

A Camera Revolver for Improved Image Stitching

Kei Utsugi*
Systems Development Laboratory,
Hitachi, Ltd.

Toshio Moriya†
Systems Development Laboratory,
Hitachi, Ltd.

Abstract

A technique for rotating a multiple-camera system for making omnidirectional movies is described that achieves better image stitching. It eliminates mismatches caused by parallax errors due to the distance between the optical centers of each camera.

1 Introduction

Immersive projection display (IPD), which display surrounding images on a special screen (cylindrical, spherical, plane, and so on), are attracting increasing attention because they create a sensation of immersive reality. However, these systems need wraparound or omnidirectional high-resolution movies of the actual scene, and it is quite difficult to shoot such movies with sufficient quality using a single camera. One way to obtain such images is to use a multiple-camera system in which each camera faces a different direction and shoots images in that direction. The images obtained are stitched together into a large high-resolution image. This process is called *stitching*[1] or *mosaicing*[2].

Theoretically, perfect stitching requires that all the cameras have a the same optical center. However, the optical centers of those cameras are usually not at exactly the same point because of the physical sizes of the lenses and camera mechanisms. The behavior and output image of an actual multiple-camera system thus differ from those of the ideal one. To reduce the mismatch in stitching due to parallax errors, we have developed a technique in which the cameras are rotated synchronously around the center of the system.

2 Omnidirectional scenes for IPDs

Two examples of our IPD systems are shown in figure 1 and 2. They consist of multiple projectors and PCs operating in parallel[3]. The one in figure 1 has a cubic screen that occupies 270 degrees of the user's viewing angle. The one in figure 2 has a spherical screen that occupies 140 degrees of the user's viewing angle.

The purpose of our research was to develop a technique for making high-resolution source movies for these IPD systems using the omnidirectional camera system. Such movies need a wide view and the highest resolution possible.

A photograph of the camera system we used is shown in figure 3. Five cameras are set around the

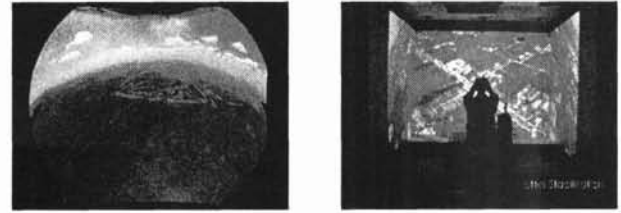


Figure 1: Fisheye image taken using 35 mm film (left), and corresponding digitized image on a cubic IPD system (right). However, the resolution of source movie is yet inadequate for this display system.

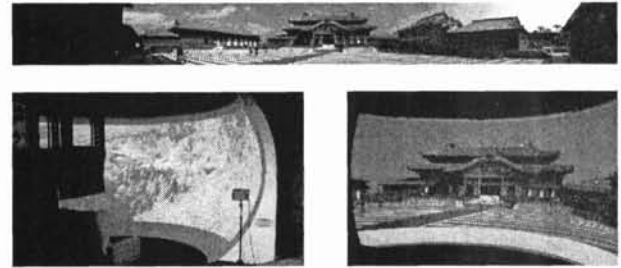


Figure 2: Omnidirectional image taken using multiple cameras (top), and corresponding digitized image spherical IPD (bottom left and right).

circumference of a cylindrical body at even intervals (72 degree); each one captures the image in its respective direction as a digital movie. A sixth camera mounted on the top shoots in the upward direction. The length of this system is about 10cm. Each camera uses a wide-angle fish-eye lens and has a resolution of 768 pixels(wide) and 1024 pixels(high).

3 Problem in stitching images

Because of the distances between the cameras' optical centers, it is impossible to perfectly stitch the images together. An example of parallaxic discordance is shown in figure 4. The discordance of background is less conspicuous than that of the person near the camera because the degree of parallax error between each

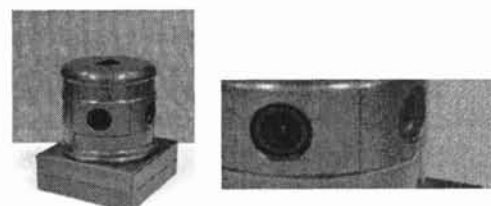


Figure 3: Omnidirectional camera system [4].

*Address: 1099 Ohzenji, Asao-ku, Kawasaki, Kanagawa, 215-0013, Japan. E-mail: utsugi@sdl.hitachi.co.jp

†Address: 1099 Ohzenji, Asao-ku, Kawasaki, Kanagawa, 215-0013, Japan. E-mail: moriya@sdl.hitachi.co.jp



Figure 4: Discordance due to parallax error is more conspicuous the closer the object to the camera.

image depends on the distance between the object and camera.

The degree of the stitching error is closely related to how the images are stitched together. In the system we used, images are stitched by mapping them onto a virtual sphere with a radius of 5 m (as shown in figure 10). This method guarantees good matching for objects 5 m from camera. A comparison of the discrepancy using this method (A) to that using another stitching method (B), in which the optical center of each camera is assumed to be exactly at the center of the system. (This graph shows the theoretical value given the conditions given in the next section; $d=20\text{mm}$, $s=5000\text{mm}$, $\theta-\phi = \cos 36^\circ$). Method (B) is equivalent to mapping images onto a sphere with an infinite radius. As shown in figure 5, method (A) is better when shooting in a confined area (i.e., a room).

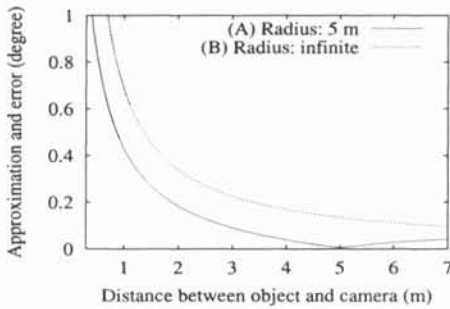


Figure 5: Relation between distance of an object from the center of camera system and discrepancy in image.

However, the nearer the object, the more the discrepancy is. Objects within 1m of camera greatly affect the quality of the stitching. Fortunately, discrepancy affects only the area around the seam ("blending area" in figure 6). Therefore, by rotating the cameras so that head-on images are captured of each object close to the system, we can obtain better stitched images.

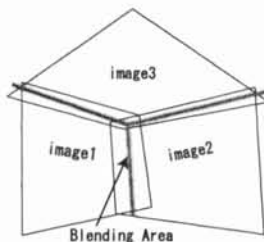


Figure 6: Discordance due to parallax error affects only the "blending areas," the areas around the seams.

4 Camera Revolver System

We call an omnidirectional camera system using our rotation technique a *camera revolver* system. The concept is illustrated in figure 7. The camera system is placed at the end of a robot arm (figure 8) so that the cameras revolve around the ideal optical center. A range finder system monitors the distance of the objects around the cameras and outputs the data in the form of a depth map (figure 9). This depth information does not have to be precise, so the range finder does not need to be placed at exactly the same position as the camera system.

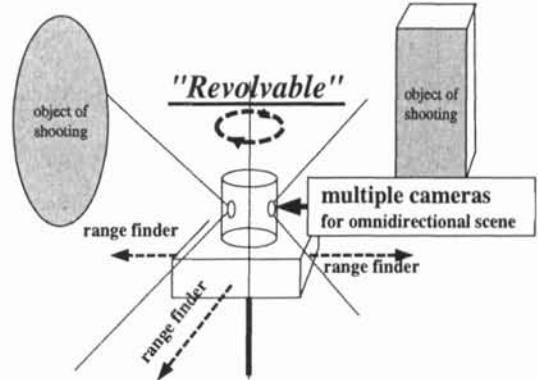


Figure 7: The concept of *Camera Revolver*.

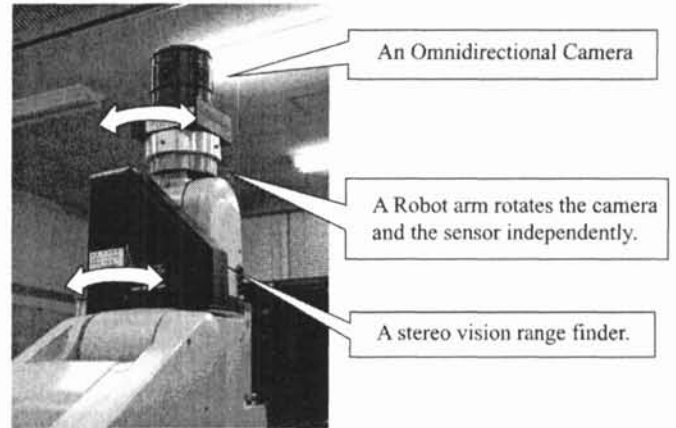


Figure 8: Photograph of *Camera Revolver*. An omnidirectional camera system and a range finder are placed on the robot arm.

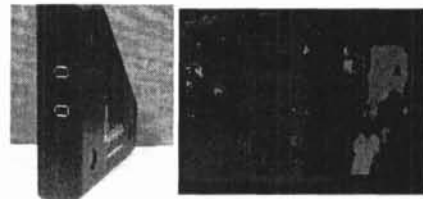


Figure 9: A range finder(left) and depth map (right).

Although it is better to measure depth information all around the camera system using multiple range finders, our current system has one range finder, which covers about 60 degrees and rotates around the center of the system, independent of the rotation of the camera. If an object comparatively close to the camera system is detected in a seam, the system rotates the cameras so as to move the object out of the seam. Since the

system shoots the 360° scene, we obtain substantially the same image regardless of rotation.

To control the rotation of the system, we need a mathematical model of the effect of a mismatch on the image. Figure 10 illustrates the mismatch due to parallax error related to the position of camera. We define the center of the system as O , the optical center of a camera as C , and the object in view as P .

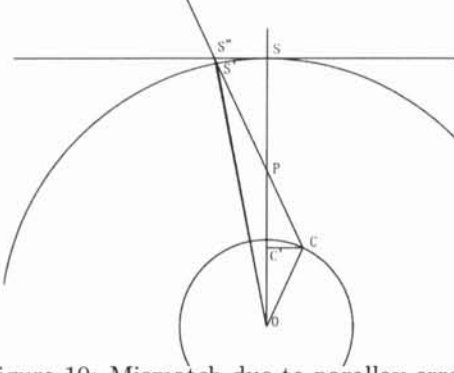


Figure 10: Mismatch due to parallax error,

We define $\vec{OC} = r\vec{\theta}$ and $\vec{OP} = d\vec{\phi}$, where $\vec{\theta}$ and $\vec{\phi}$ are vectors, each with a length of 1. (In our estimation $r \equiv 20mm$,) The virtual sphere, S_0 , on which the images are mapped is centered at O and has a radius, s , of $5m$. The intersection of line \vec{OC} and the sphere S_0 is defined as point S . Point S represents the image of object P mapped on the sphere S_0 when the omnidirectional picture is shot with an ideal optical center O . The intersection of line \vec{CP} and the sphere S_0 , point S' represents the image of object P mapped onto the virtual sphere S_0 using the actual optical center of a camera, C . Angle $\angle SOS'$ represents the degree of discrepancy between the ideal image and the actual one. Plane S_1 is defined by point S and vector $\vec{OP}/|OP|$ as its normal. The intersection of line \vec{CP} and plane S_1 is defined as Point S'' . As angle $\angle SOS'$ must be described using complex calculation, we use $\angle SOS''$ as its approximate value. (The degree of error from this approximation is calculated in the Appendix)

$$\begin{aligned} |SS''| &= \frac{|CC'||SP|}{|PC'|} = \frac{|r \sin \angle COC'| |s - d|}{d - r \cos \angle COC'} \\ &= \frac{r\sqrt{1 - |\vec{\theta} \cdot \vec{\phi}|^2} |s - d|}{d - r\vec{\theta} \cdot \vec{\phi}} \end{aligned} \quad (1)$$

Therefore, the discrepancy is approximately

$$\angle SOS'' = \arctan \frac{r|s - d|\sqrt{1 - |\vec{\theta} \cdot \vec{\phi}|^2}}{s(d - r\vec{\theta} \cdot \vec{\phi})}. \quad (2)$$

For reasons of practical, we use a simple performance function:

$$\angle SOS'' \approx \frac{r|s - d|\sqrt{1 - |\vec{\theta} \cdot \vec{\phi}|^2}}{sd} \quad (3)$$

where $1 \gg \frac{r}{d}$

The images shot by each camera are stitched into an omnidirectional image. An alpha map of each camera's image is prepared prior to blending the images in the



Figure 11: Image before rotating camera (top) and after (bottom).

seam areas ($A_m[0 : 1]$, where m is the camera ID). The weight parameter for a seam area as $W_m(\vec{\phi})$.

$$W_m(\vec{\phi}) = 0.5 - |A_m(\vec{\phi}) - 0.5| \quad (4)$$

Parameter $\vec{\phi}$ represents the direction from the center of the system, and the depth map around the system is represented as $d(\vec{\phi})$. The equation below $P(\vec{\psi})$ is a performance function representing the degree of discrepancy in the blending area, where the rotation of the system is represented by $\vec{\psi}$.

$$P(\vec{\psi}) \equiv \sum_m \iint \frac{I(\vec{\phi}) W_m(R(\vec{\psi}, \vec{\phi})) |s - d(\vec{\phi})| \sqrt{1 - |\vec{\theta}_m \cdot \vec{\phi}|^2}}{sd(\vec{\phi})} d\phi \quad (5)$$

$R(\vec{\psi}, \vec{\phi})$ represents vector $\vec{\phi}$ after rotation of camera system $\vec{\psi}$. The weighting parameter, $I(\vec{\phi})$, is additional information about the shape of the display system. (For example, we do not need upper and lower images for a spherical IPD shown in figure 2.)

Although we use $d(\phi)$ as a continuous function in the equation 5, the directions in which the sensor can measure distances between a camera and an object are generally given as discrete samples of vectors, $\vec{\phi}_i$, where i represents a sample of measurements. To obtain a continuous depth map, $d(\vec{\phi})$, from discrete depth data, $d(\vec{\phi}_i)$, we make a Delaunay diagram of $\vec{\phi}$ on the virtual sphere. When $\vec{\phi}_{T(0)}, \vec{\phi}_{T(1)}, \vec{\phi}_{T(2)}$ are the vertices of a Delaunay triangle in which the direction $\vec{\phi}$ we needs, we define $d(\vec{\phi})$ as :

$$d(\vec{\phi}) \equiv \sum_i \beta_i d(\phi_{T(i)}) \quad (6)$$

$$\beta_i \equiv \frac{|(\vec{\phi}_{T(j)} - \vec{\phi}) \times (\vec{\phi}_{T(k)} - \vec{\phi})|}{|(\vec{\phi}_{T(j)} - \vec{\phi}_{T(i)}) \times (\vec{\phi}_{T(k)} - \vec{\phi}_{T(j)})|} \quad (7)$$

where $i \neq j \neq k$.

Several angles $\vec{\psi}_n$ that locally minimize this above performance function may be found, and one of which should be selected according to the another total performance function below. Rotating the camera system reduces image quality because of the mechanical vibration, and it also changes the stitching parameters for

the background images. The total performance function to choose $\vec{\psi}_n$ considers these effects:

$$P(\vec{\psi}_n) - P(\vec{\psi}_0) - V_A |t_n - t_0| - V_B \int_{t_0}^{t_n} P(\vec{\psi}(t)) dt \quad (8)$$

where $\vec{\psi}(t_0) = \vec{\psi}_0$ represents the current direction of the system by a time parameter t_0 , and $\vec{\psi}(t_n) = \vec{\psi}_n$ represent the direction after rotation. Weight parameter V_A describes derogation due to the mechanical vibration during rotation. Another weight parameter, V_B , describes the effect of stitching mismatch during rotation. When this performance function is small enough, the camera system should not be rotated because the benefit of rotation does not exceed the negative effect.

5 Conclusion

Several studies have investigated ways to stitch images into a large panoramic image, but the discordance of the optical centers makes it impossible to obtain perfect stitching. As the degree of discordance depends on the distance between the cameras and the target object, we developed a technique for rotating multiple cameras so as to keep the objects from falling into the seams between images. To improve this method so as to produce highly accurate stitched images, we need to develop a stabilization technique that eliminates the effects of camera shaking when the cameras are rotated.

Acknowledgements

This work was supported by the Telecommunications Advancement Organization of Japan.

Appendix: Precise Discordance

We approximated the degree of discordance using $\angle SOS''$ instead of $\angle SOS'$ in section 4. The disparity between $\angle SOS'$ and $\angle SOS''$, $\angle S'OS''$ is calculated as follows. (See also figure 12)

$$\angle S'OS'' = \angle OS'P - \angle OS''S' \quad (9)$$

$$s \sin \angle OS'P = d \sin \angle OPC = |OC''| \quad (10)$$

$$\begin{aligned} |OC''| &= |C'C| \cdot |PO| / |PC| \\ &= \frac{r \sqrt{1 - |\vec{\theta} \cdot \vec{\phi}|^2} \cdot d}{\sqrt{(d - r\vec{\theta} \cdot \vec{\phi})^2 + r^2(1 - |\vec{\theta} \cdot \vec{\phi}|^2)}} \quad (11) \end{aligned}$$

$$\angle OS'P = \arcsin \left(\frac{r d}{s r} \sqrt{\frac{1 - |\vec{\theta} \cdot \vec{\phi}|^2}{(\frac{d}{r})^2 + 1 - 2(\frac{d}{r})|\vec{\theta} \cdot \vec{\phi}|}} \right) \quad (12)$$

$$\angle OS''S' = \angle OS''P = \angle OS''S - \angle PS''S \quad (13)$$

$$= \arctan \left(\frac{|SS''|}{s} \right) - \arctan \left(\frac{|SS''|}{|PS|} \right) \quad (14)$$

$$\begin{aligned} &= \arctan \frac{r|s-d|\sqrt{1-|\vec{\theta} \cdot \vec{\phi}|^2}}{(d-r\vec{\theta} \cdot \vec{\phi})s} \\ &\quad - \arctan \frac{r\sqrt{1-|\vec{\theta} \cdot \vec{\phi}|^2}}{d-r\vec{\theta} \cdot \vec{\phi}} \quad (15) \end{aligned}$$

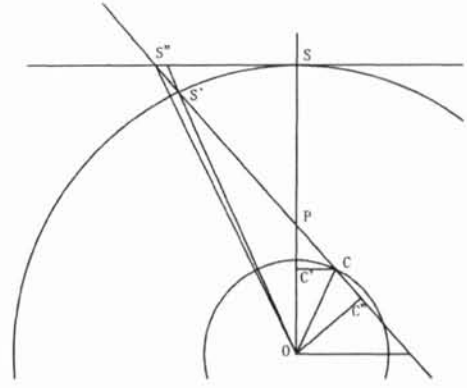


Figure 12: Mismatch due to parallax error.

- O: Center of system
- C: Optical center of camera
- P: Object in view

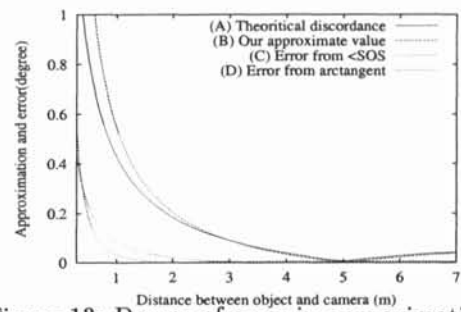


Figure 13: Degree of error in approximation

The error from approximation, $E(d, r, s, \vec{\phi}, \vec{\theta})$, is calculated as the following equation 16. The graph in figure 13 shows the degree of this error (C) compared to the other error factor (D) which is the error between equations 2 and 3.

$$\begin{aligned} E(d, r, s, \vec{\phi}, \vec{\theta}) &= \arcsin \left(\frac{r d}{s r} \sqrt{\frac{1 - |\vec{\theta} \cdot \vec{\phi}|^2}{(\frac{d}{r})^2 + 1 - 2(\frac{d}{r})|\vec{\theta} \cdot \vec{\phi}|}} \right) \\ &\quad - \arctan \frac{r|s-d|\sqrt{1-|\vec{\theta} \cdot \vec{\phi}|^2}}{(d-r\vec{\theta} \cdot \vec{\phi})s} \\ &\quad + \arctan \frac{r\sqrt{1-|\vec{\theta} \cdot \vec{\phi}|^2}}{d-r\vec{\theta} \cdot \vec{\phi}} \quad (16) \end{aligned}$$

References

- [1] S. E. Chen, "QuickTime VR — An image-based approach to virtual environment navigation," In Proc. SIGGRAPH 95, pages 29-38, 1995.
- [2] S. Pelg and J. Herman, "Panoramic Mosaics by Manifold Projection," In Proc. IEEE CVPR 1997, pages 338-343, 1997.
- [3] H. Takeda, M. Yamasaki, T. Moriya, T. Minakawa, F. Beniama, and T. Koike, "A video-based virtual reality system," In Proc. ACM VRST99, pages 19-25, 1999.
- [4] <http://www.ptgrey.com/products/ladybug/>
- [5] T. Wada, N. Ukita, and T. Matsuyama, "Fixed viewpoint pan-tilt-zoom camera and its application," Trans. of the Institute of Electronics, Information and Communication Engineers, vol. J81-DII, no. 6, pages 1182-1193, 1998 (in Japanese).