Computing the Field-of-View of a Stitched Panorama to Create FoV Sensitive Virtual Environments

Kenji Mase Hikaru Nishira
ATR Media Integration & Communications Research Laboratories
Seika-cho, Soraku-gun, Kyoto 619-02 JAPAN
81-774-95-1440(phone), 81-774-95-1408(fax)
E-mail: `mase@mic.atr.co.jp`

Abstract

The representation of a visual scene is a key issue in navigating through image-based virtual environments. Using affine and projective invariants, sequential/multiple images from a pivoting camera can be easily mosaiced into a planar sheet. We discuss the importance of field-of-view to share the sensation of the scene observer (i.e. camera person) in a virtual environment and propose a method of obtaining the total field-of-view angle of image mosaics. The extension to an environment with moving objects as well as applications are presented.

1. Introduction

Recent progress in computer graphics technology, especially in the texture-mapping (e.g. environment-mapping) of pixel images is closing the gap between 3-D scene reconstruction from image sequences and the creation of geometrically modeled virtual environments. Due to the visual perception ability of humans, we can "feel" reality by mapping scene images (texture) onto roughly reconstructed geometric models such as human faces and buildings. This benefit is also applicable to general visual scenes for such purposes as geographical mapping, driving through towns, walking through rooms in museums, etc. Even image mosaics, i.e. stitched panoramas, constructed from pivoting camera shots along a navigation path can provide good interactivity and an explorative experience in many cases [1, 2, 5, 10].

Naturally, in the next step we would like to perceive new visual sensations of reality such as expansion-of-view, i.e. field-of-view (FoV). For example, the sensation of presence at the same position of a camera-person in a scenery is indispensable for sharing and experiencing the virtual world beyond time and place. The capability of sharing such a sensation as well as the freedom to explore the space are important in virtualized communications. However, it can also be distracting if the picture is projected with a different FoV. If we keep or obtain the correct FoV angle, we can project the picture on a screen in the same dimensions the camera person sensed while location hunting. Vision-based navigation research has overlooked utilizing the obtained rotational parameters between frames to estimate the entire FoV. In this paper, we propose a FoV conscious image projection method based on the estimation of the entire FoV through the image sequence. The estimation of the entire FoV is possible by accumulating rotational parameters and the estimated field-of-view angle of the lens.

Methods of constructing image mosaics have recently been gaining attention, especially in virtual environment creation [2]. Mosaics can, of course, be easily constructed from an image sequence if the camera parameters are sensed and recorded carefully. This is, however, a time-consuming task and camera information is not usually provided with video programs. Automatic image registration methods of image sequences based on the traditional passive navigation and structure-from-motion paradigm [3] have long been investigated. However, the geometric approaches are not robust to noise. New formulations [6, 9] of image registration methods based on an image-based paradigm have recently been proposed and fit this mosaic problem.

In the following, a mosaicing method based on image correlation and the FoV of mosaics is described in Section 2. A method of obtaining a panoramic movie by using a layered representation and a three-step segmentation of moving objects is introduced in Section 3. Finally, the concept and the system configuration of FoV sensitive virtual environments are described in Section 4 followed by a discussion of applications.

1015-4651 / 96 $5.00
Proceedings of ICPR '96

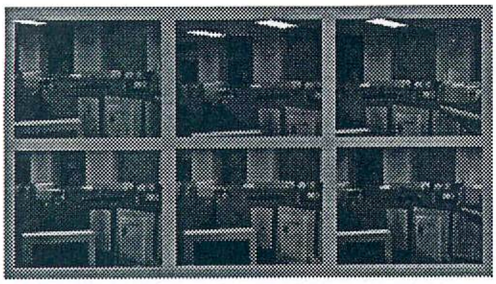


Figure 1. Input example pictures for mosaic.

2. Mosaics of Multiple Images

2.1. Relations of Images from Pivoting Camera

We will follow the mosaicing method described in [9]. Two images are related by a transformation:

$$\mathbf{x}' = \mathbf{M}\mathbf{x}, \tag{1}$$

where $\mathbf{x} = (x, y)$ and $\mathbf{x}' = (x', y')$ are corresponding points of images. M is the best transformation with which the maximum correlation is accounted by minimizing the following error equation in terms of the sum of squared intensity differences:

$$E = \sum_i [I'(x'_i, y'_i) - I(x_i, y_i)]^2. \tag{2}$$

When the composition is a panoramic mosaic of the image collections obtained from an ideally pivoting camera, the relationship between corresponding points in the two images follows Kanatani's equation [4]. For an initial inclination (θ) of the camera system and a small amount of pan (α) and tilt (γ), that is, rotations around the x- and the y-axis (i.e. horizontal and vertical axes of the camera view plane), respectively, equation 1 can be rewritten as:

$$x' = \frac{m_0 x + m_1 y + m_2}{m_6 x + m_7 y + 1} = f\frac{x + \alpha \sin\theta y + f\alpha\cos\theta}{-\alpha\cos\theta x + \gamma y + f} \tag{3}$$

$$y' = \frac{m_3 x + m_4 y + m_5}{m_6 x + m_7 y + 1} = f\frac{-\alpha \sin\theta x + y - f\gamma}{-\alpha\cos\theta x + \gamma y + f} \tag{4}$$

where f is the fixed focal length[7].

Figure 1 shows example pictures of a room scene taken from a hand-held still camera moving around while shooting and figure 2 is an example result of an image mosaic with this approach obtained. In order to avoid becoming stuck in local minima, a hierarchical approach was taken to compute the transformation matrix M by employing two SSD methods, i.e. the first one used simple local correlations to compute a large translational displacement factor and then the Levenberg-Marquardt method[9, 8] was used for nonlinear iterative minimization of the target equations.

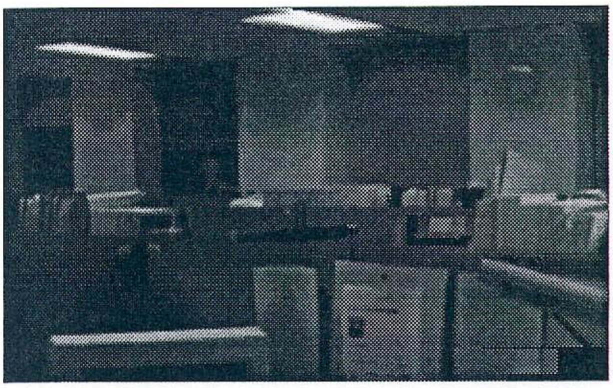


Figure 2. An example of mosaic room.

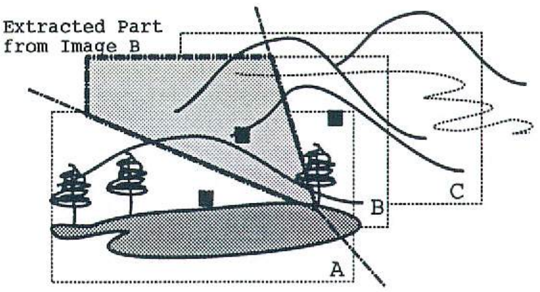


Figure 3. Voronoi blending.

2.2. Blending Tiles

In this paper, we focus on the alignment problem of mosaicing multiple images, although blending of the images after the alignment is also an important phase of composition. Simple intensity morphing using linear interpolation between adjacent image tiles provides better quality. We employ the voronoi polygon (see Figure 3) set to define the area and the picture combination for interpolation. Voronoi tessellation can be obtained by computing the bi-secting perpendicular line of a segment connecting the two centers of the images. The mosaic in Figure 2 is obtained from six still pictures taken with a digital still camera of 640 × 512 image resolution with the voronoi blending method. Some discontinuous boundary is still observed at the edge of the picture tiles. This happens unless there exists a counterpart for blending.

2.3. Total FoV

In later stages of image acquisition and mosaicing, the rotational parameters between two successive

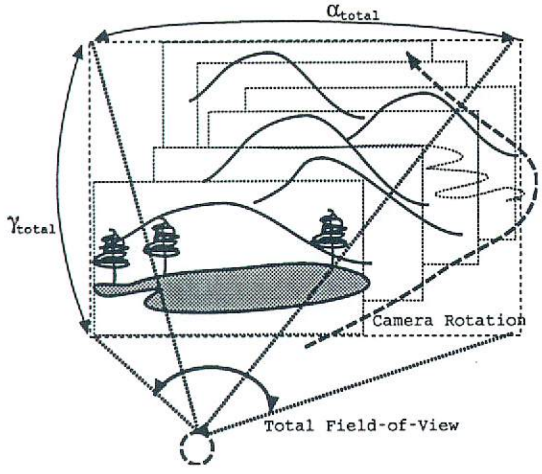
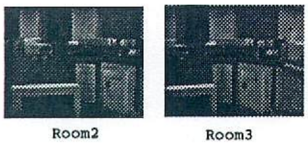
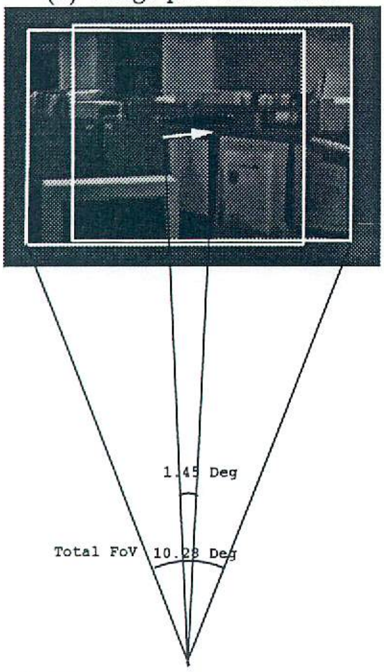


Figure 4. Total field-of-view.



(a) Image pair for mosaic



(b) Mosaic and computed FoV

Figure 5. Experimental result of computing FoV.

frames, i.e. $\alpha_{i,i+1}$ and $\gamma_{i,i+1}$, are computed and the current total FoV ($\alpha_{total}(T)$, $\gamma_{total}(T)$) changes. With the latest update of FoV and the FoV sensitive projection onto the screen, the observer can get the same sensation of expansion-of-view as the movie shooter. The computation of the total FoV can be done in the following way:

$$\alpha_{total}(T) = \max_{0 \le t \le T}(\sum_{0 \le i \le t} \alpha_{i,i+1}) - \min_{0 \le t \le T}(\sum_{0 \le i \le t} \alpha_{i,i+1}) + \frac{fov_h^l}{2} + \frac{fov_h^r}{2} \quad (5)$$

$$\gamma_{total}(T) = \max_{0 \le t \le T}(\sum_{0 \le i \le t} \gamma_{i,i+1}) - \min_{0 \le t \le T}(\sum_{0 \le i \le t} \gamma_{i,i+1}) + \frac{fov_v^t}{2} + \frac{fov_v^b}{2} \quad (6)$$

where fov_h^l and fov_h^r are the horizontal elements of FoV of the image located at the left extreme and right extreme, respectively. Analogously, fov_v^t and fov_v^b are the vertical elements at the upper extreme and lower extreme, respectively.

Figure 5 is an example result of computing the total FoV from a pair of images (a). The motion parameter computed for this image pair is:

$$\mathbf{M} = (1.0033, -0.000815, -105.53, 0.001737, 1.0038, -8.3860, 0.000006, -0.000003)$$

The mosaic is shown in Figure 5(b) in which the FoV is $(\alpha_{total}, \gamma_{total}) = (10.28, 6.74)$ in degree and the $f = 1:6.48$.

3. Panoramic Movie

A natural scene is not always static, for example, scenes of a lake with flying birds, or people in action such as when playing games in a field. This section describes how to deal with multiple images of a scene that contains moving objects.

3.1. Layered Scene Representation

We assume the scene to be layered by the combination of a static background and moving foreground objects which may move along any path between the background and the camera position as shown in Figure 6. In other words, the sufficient portion of an image which is supposed to represent the static background is available to be used as a reference in mosaicing.

The layering is done by an image subtraction between successive images after the registration for mosaicing described in Section 2. In order to avoid in-correct registration for images that contain moving objects, we employ a three-step algorithm: first, an image is separated into evenly blocked areas and then the piecewise registration is performed. Second, the registration parameter sets {M} are evaluated to find a dominant motion, i.e. motion due to the moving camera. Third, using the obtained parameters, the images are aligned and subtracted.

3.2. Segmentation of Moving Objects

At the stage of subtraction, we have two background images with holes, B_t and B_{t+1}, and a common region containing foreground objects, $O_{t,t+1}$. The object area is obtained as a logical OR at time t and $t+1$. In order to separate the object region of each time image, O_t, we exploit the temporal smoothness of object motion and take the intersection of two OR'ed regions taken from three frames;

$$O_t = O_{t-1,t} \cap O_{t,t+1}. \qquad (7)$$

The holes in the background image are filled up partly by the process above and with the image in which the portion contains the background scene.

4. FoV Sensitive Virtual Environments

Why should we care so much about the FoV?

1. We want to share the sensation of the observer (camera-person).
2. FoV is an important criteria in viewing image mosaics.
3. The scale and display of an entire image mosaic on a limited display are destructive to the sensitivity of the content, although they are good for an overview of the information contained.

We prefer to use wider angle image displays (or projectors) for image mosaics especially for panoramic mosaics. This means that mosaicing breaks the framing of a camera shot. After breaking the frame of an image, i.e. the FoV of a frame, we cannot define the size of the displayed image.

It is time that we abandon the notion of frames which are bound to the size of the display and pixels-per-inch. What should bind us are the (true or virtual) laws of nature and our intention of how we enter and navigate in the environment. The real world shows us details if we come closer to objects. Let us suppose we are approaching an object in order to examine its details. We expect to get some immersive sensation as well as a detailed view. However, as shown in Figure 7, the viewing cone covers a smaller area of the projection plane for a certain part of the scene, thus the viewer will get fewer pixels. This is not a good response to the viewer who has approached the object in order to get more pixels. An FoV sensitive display can solve the problem by showing the image in different resolutions per unit size and FoV's.

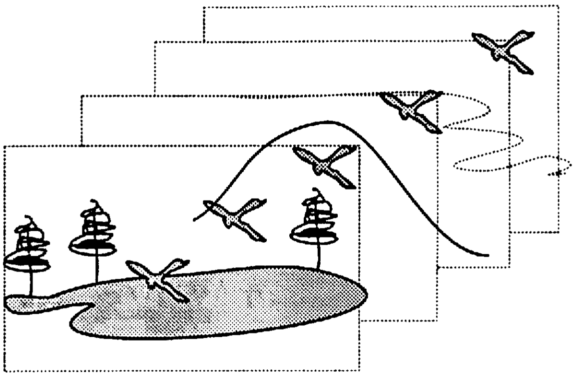


Figure 6. Panoramic movie.

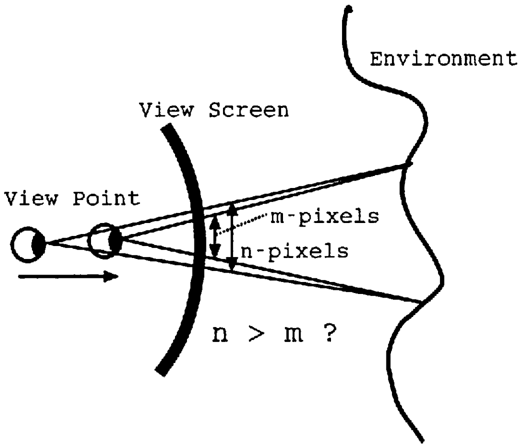


Figure 7. Viewing angle and resolution.

4.1. System Configuration and Response to Viewer's Action

We are constructing an FoV sensitive virtual environment which incorporates the 3D positioning of the viewer and image mosaic rendering (Figure 8). The viewer's position can be obtained by a vision-based position detection algorithm or a pressure sensitive floor. The system initially informs the viewer of the best position where the viewer can enjoy the same view as the observer did when capturing the scene, with the position indicated by means of visible lights on the floor.

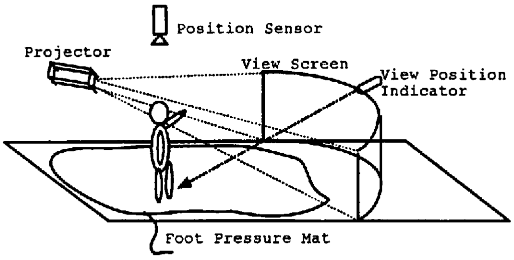


Figure 8. A plan of FoV sensitive virtual environments.

4.2. Applications

Creating image-based virtual environments has many potential applications because they can provide realistic details of visual information. Once we virtualize the real environment in terms of image mosaics or panoramic movies, we can overlay geometrical graphics objects or real objects onto the scene.

Panoramic movies can provide the sensation of wide spread views, while at the same time the objects may move across the view. This can have many applications, including tele-watching of sports events (e.g. virtual soccer stadiums, virtual ski jump slopes) and geographic scenery (e.g. virtual Yosemite National park), and walk-throughs of existing towns and buildings (e.g. virtually-enhanced meta-museums, cyber-towns).

In order to realize walk-throughs of panoramas, the recovery of depth sensation is essential. This has been a key issue in computer vision research. However, when the observer is allowed to navigate within a small offset from the camera path used in the shooting, the sensation of depth can be perceived. This can be done by projecting the environment with a wide field-of-view and small changes in projections. A simple interpolation between two spherical panoramas will give such a sensation. Image-based rendering based on the plenoptic function [6] can be applied to this problem.

5. Conclusion

In this paper, we pointed out the importance and effectiveness of field-of-view for image mosaics. Reconstructing image mosaics, panoramic views and panoramic movies is made possible by recent technological advances in areas such as image registration, camera parameter restoration, depth from motion, and image-based rendering. By incorporating the FoV, we can experience a richer sensation of virtual environments. With the total FoV parameters we can decide the appropriate screen size when showing a mosaic to provide viewers the same sensation of existence as the movie shooter had.

We are also aware of the destructive effects of presentation when printing mosaics on paper. Presentation by a curved projection and a view-point will help in the recognition of the environment even on printed paper.

An FoV sensitive system is under development at ATR which incorporates the algorithms of panoramic movies. Progress in implementation and experimental results of providing panoramic movies will be provided soon.

Acknowledgments

The authors would like thank to Dr. Ryohei Nakatsu and Dr. Kouhei Habara for their support and encouragements in this research and the members of the ATR MI&C lab department 2 for their creative discussions. Nishira is currently an undergraduate student of the University of Tokyo and contributed to this work while his student internship stay at ATR.

References

[1] A. Akutsu and Y. Tonomura. Video tomography: An efficient method for camerawork extraction and motion analysis. *in Proc. ACM Multimedia'94*, pages 349–356, 1994.

[2] S. E. Chen. Quicktime VR – an image-based approach to virtual environment navigation. *Proc. SIGGRAPH'95 (Los Angeles, CA)*, pages 29–38, Aug. 1995.

[3] B. K. P. Horn. *Robot Vision*. The MIT Press, 1985.

[4] K. Kanatani. Camera rotation invariance of image characteristics. *Comput. Vision, Graphics, & Image Processing*, 39(3):328–354, Sept. 1987.

[5] A. Lippman. Movie-maps: An application of the optical videodisc to computer graphics. In *SIGGRAPH'80 Proceedings*, 1980.

[6] L. McMillan and G. Bishop. Plenoptica modeling: An image-based rendering system. *Proc. SIGGRAPH'95 (Los Angeles, CA)*, pages 39–46, Aug. 1995.

[7] D. Murray and A. Basu. Motion tracking with an active camera. *IEEE Trans. on PAMI*, 16(5):449–459, May 1994.

[8] W. H. Press et. al. *Numerical Recipes in C*. Cambridge University Press, 1992.

[9] R. Szeliski. Video mosaics for virtual environments. *IEEE Computer Graphics and Applications*, 16(2):22–30, March 1996.

[10] L. Teodosio and W. Bender. Salient video stills: Content and context preserved. *in Proc. ACM Multimedia'93*, pages 33–46, 1993.