1998 International Computer Symposium
Workshop on Computer Graphics and Virtual Reality
December 17-19, 1998, N.C.K.U., Tainan, Taiwan, R.O.C.

BUILDING FULL VIEW SPHERICAL PANORAMAS VIA AN IMAGE-BASED APPROACH

Ju-Wei Chen, Shu-Cheng Huang, Tse Cheng, Der-Lor Way, and Li-Sheng Shen*

U500 Computer & Communications Research Laboratories
Industrial Technology Research Institute
Chutung, Hsinchu, Taiwan 31015, R.O.C.
*Email: jwchen@ccl.itri.org.tw

Abstract

The paper presents an image-based approach for building spherical panoramas for image-based virtual reality systems. The proposed approach includes the design of the omni-directional picture-taking manner, the image warping from a photographic image to the spherical environment map, and the image mosaic algorithm to seam these photographic images together to build a seamless spherical panorama. A parametric spherical environment map is used to store the image data of the spherical panorama. The image mosaic algorithm, through the horizontal and vertical image registration accompanied with fine-tuning, can compute the accurate sight direction of each picture. Thus, it can tolerate some errors in sight directions predefined, which are resulted from the imperfect photographic equipment. When the stitched spherical panorama is navigated, an image browser can retrieve the image data of the scene based on the desired view direction from the spherical environment map.

1 Introduction

For virtual reality systems, the synthesis and navigation of a virtual environment are usually accomplished by one of the following approaches: *3D modeling and rendering*, *branching movies*, and *image-based approach* [1]. From the consideration of creating object, rendering speed, and storage space requirement, the *image-based approach* seems to be a reasonable and feasible approach, especially for rendering complicated scenes, such as the real-world scenery. In this paper, we propose a new approach for building image-based virtual environments.

Many methods of building image-based panoramas have been proposed. The QuickTime VR system contains the function of creating a seamless *cylindrical* panoramic image from a set of overlapping pictures [1]. Shieh *et al.* further improved the stitching method [2]. However, the cylindrical panoramic image cannot be used to store the full-view data of a scene. The function of building an omni-directional image-based panorama has been mentioned in the developer manual of $InfinitePictures^{TM}$, Inc. [3], where the pictures are taken using the fisheye lens. The fisheye lens is more expensive than an ordinary extreme wide-angled lens and is not popular in people's daily life. Szeliski and Shum proposed a numerical approach of image registration, which can cope with arbitrary rotation in camera motion, but the translation is still assumed to be fixed [4]. The rotation free makes the method more time-consuming.

In this paper, we propose a practical and fast image-based approach. The pictures, used to build a full view panorama, are only taken by the camera with an ordinary extreme wide angled lens. First, we propose a design of picture-taking manner. Then, the proposed algorithm of image mosaic is presented. It can tolerate errors in sight directions predefined. Finally, the experimental results are shown to prove the effectiveness of the proposed approach.

2 Image Warping

Many types of environment maps can be used to store the scene data of surrounding environments [5]. For spherical panoramas, spherical environment maps are first taken into consideration. The two types of particular interest are "spherical reflection maps" and "parametric spherical environment maps". "Spherical reflection maps" store the image of an environment as an orthographic projection of a sphere [6]. The orientations near the silhouette of the sphere are sampled very sparsely. This will increase the difficulty of rendering

*Responsible for correspondence.

the spherical reflection maps. Oppositely, "parametric spherical environment maps *(PSEM)*" store the data of any environment into a rectangular image, in which the coordinates of each pixel can linearly map to the corresponding point on a sphere surface. The *PSEM* has many other useful properties. The whole environment data of a spherical panorama can be included in a single contiguous image. All regions are always sampled at least as much as the regions at equator. "Translation" along the equatorial direction in the map corresponds to "rotation" around the axis of two poles of a sphere. Therefore, we adopt the *PSEM* to store the scene of a full view panorama.

The image warping from a photographic image to the *PSEM* is essential when we utilize the *PSEM* to store the scene data. In our proposed method, all photographic images are first transformed onto the space of the *PSEM*. Then, a series of image registrations are performed on these warped images to create panoramic image mosaics. The formula of the image warping is derived as follows.

In the Cartesian (or rectangular) coordinate system (R^3), a sphere with radius ρ can be described as the following equation

$$x_R{}^2 + y_R{}^2 + z_R{}^2 = \rho^2, \tag{1}$$

where x_R, y_R, and z_R are the coordinates of any point on the sphere surface.

On the other hand, any point on a sphere with radius ρ can be denoted by (ρ, θ, ϕ) in the azimuthal coordinate system, where $0 \le \theta < 2\pi$ and $0 \le \phi < \pi$ if we represent the angles in units of radians, or $0° \le \theta < 360°$ and $0° \le \phi \le 180°$ if in units of degrees. For a point on the sphere surface, the coordinates in the two coordinate systems have the following relations

$$\begin{cases} x_R = \rho \sin\phi \cos\theta \\ y_R = \rho \sin\phi \sin\theta \\ z_R = \rho \cos\phi. \end{cases} \tag{2}$$

In other words, the spherical surface of a predefined radius has only two degrees of freedom, represented by two parameters: θ and ϕ, respectively. In the *PSEM*, we utilize two orthogonal axes, X and Y, to represent θ and ϕ, as shown in Figure 1. The resolutions of respective axes, that is, pixels/per degree, should be defined first. Typically, the resolutions of the two axes are defined to be the same. Therefore, the width of the *PSEM* will be twice as long as the height. However, they can also be different. Assume that $x\_resolution\_PSEM$ and $y\_resolution\_PSEM$ are the predefined resolutions of x and y axis, respectively. For any pixel in the *PSEM* with coordinates (x_m, y_m), we can obtain the corresponding azimuthal coordinates θ and ϕ from (x_m, y_m) as follows

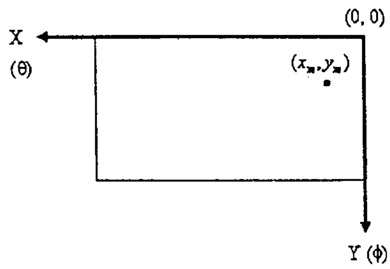


Figure 1: The rectangular image of a parametric spherical environment map.

$$\begin{cases} \theta = x_m / x\_resolution\_PSEM \\ \phi = y_m / y\_resolution\_PSEM. \end{cases} \tag{3}$$

Assume that a spherical panorama can be built from M pictures. Basically, these pictures are taken from a stationary viewpoint but from different sight directions. The sight direction of picture P_i is denoted by (θ_i, ϕ_i) $(1 \le i \le M)$. The spatial relation between picture P_i and the optical center of the camera lens can be illustrated by Figure 2, where the radius of the sphere is the focal length f, and the optical center of the camera lens is located at the center of sphere S. The M sheets of the picture film just tangent the sphere S on respective picture centers.

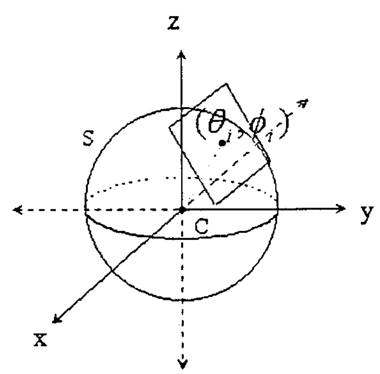


Figure 2: The film of picture P_i with azimuth angle (θ_i, ϕ_i) tangents sphere S on the picture center.

The viewing plane of a picture P_i can be described by a 2D Cartesian coordinate system, as shown in Figure 3. The original of the 2D Cartesian coordinate system is defined to be at the center point of the picture, x axis is defined along the direction of film width and y axis is defined along that of film height. If the film width is designated as "a" and the film height designated as "b", the coordinates of four corner points of a picture - $P1$, $P2$, $P3$, and $P4$ will be $(b/2, a/2)$, $(-b/2, a/2)$, $(-b/2, -a/2)$ and $(b/2, -a/2)$; the coordinates of center points of the four bounding edges - $Q1$, $Q2$, $Q3$, and $Q4$ will be $(0, a/2)$, $(-b/2, 0)$, $(0, -a/2)$ and $(b/2, 0)$, respectively.

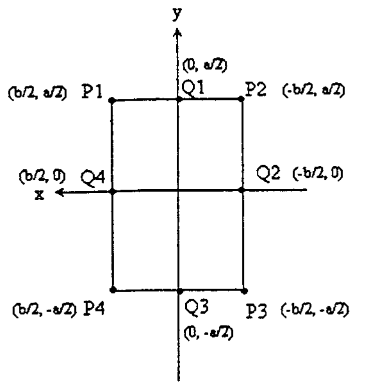


Figure 3: A 2D Cartesian coordinate system represents a picture; the film height is denoted by "a" and the film width is denoted by "b".

Another 3D Cartesian coordinate system ${R^{(i)}}^3$ can describe the spatial relation between the film of P_i and the optical center of the camera lens, as shown in Figure 4. The optical center of the lens is located at the original of ${R^{(i)}}^3$, and the film is located on plane $z^{(i)} = f$. Therefore, the coordinates of the four corner points $P1$, $P2$, $P3$, and $P4$ will be $(b/2, a/2, f)$, $(-b/2, a/2, f)$, $(-b/2, -a/2, f)$ and $(b/2, -a/2, f)$; the coordinates of the four center points $Q1$, $Q2$, $Q3$ and $Q4$ will be $(0, a/2, f)$, $(-b/2, 0, f)$, $(0, -a/2, f)$ and $(b/2, 0, f)$, respectively. The M pictures can be described by their respective 3D Cartesian coordinate systems ${R^{(i)}}^3$ in this way.

The spatial relation between the global (or common) coordinate system R^3 in Figure 2 and the coordinate system ${R^{(i)}}^3$ of picture P_i can be illustrated by Figure 5. Through a series of coordinate transformations, we can derive the pixel coordinates of the global coordinate system R^3 from respective ${R^{(i)}}^3$ of each picture, and *vices verse*. The corresponding azimuthal coordinates (ρ, θ, ϕ) can also be derived easily,

The warped images of the respective photographic images on the *PSEM* are generated based on the formula of image warping. Each pixel of the *PSEM* is mapped from at least one pixel of one photographic image. Using the transformation formula, for each pixel on the *PSEM*, we can derive the coordinates of the corresponding pixel of the photographic image. The coordinates of the pixel on the *PSEM* can linearly map to the azimuthal coordinates (θ, ϕ) as described in Equation 3. In the following, we will derive the relation between the azimuthal coordinates (θ, ϕ) of one pixel on the *PSEM* and the 2D Cartesian coordinates (x, y) of the corresponding pixel on the photographic image based on the sight direction of the picture.

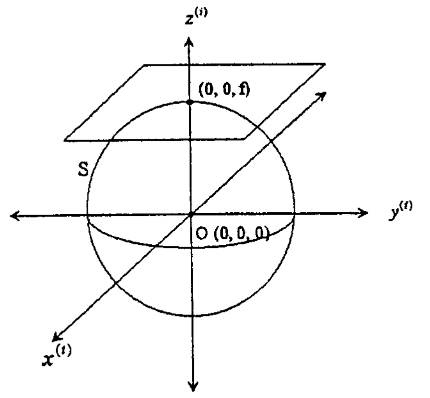


Figure 4: A 3D Cartesian coordinate system describes the spatial relation between the film and the lens within camera.

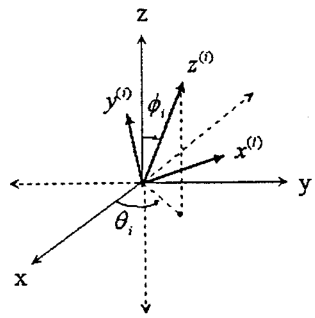


Figure 5: The relation between two Cartesian coordinate systems R^3 and ${R^{(i)}}^3$.

The spatial relation between the global coordinate system R^3 in Figure 2 and the respective coordinate system ${R^{(i)}}^3$ of picture P_i is illustrated by Figure 5. The sight direction of picture P_i is denoted by (θ_i, ϕ_i). In order to obtain ${R^{(i)}}^3$ from R^3, we can assume ${R^{(i)}}^3$ and R^3 being superimposed at first. Next, ${R^{(i)}}^3$ rotates ϕ_i degrees counterclockwise around x axis of R^3, which is designated as $R_x(\phi_i)$. Finally, it rotates $(\theta_i + 90°)$ degrees around z axis of R^3, designated as $R_z(\theta_i + 90°)$. Conversely, we can also derive the coordinates in ${R^{(i)}}^3$ from the coordinates in R^3 through the transformations $R_z(-(\theta_i + 90°))$ and $R_x(-\phi_i)$. For any pixel on the sphere surface with radius f, the coordinates of the pixel in $R^{(i)}$[3] can be derived from the azimuthal coordinates (θ, ϕ) and the sight direction (θ_i, ϕ_i) as follows.

$$\begin{cases} x^{(i)} &= f \sin\phi \sin(\theta - \theta_i) \\ y^{(i)} &= f[\sin\phi_i \cos\phi - \cos\phi_i \sin\phi \cos(\theta - \theta_i)] \\ z^{(i)} &= f[\cos\phi_i \cos\phi + \sin\phi_i \sin\phi \cos(\theta - \theta_i)]. \end{cases} \quad (4)$$

To obtain the coordinates of the corresponding pixel on the photographic image, the pixel on the sphere surface S should further be perspectively projected onto the viewing plane: $z^{(i)} = f$. The film of the picture is just on the viewing plane. The coordinates of each pixel projected on plane $z^{(i)} = f$, designated as (x^p, y^p, z^p), can be derived by the following equation

$$\begin{cases} x^{(p)} &= \frac{f \sin\phi \sin(\theta-\theta_i)}{\cos\phi_i \cos\phi + \sin\phi_i \sin\phi \cos(\theta-\theta_i)} \\ y^{(p)} &= \frac{f[\sin\phi_i \cos\phi - \cos\phi_i \sin\phi \cos(\theta-\theta_i)]}{\cos\phi_i \cos\phi + \sin\phi_i \sin\phi \cos(\theta-\theta_i)} \\ z^{(p)} &= f. \end{cases} \quad (5)$$

The warped image of a rectangular image on the *PSEM* is not rectangular. The actual shape depends on the elevation angles ϕ. Conventionally, an images is represented by a 2D array. Using rectangular images is more convenient for most algorithms of image processing. In the proposed algorithm of image mosaic, many techniques of image processing will be utilized, including the image registration and the image blending. Therefore, we proposed a method to cut all these warped images of irregular shapes into rectangular images.

Basically, the shapes of warped images on the *PSEM* will depend on the elevation angles ϕ's. The nearer anyone pole of a sphere the sight direction is, the wider the warped image is. Therefore, the manner of image cutting should also depend on the shape of a warped image, that is, the elevation angle ϕ of the sight direction. The important part of an image must be included in the warped image after cutting. Taking the focal length *18mm*, the film width *24mm*, and the film height *36mm* as the example, the rules of image cutting are listed as follows.

- If ϕ is less than 10° or greater than 170°, the warped image will be cut into a rectangular region by one horizontal line, as shown in Figure 6(a). This manner of image cutting is named *horizontal cutting*.
- If ϕ is between 10° and 170°, the image region will be cut into a rectangular region by a pair of horizontal lines and a pair of vertical lines, as shown in Figure 6(b). The manner is named *vertical cutting*.

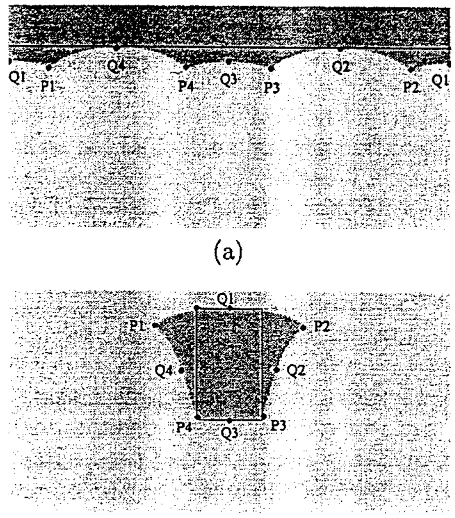


Figure 6: Two image regions after warping are cut into rectangles; (a) the sight direction with azimuth angle (180°, 0°); (b) the sight direction with (180°, 60°).

The warped images after horizontal cutting or vertical cutting mentioned above will be processed in the later image mosaic algorithm.

3 Design of Picture-Taking Manner

The spherical panorama in this paper is built from photographic images captured from a standard camera with an ordinary lens. Therefore, before we take pictures, we should design the picture-taking manner. The number of pictures sufficing to build a panorama will depend on the film width, film height, the focal length of the camera lens, and the overlapping ratio between two contiguous pictures to be used in the image registration.

Similarly, let the focal length be designated as "*f*", the overlapping ratio between contiguous pictures as "*k%*", the film height as "*a*", and the film width as "*b*", where "*a*" is greater than "*b*". We take two respective pictures for sight directions along the north and south poles. If the pictures are taken in the style of landscapes, the least number of circles of pictures needed to be taken, denoted by *NumCirclesLandscape*, will be

$$NumCirclesLandscape = \lceil \frac{180°}{(100-k)\% * [2\tan^{-1}(\frac{a}{2f})]} - 1 \rceil. \quad (6)$$

On the other hand, if these pictures are taken in the style of portraits, the least number of circles will be

$$NumCirclesPortrait = \lceil \frac{180^\circ - 2\tan^{-1}(\frac{b}{2f}) + [2k\% * \tan^{-1}(\frac{a}{2f})]}{(100-k)\% * [2\tan^{-1}(\frac{a}{2f})]} \rceil. \quad (7)$$

Based on the number of circles predefined, we can determine the elevation angle ϕ of the sight direction of each circle. We take the same number of pictures for each circle for the reason of the algorithm design of the image mosaic.

Before the image stitching, each photographic image is transformed onto the space of *PSEM* and is cut into a rectangular warped image. The width and height of one rectangular warped image also depend on the elevation angle of respective sight direction. The number of pictures of one circle is determined by the focal length, the image width and the overlapping ratio between contiguous pictures. The photographic equipment is not so perfect that the elevation angles of the sight directions of pictures cannot be very accurate. Therefore, we utilize the possible narrowest width of rectangular warped images to derive the number of pictures to be taken for one circle. The width of the rectangular warped image is a function of the elevation angle ϕ. The narrowest width will appears on the first order deviation of the function to ϕ vanishing. If we adopt the style of portrait, the ϕ_n, at which the narrowest width appears, can be derived as follows

$$\phi_n = \tan^{-1}(\frac{2f}{a}). \quad (8)$$

If the overlapping ratio between contiguous pictures is defined to be $k\%$, the number of pictures of one circle, $NumPictures(\phi_n)$, will be

$$NumPictures(\phi_n) = \lceil \frac{360^\circ}{\phi_n * (100-k)\%} \rceil. \quad (9)$$

The pictures, used to build a panorama, can be taken according to above equations.

4 Image Mosaic Algorithm

The proposed algorithm of image mosaic is to stitch sequences of overlapping photographs into a seamless image of a spherical panorama. How to take these pictures has been introduced in the previous section. The image mosaic algorithm contains three major parts: *computing accurate sight directions of pictures*, *adjusting intensities of photographic images*, and *stitch processing*.

In stitching these overlapping photographic images, computing the accurate sight direction of each picture is the first and indispensable stage. Each picture is taken based on the sight direction predefined. There may still exist some errors in the sight directions because the photographic equipment is not so perfected or the equipment is operated improperly. The stage of computing accurate sight directions of pictures also contains three parts: *horizontal (or latitudinal) image registration*, *vertical (or longitudinal) image registration*, and *fine tuning of the sight directions of pictures*. We take the same number of pictures for each circle. The spatial relationships among these sequences of pictures can be described by a 2D array, as shown in Figure 7. Basically, the pictures in the same row are taken

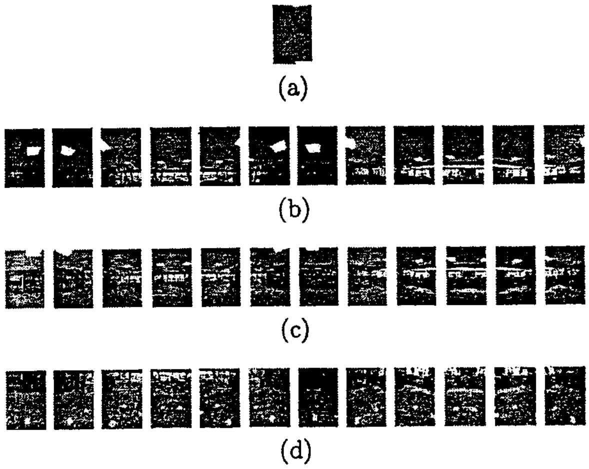


Figure 7: The 37 pictures used to build a spherical panorama in our experiment; (a) the picture with sight direction along the north pole; (b) the 12 pictures of the circle above the equator; (c) the 12 pictures around the equator; (d) the 12 pictures of the circle below the equator.

from the same latitudinal degree, and the same column taken from the same longitudinal degrees. The picture along the north pole is aligned with the leftmost column. To obtain a seamless image of a spherical panorama, we perform a series of image registrations to derive the accurate sight directions of these pictures.

In the 2D arrays of photographic images, the contiguous images have left-right spatial relations in the same row, but have up-down relations in the same column. The image registration, performed on the images in the same row, is named *horizontal image registration*; that performed on the images in the same column is named *vertical image registration*. The *horizontal image registration* is to determine the horizontal relations among these images, that is, to determine the horizontal rotation angles θ's in sight directions. On the contrary, the *vertical image registration* is to determine the vertical relations among these images, that is, to determine the elevation angles ϕ's in sight directions.

The *PSEM* has one useful property mentioned in Section 2: "translation" along the equatorial direction in the map corresponding to "rotation" around the axis of two poles of a sphere. We therefore adopt an image registration algorithm which searches for the best alignment position of two images through a series of translations [2].

The pictures on the equatorial circle have the least variance of elevation angles in sight directions among all circles. Therefore, the *horizontal image registration* is performed on the pictures of the equatorial circle to determine the left-right spatial relationships among these pictures. Before performing the image registration, we transform the photographic images onto the *PSEM*. Assume that image A and B are two warped images through image cutting and have already been registered, as shown in Figure 8. The vector from the center of image A to the center of image B, is designated as $(\Delta x_{BA}, \Delta y_{BA})$. The difference of their sight directions, $(\Delta\theta_{BA}, \Delta\phi_{BA})$, can be derived as follows

$$\begin{cases} \Delta\theta_{BA} = -\Delta x_{BA}/x\_resolution\_PSEM \\ \Delta\phi_{BA} = \Delta y_{BA}/y\_resolution\_PSEM. \end{cases} \quad (10)$$

The accurate sight directions of all the pictures around

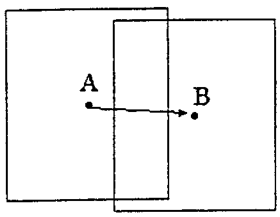


Figure 8: Two images A and B have been registered .

the equator, designed as (θ_e, ϕ_e)'s, will be obtained after the horizontal image registration.

The pictures with different elevation angles in the same column are taken from almost the same longitudinal degree. The errors in longitudinal degrees predefined can be kept within five degrees. However, the elevation angles in sight directions cannot be accurately adjusted by our equipment. The errors in the elevation angles of sight directions may be very large. The accurate sight directions are derived through a series of image registrations. For any two pictures with up-down relation, all possible elevation angles should be taken into consideration to find the result of the image registration. For one photographic image, the computation would be repeated for respective warped images of all possible elevation angles based on Equation 5. It is very time consuming. Therefore, we propose an algorithm to speed up the *vertical image registration*.

The *PSEM* has the useful property mentioned above: "translation" along the equatorial direction in the map corresponding to "rotation" around the axis of two poles of a sphere. Therefore, we utilize a coordinate system transformation such that a vertical image registration in the original coordinate system (R^3) becomes a horizontal image registration in the new coordinate system. In the new coordinate system, the difference of sight directions around the polar axis corresponds to that along the longitudinal direction in the original coordinate system. The result of the image registration, that is, the accurate elevation angles, can be obtained through a series of image shifting and correlation computation. The respective warped images of different elevation angles ϕ's need not be generated anew during image registration. The recomputation of warped images can be avoided. The vertical registration is then dramatically speeded up.

The vertical image registration includes the following four steps.

1. Rotate each photographic image 90° clockwise. This rotation can be accomplished by storing one row of image pixels into a column of another image array indexed in the inverse sequence.

2. Apply "image warping" and "image cutting" mentioned in the previous section to the rotated image by setting the ϕ_L to 90°.

3. Apply a series of horizontal image registrations to them.

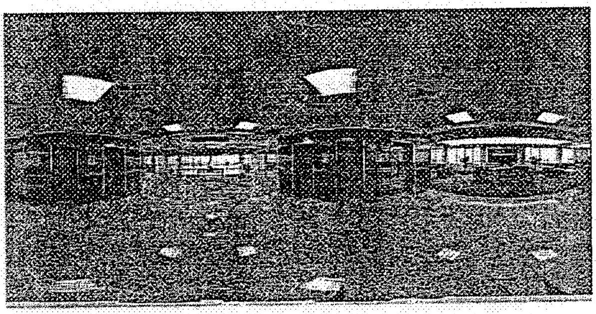
4. Derive accurate sight directions from the image positions of image registration.

Figure 9 shows the leftmost picture of Figure 7(a), (b), (c), and (d), respectively. These pictures have their respect predefined θ's with possible errors within five degrees, but the possible errors of ϕ's may be in a wider range. The four pictures are to be registered along the longitudinal direction. Utilizing the coordinate transformation, we represent the four pictures by the new coordinate system. As shown in Figure 10, they are registered from left to right.

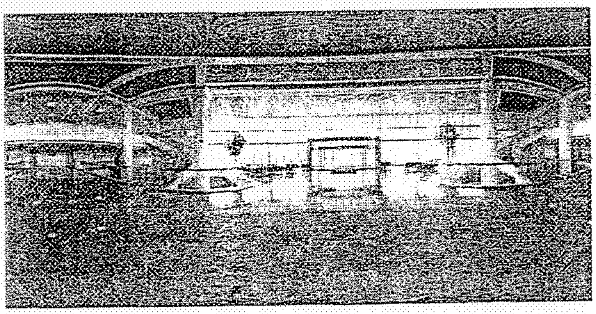
Assume that the sight direction of one picture on the equator is designated as (θ_e, ϕ_e). The contiguous picture with the similar longitudinal degrees above the equator is designated as (θ_a, ϕ_a). Through the coordinate transformation and a horizontal image registration, the difference of the two sight directions in the new coordinate system, denoted by $(\Delta\theta^{(n)}, \Delta\phi^{(n)})$, can be obtained. The sight direction of the picture above the equator can be derived as follows

$$\begin{cases} \theta_a = \theta_e - \Delta\phi^{(n)} \\ \phi_a = \phi_e - \Delta\theta^{(n)}. \end{cases} \quad (11)$$

bottom picture along the south pole would be hidden by the camera tripod. It was replaced by a pattern. Figure 7 shows the other 37 pictures in the first experiment. All of them were processed by the image warping and stitched together as a seamless image of a spherical environment map, as shown in Figure 12(a). In the similar way, we also used another 37 pictures to build the other spherical panorama in the second experiment. The stitched result is shown in Figure 12(b).



(a)



(b)

Figure 12: Two *PSEM* images of spherical panoramas built from 37 pictures, respectively; (a) the result of the first experiment; (b) the result of the second experiment.

6 Conclusions

The paper proposes a practical image-based approach of building spherical panoramas, including the design of the picture-taking manner, the image warping from photographic images to the spherical environment map, and the image mosaic algorithm. A parametric spherical environment map (*PSEM*) is utilized to store the scene data of one spherical panorama. The image mosaic algorithm include three major stages: *computing accurate sight directions of pictures*, *intensity adjusting of picture images*, and *stitch processing*. The algorithm can tolerate errors in sight directions predefined.

Basically, the optical center is assumed to be kept stationary in this paper. The assumption is reasonable when the photographer is not in motion, which is also adopted by most of previous researchers and developers. However, the optical center will move when pictures are taken by a handheld camera or from an airplane in the flight. How to build panoramas efficiently from these pictures and extract perspective 3D scene data from the panorama are very interesting issues and worth investigating in the future.

Acknowledgment

The authors would like to thank the research grant of the Video-based Virtual Reality Project (project no. 3N22400) supported by the Minister of Economic Affairs. The apparatus, designed by using proposed approach, has been filed in application of U.S. and R.O.C. patent.

References

[1] S. E. Chen, "Quicktime VR — An Image-based approach to virtual environment navigation," *ACM SIGGRAPH*, pp. 29-38, 1995.

[2] J. W. Hsieh, *et al.*, "An Intelligent stitcher for panoramic image-based virtual worlds," in application of U.S. Patent, numbered 08/933,758, 1996.

[3] "SmoothMove panorama web builder," Developer Manual (Version 2.0) of $InfinitePictures^{TM}$, Inc., 1996.

[4] R. Szeliski and H. Y. Shum, "Creating full view panoramic image mosaics and environment maps," *ACM SIGGRAPH*, pp. 251-258, 1997.

[5] Ned Greene, "Environment mapping and other applications of world projections," *IEEE Computer Graph and Applications*, Vol. 6, No. 11, Nov. 1986, pp. 21-29.

[6] Zimmermann, Steven D., "Omniview motionless camera orientation system," U.S. Patent, numbered (5,185,667), 1993.