 RAM for implementing. The size of each stereo pair image is 640 x 480. Image (a) and (b) of Figure 4-1 are left and right image of stereo pair. Our target object area in right eye image is image (c) of figure 4.1

(a) Left Eye Image

(b) Right Eye Image



(c) Selected Area



(d) Final Result

Figure 4-1 Matching Results

At area-based matching step, we can get a rough area which has pixels of similar x coordinates offset between stereo pair images. At this step, some pixels may be mis-calculated due to noise, intensity difference, or repetitive pattern. This kind of pixels are tested using relative distance to the pixels whose position are determined correctly and merged to appropriate area.

We need to calculate similarity curve values for each pixels which are located at boundary area. The graph in figure 4-2 shows sample similarity curve of a pixel at the boundary. This graph has MSE(Minimum Square Error) values. In the case of boundary pixel, error increasing rate was much higher than that of inner pixel. Smaller threshold can produce more accurate boundary pixels, however, noise can affect more. On the contrary, bigger threshold will be affected less by noise, however, boundary will be blurred more.

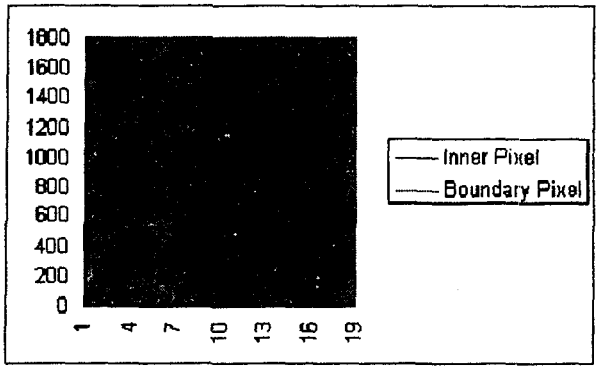


Figure 4-2 Similarity Curve

5. Conclusion and Future Research

The proposed method showed its separability of the interested object image from complicated background. This method can be applied to computer vision areas such as target recognition, dressing simulation, and security system. In the future, we are planning to combine feature-based method to the proposed method and refine bacground elimination step.

References

[1] Aaron F. Bobick, "large Occlusion Stereo," International Journal of Computer Vision, 33(3), pp 181-200, 1999

[2] C. Baillard, H. Maitre, "3-D Reconstruction of Urban Scenes from Aerial Stereo Imagery: A Focusing Strategy," Computer Vision and Image Understanding Vol. 76, No. 3, December, pp. 244-258, 1999

[3] Eric Marchand, Francois Chaumette, "An Autonomous Active Vision System for Complete and Accurate 3D Scene Reconstruction," International Journal of Computer Vision 32(3), 171-194 (1999)

[4] Kenji Nagao, "Using Photometric Invariants for 3D Object Recognition,"Computer Vision and Image Unserstanding, Vol. 71, No. 1, July, pp74-93, 1998

[5] Ali Reza Mirhosseini and Hong Yan, "Human Face Image Recognition: An Evidence Aggregation Approach,"Computer Vision and Image Understanding, Vol. 71, No. 2, August, pp 213-230, 1998

[6] Ebroul Izquierdo M. and Silko Kruse, "Image Analysis for 3D Modeling, Rendering, and Virtual View Generation," Computer Vision and Image Understanding Vol. 71, No. 2, August, pp. 231-253, 1998.