Generation of Spherical Image Mosaic Using 3D Information Obtained from Stereo Omni-directional System (SOS)

Hideki Tanahashi† Caihua Wang† Yoshinori Niwa† Kazuhiko Yamamoto‡
†Office of Regional Intensive Research Project, Softopia Japan, JST
Kagano 4–1–7, Ogaki City, Gifu 503-8569, JAPAN
{tana, c-wang, niwa}@softopia.pref.gifu.jp
‡ Department of Information Science, Faculty of Engineering, Gifu University
Yanagido 1–1, Gifu City, Gifu 501-1193, JAPAN
yamamoto@info.gifu-u.ac.jp

Abstract

In general, it is difficult in a system that acquires omni-directional image using multiple cameras to generate a single center of projection image as each camera center differs from the system center. We have developed a system called Stereo Omni-directional System (SOS). In this system, the stereo units, which are composed of three cameras, are arranged on each plane of regular icosahedrons. This system can acquire all directional color and stereo pair images, in real time. The images acquired from the SOS are images with a different viewpoint. Hence, by using three-dimensional information that are calculated from these stereo pair images, it is possible to generate images as seen from the center of the SOS. A single center of projection image can be generated by integrating them. In this paper, we propose a method of generating a three-dimensional spherical image mosaic from a single viewpoint using the SOS.

1. Introduction

In recent years, due to the availability of high function and low cost PCs and image sensors, the research to acquire the image of a spacious dynamic real environment using multiple cameras have been proposed. The high-resolution image of an environment can be acquired by using multiple cameras. Such developments have invented a new variety of applications in surveillance and monitoring, robotics, visualization and virtual reality.

Generally, there are two methods to acquire the image and three-dimensional information of a wide and dynamic real environment using multiple cameras. The first method is to acquire the image of an environment by viewing the inside from the outside viewpoint. This method acquires information by arranging the multiple cameras in surroundings of the environment. Kanede et. al [8] digitized whole real world events including three-dimensional shape information of a dynamic environment. Matsuyama [10] proposed a system which generated three dimension image of human's action, such as dance, physical exercise and dactylology. Nebel and Cockshott [6] proposed a dynamic 3D whole body scanner that is arranged eight pods composed of three cameras at the corners of a parallelepiped. These methods reconstruct the three-dimensional model using the images obtained from multiple cameras and it is necessary to arrange the cameras accurately.

The second method acquires the image of an environment by viewing the outside from a center viewpoint. This method arranges multiple cameras at the center of the environment in a different direction. Moving the system becomes possible by accurately arranging multiple cameras at once and the acquisition of wide environmental information becomes possible. One of the important characteristics for imaging system using this method is a single viewpoint property [13]. The group of Bell Lab. AT&T [11] proposed an omni-directional viewer that is composed of four cameras and four triangular mirrors. Kawanishi et. al [9] proposed a stereo omni-directional imaging sensor that is composed of twelve cameras and two hexagonal pyramidal mirrors. These systems can acquire high-resolution 360 deg. panoramic images of dynamic environment with single viewpoint constraint. However, in order to satisfy this constraint, these systems have used planar mirror. So the systems are large and there are some blind spot caused by the reflection of camera, and the position of the sensor and the object. On the other hands, Kimuro et. al [12]

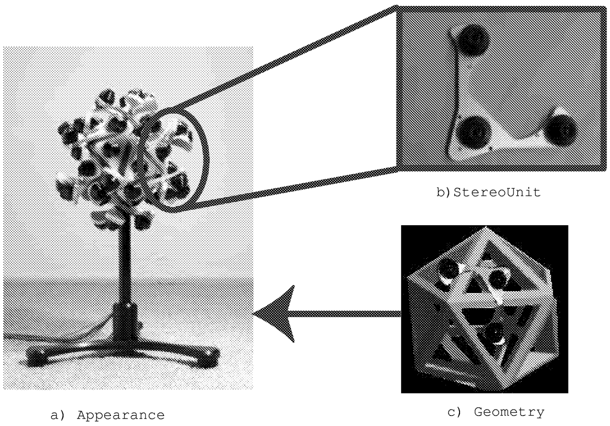


Figure 1. Stereo Omni-directional System (SOS)

proposed an omni-directional camera system that reduced blind spot. They used four cameras equipped with fish-eye-lens. Three of these cameras are fixed horizontally in equi-angular distance, and the other is fixed to upright direction. This system can get omni-directional view at once except for right down direction. Fermuller et. al [2] proposed a system which sampled the whole visual field with high resolution. Their system consist of six cameras looking in different directions. However, because each camera center differs from the system center, a spherical image mosaic generated from the images does not satisfy single viewpoint constraint. Therefore it is necessary to acquire the images of the dynamic environment in all direction without blind spot with single viewpoint constraint.

In this paper, we proposed a new method to generate a spherical image mosaic satisfied single viewpoint constraint using Stereo Omni-directional System (SOS) which we developed [5]. This system can acquire color and stereo images of all direction from the system, in real time by using a component constructed of sixty cameras. The cameras are separated into twenty stereo units. Each unit is composed of three cameras, which are arranged on each plane of regular icosahedrons. In this system, each camera center differs from the system center and a spherical image mosaic generated from the images does not satisfy single viewpoint constraint. However, this system acquires not only color image but also stereo image pairs in each direction. So, using the three-dimensional information calculated from these stereo image pairs, a spherical image mosaic satisfied single viewpoint constraint can be generated.

2. Stereo Omni-directional System (SOS)

Figure 1 shows a prototype of the Stereo Omni-directional System (SOS). The system is capable of obtaining color images and depth maps in all directions by using a component constructed of sixty cameras. All cameras of this system operate concomitantly, and all images can be obtained at the same time. The cameras are separated into twenty stereo units. Each stereo unit is composed of three cameras that are arranged to form two stereo pairs, top/bottom and left/right as shown in Figure 1 b). By arranging one standard camera vertically and horizontally, the three cameras have two baselines. In each stereo image pair, the epipolar lines are either vertical or horizontal in position. Therefore, if a match cannot be obtained by one of the stereo pairs, it is possible to obtain it with another pair. As a result, it is possible to obtain a larger amount of 3-D information using the three-cameras system in comparison to any other two-cameras system. In addition, the three-cameras system can obtain more accurate and reliable depth values from these matches. The stereo camera calibration, which becomes a problem when the depth values are generated, has been done individually with each unit. In this system, the stereo units, which have these characteristics, are arranged on each plane of regular icosahedrons (Figure 1 c). Using any regular polyhedrons, it is possible to evenly sample 3-D space while maintaining the same resolution. To acquire the high-resolution images, we chose regular icosahedrons. Table 1 shows specifications of a stereo unit.

Table 1. Specifications of SOS

Image Sensor	1/3’ CMOS Color Image Sensor
Effective Resolution	640 (H) *480 (V)
Focal Length	2.9 (mm)
Feild of View	96.6 deg (H) * 71.9 deg (V)
Baseline Length	90mm

In the stereo processing, the rang precision influences the baseline length. Placing too much emphasis on maintaining the baseline length within each face of a regular icosahedrons will increase the size of the overall system. As a result, the system will be unable to see objects close to it. Therefore, each stereo unit is arranged so that it does not obstruct the view of other units. As a result, we succeeded in reducing the size of the entire system (diameter: 27cm; weight: 4.5 kg) securing the base line length (90 mm) to generate 3-D information with enough precision. In these conditions, the depth resolution became 15cm at 2m given an image resolution 640x480. This system can acquire a space over 400mm with an overlap of image from the center.

From each unit, a color image and two monochrome images can be obtained. As a result, twenty color images and forty monochrome images are obtained. These images are stored in a memory unit, and each pair of stereo images is translated to a PC where stereo image processing has been performed. The color image and the depth map of each direction are acquired at 15 frames per second. The SOS has the following characteristics:

1. An all-directional view from an observation point can be acquired at the same time.
2. Information can be acquired in real time.
3. Stereo images can be acquired.
4. There is no difference in resolution depending on the direction.
5. A high-resolution evaluation image can be acquired and the acquired range precision is high.

In the following sections, we describe the method to generate the spherical image mosaic using the SOS.

3. Spherical Image Mosaic

As described in the above section, color images and stereo pair images of a dynamic environment in all direction are acquired from the SOS via its twenty stereo units at the same time. The three-dimensional reconstruction of the dynamic environment can be carried out using the stereo matching results of each stereo pair images [4]. However, these amounts of data are huge and all data cannot display not to use the special display system, which are surrounded with many screens arranged in all direction, at the same time. We select to build a more simple representation for the images in all direction. A spherical image mosaic representation is the most suitable representation for a dynamic environment in all direction. We have proposed a method of generation of a spherical imagemosaic by which each image was cut into triangular regions and integrated them into an image mosaic which corresponds to polar expansion of icosahedrons [1]. This method can generate the spherical image mosaic at high speed by using only color images in all directions. So, this method is suitable for live event viewing of a dynamic environment [3]. However, as the size of each triangular region is decided by the constant depth value of the scene which we input, lack areas are caused in objects which are nearer than the depth given and overlap areas are caused in object which are far.

In this paper, we generate an images as seen from the center of the SOS using color images and three-dimensional information obtained from the SOS and finally generate spherical image mosaic.

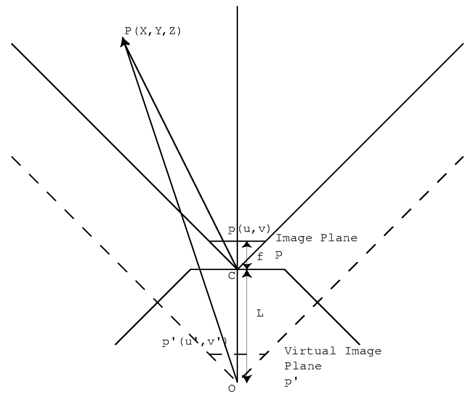


Figure 2. Relation between an acquired image and an image seen from a center of SOS

3.1. Generation of image as seen from the center of the SOS

Figure 2 shows the relation between an image obtained from a stereo unit and an image seen from the center of the SOS. O is the center of the SOS and C is the camera center. A three-dimensional point $P(X, Y, Z)$ is projected on the image plane of an actual camera and is represented on the virtual image plane as seen from the center of the SOS as follows:

$$\begin{pmatrix} u - u_0 \\ v - v_0 \end{pmatrix} = \frac{f}{Z} \begin{pmatrix} X \\ Y \end{pmatrix} \quad (1)$$

$$\begin{pmatrix} u' - u_0 \\ v' - v_0 \end{pmatrix} = \frac{f}{Z + L} \begin{pmatrix} X \\ Y \end{pmatrix} \quad (2)$$

where u, v, u', v' take between -0.5 to 0.5. u_0, v_0 are image centers. f is the focus length and L is the distance from the center of the SOS to the center of plane composed a regular icosahedrons. Using the equations 1 and 2, the pixel position $p(u, v)$ on the actual image plane is shifted to the pixel position $p'(u', v')$ on the virtual image plane as follows:

$$\begin{pmatrix} u' - u_0 \\ v' - v_0 \end{pmatrix} = \frac{Z}{Z + L} \begin{pmatrix} u - u_0 \\ v - v_0 \end{pmatrix} \quad (3)$$

The pixel position on the image shifts toward the image center according to the three-dimensional distance to the object and the two-dimensional distance to the center of image as described. The point on the image does not shift if the distance (Z) to the object is larger than the distance (L) from the center of the SOS to the center of plane. Figure 3 shows the relation between distance and shift of pixel. Each line is shows the pixel position from the image center. $L = 0.1(m)$ according to design of the SOS. For example, when the pixel position from the image center is $100(pixel)$ and the distance to the object is $9.5(m)$, the pixel will shift

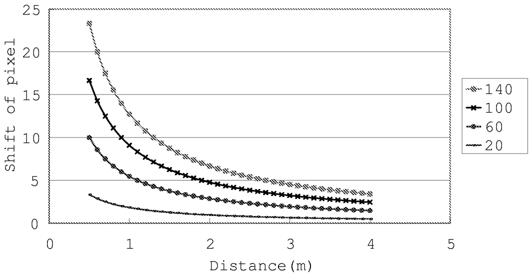


Figure 3. Relation between distance and shift of pixel

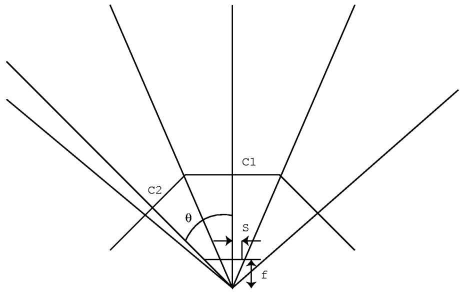
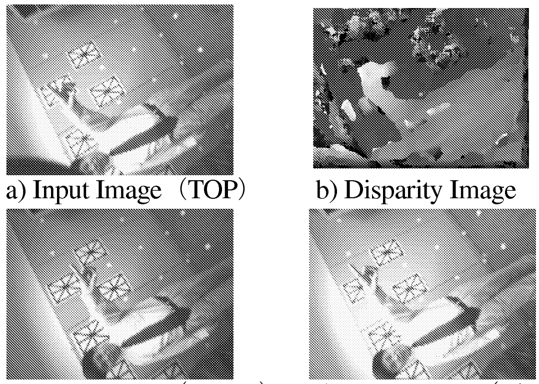


Figure 4. Geometry of center of icosahedrons



a) Input Image (TOP)　b) Disparity Image

c) Input Image (Center)　d) Input Image (Right)

Figure 5. Example images from a stereo unit

within 1 pixel. However when the distance to the object is greatly different by the adjoined pixel, the occlusion area is caused. The occlusion area is large when the pixel position is far from the image center and the object is near to the system. The occlusion area is represented as follows:

$$\begin{pmatrix} u'_2 - u'_1 \\ v'_2 - v'_1 \end{pmatrix} = \frac{Z_2}{Z_2 + L}\begin{pmatrix} u_2 - u_0 \\ v_2 - v_0 \end{pmatrix} - \frac{Z_1}{Z_1 + L}\begin{pmatrix} u_1 - u_0 \\ v_1 - v_0 \end{pmatrix} \quad (4)$$

In this paper, the occlusion area is filled with the color of the pixel with larger distance.

3.2. Generation of spherical image mosaic

As described in the previous subsection, images in all directions as seen from the center of the SOS are obtained by each image. We can cut out triangular region from these image and generate a spherical image mosaic by integrating them. Because each camera is calibrated and the geometries of the SOS are known, the center and the orientation of each triangular region in the image of each unit can be determined directly. The size of each trianglar region can be calculated as follows:

$$S = f\tan(\frac{\theta}{2}) \quad (5)$$

where f is the focus length of each camera, and θ is the radial angle between the normal vectors of two neighboring plane on the icosahedrons, and we know that $\theta = 42.1(degrees)$ from the geometry of an icosahedrons. Figure 4 shows the geometries for the computation involved in the equation 5.

4. Experiments

The experiment apparatus was a commercial PC (2CPUs Pentium III, 800MHz CPU) for capturing all-direction images and generating a spherical image mosaic. Two stereo pair images are acquired from the SOS in each direction. These input images are rectified to compensate for lens distortions. Basically the rows of images digitized from horizontally displaced cameras are aligned, and similarly the columns of images obtained from vertically displaced cameras are aligned. After that, each pixel in the reference image is compared with pixels along the epipolar lines in the top and left images and the disparity image has been generayed. The comparison measure used is the sum of absolute differences (SAD). The size of each image is 320 by 210 (pixels) and that of the window in which the SAD is calucuated is 7 x 7. These processes used the Triclops Software Development Kit (SDK) [7]. In these conditions, the depth resolution becomes 8cm at 1.5m.

Figure 5 show images and the depth map acquired from a stereo unit of the SOS. Figure 5 c) is a color image and figure 5 a) and d) are monochrome images. Figure 5 b) is a disparity image obtained from these images. Objects nearer to the camera appears in higher gray levels. These high-resolution images in all direction are acquired from the SOS.

Figure 6 shows a processed image as seen from the center of the SOS by using the method described in Section

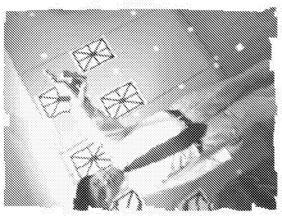


Figure 6. Image seen from a center of SOS

3.1. From this figure and figure **??** c), it seems that pixels are shift to the image center according to the distance and the occlusion areas which are caused by shifted pixel are filled up by the color of far pixel from the stereo unit with compatibility. However discontinuity is caused at miss-corresponding point which is lacked texture and occlusion area at a stereo processing.

The images as seen from the center of the SOS are generated from the images of each stereo unit and a spherical image mosaic which integrate them is shown in figure 7. To compare the present method with the previous results, a spherical image mosaic, by settingthe depth value as 100 cm distance has been acquied from the SOS (Figure 8). Figure 7 b), c) and d) are magnified images by which the regions are cut out from the image shown in figure 7 a) and figure 8 b), c) and d) are magnified images of same area shown in figure 8 a). Figure 7 b) and figure 8 b) shows good mosaic of person's image. However, figure 8 b) shows that the background area has an overlap area as the distance to the background is farther than the set distance. Thus the images seen from the center of the SOS can be generated well by using the distance to the object. A spherical image mosaic seen from the single viewpoint is generated with satisfied accuracy for surveillance and monitoring.

5. Conclusions

In this paper, we have proposed a method to generate a spherical image mosaic of all directional images with different viewpoint. The SOS can acquire all directional color and stereo pair images, in real time, concurrently. The images acquired from the SOS are images with a different viewpoint. Hence, by using three-dimensional information which are calculated from stereo pair images, it is possible to generate images as seen from the center of the SOS. A spherical image mosaic which is projected from a single viewpoint can be generated by integrating them. Since the 2-D spherical image mosaic represents the dynamic environment from the center of the system, various processing can be easily performed on the spherical image mosaic. The spherical image mosaic is useful for surveillance and monitoring, visualization and virtual reality applications.

As the future works, we will construct a PC cluster to support real time applications. We also planed to introduce recognition function using a spherical image mosaic to support more complicated user interactions.

References

[1] B.B.Madhavan, H.Tanahashi, C.Wang, K.Yamamoto, and Y.Niwa. Induction of a real-time omni directional stereo images acquisition system for virtual environment application. *CGIM2000*, pages 83–86, November 2000.

[2] C.Fermuller, Y.Aloimonos, P.Baker, R. Pless, J. Neumann, and B. Stuart. Multi-camera networks: Eyes from eyes. *Proc. IEEE Workshop on Omnidirectional Vision*, pages 11 – 18, June 2000.

[3] H.Tanahashi, C.Wang, Y.Niwa, and K.Yamamoto. Live events accessing for multi users with free viewpoints using stereo omni-directinal system. *Proc. 2nd IEEE Pacific-Rim Conf. on Multimedia*, page to appear, October 2001.

[4] H.Tanahashi, D.Shimada, K.Yamamoto, and Y.Niwa. Acquisition of three-dimensional information in a real environment by using the stereo omni-directional system(sos). *Proc. 3rd Int.Conf. on 3-D Digital Imaging and Modeling*, pages 365 – 371, May 2001.

[5] H.Tanahashi, K.Yamamoto, C.Wang, and Y.Niwa. Development of a stereo omni-directional imaging system (sos). *Proc. IEEE Int. Conf. on Industrial Electronics, Control and Instrumentation*, pages 289–294, October 2000.

[6] J.C.Nebel, F.J.R.Miguel, and W.P.Cockshott. Stroboscopic stereo rangifinder. *Proc. 3rd Int.Conf. on 3-D Digital Imaging and Modeling*, pages 59–64, May 2001.

[7] PointGreyResearch. Triclops stereo vision software development kit (sdk) version 2.2.2 - user's guide and command reference-.

[8] T.Kanede, P.Rander, and P.J.Narayanan. Virtualized reality: Constructing virtual worlds from real scenes. *IEEE Multimedia*, 4(1):34–47, January 1997.

[9] T.Kawanishi, K.Yamazawa, H.Iwasa, H.Takemura, and N.Yokoya. Generation of high-resolution stereo panoramic images by omnidirectional imaging sensor using hexagonal pyramidal mirrors. *Proc. 14th IAPR Int. Conf. on Pattern Recognition*, 1:485 – 489, 1998.

[10] T.Matsuyama. Cooperative distribution vision - dynamic integration of visual perception, action and communication. *Proc. of Image Understanding Workshop*, pages 365–384, 1998.

[11] V.Nalwa. Bell labs solid-state video invention, ccd, comes full circle. *http://www.lucent.com/press/0196/060103.bla.html.*

[12] Y.Kimuro, T.Yuuba, T.Kiriki, and T.Hasegawa. Moving object detection using an omni-directional vision system. *Proc. 4th Japan-France / 2nd Asia-Europe Congress on Mechatronics*, pages 546–551, 1998.

[13] Y.Yagi. Omnidirectional sensing and its applications. *IEICE Trans. Inf. and Syst.*, E82-D(3):568 – 579, March 1999.

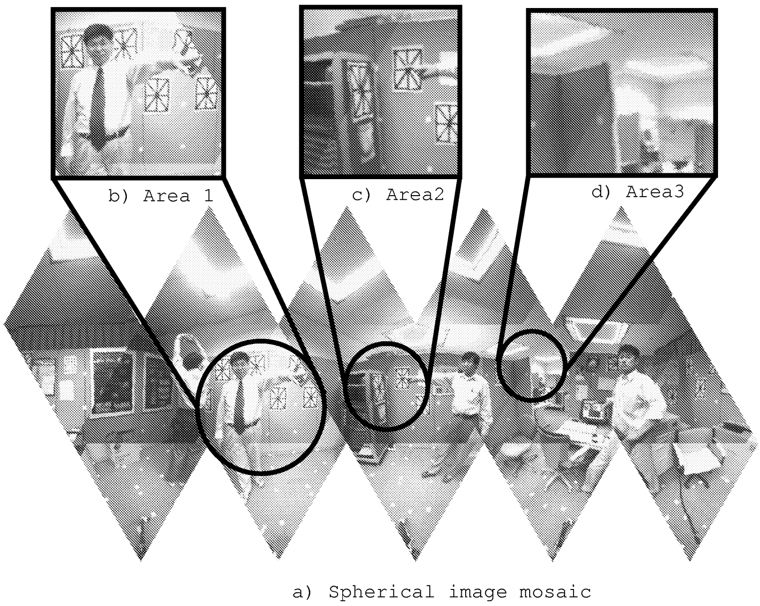


Figure 7. Spherical image mosaic using proposed method

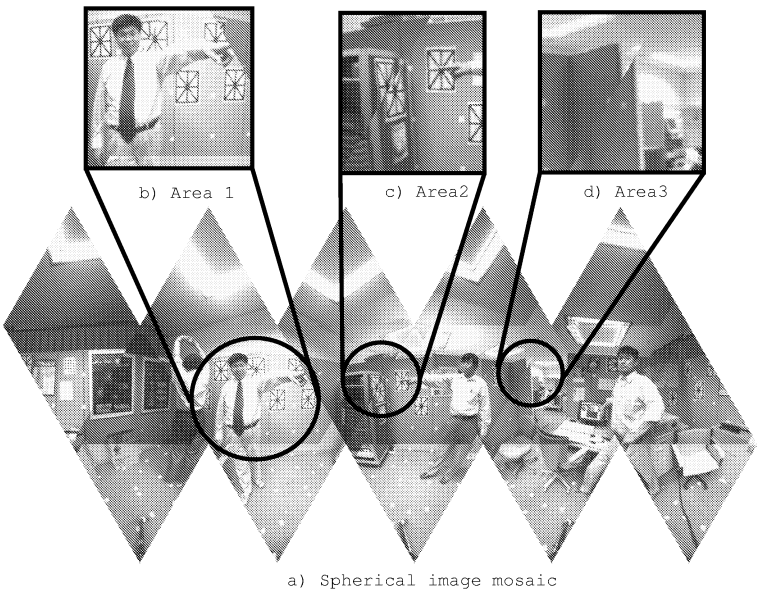


Figure 8. Spherical image mosaic using previous method