Disparity Morphing and Automatic Generation of Stereo Panoramas for Photo-Realistic Virtual Reality Systems

Ho-Chao Huang and Yi-Ping Hung
Institute of Information Science, Academia Sinica, Taipei, Taiwan, ROC
Email: jet@iis.sinica.edu.tw, hung@iis.sinica.edu.tw

Abstract

Stereo perception is important for a truly immersive virtual reality (VR) system. Recent improvements in computer and video technology have made stereoscopic display systems more prevalent than ever. However, most panoramic imaging systems which can display photo-realistic VR world still have difficulties in automatic production of high-quality stereo panoramas. This paper presents a stereo panoramic imaging system (SPISY) that can automatically generate high-quality photo-realistic stereo panoramas. The major difficulty of generating a stereo panorama from a sequence of stereo images taken by two cameras comes from the image disparity between the two adjacent images to be stitched together, which is caused by the inherent restriction that only one of the two cameras can have its lens center passed by the rotation axis. In this paper, we first analyze the image disparity problem caused by the dislocation of the lens center from the vertical rotation axis, and then propose a disparity morphing algorithm to solve it. Other characteristics of the SPISY include the capability of generating complete-focus views and balanced-exposure views, and of correcting the effect caused by camera tilt and epipolar line inconsistency. With the stereo and automatic correction features, the SPISY can easily provide realistic 360-degree panoramic stereo views for image-based VR systems.

1 Introduction

Due to the capability of generating photo-realistic views, image-based virtual reality (VR) approach[1, 2, 3, 4, 5] has become an important mechanism for constructing the VR world. This paper deals with automatic generation of stereo panoramic views for image-based VR systems. There are several commonly used method for creating panoramic images. The first one is to use panoramic cameras, e.g., Roundshot, Cirkut Camera, Widelux, Hulcharama, etc. Panoramic cameras have the capability of capturing a 360-degree panoramic image without stitching multiple images. The major advantage of using panoramic cameras is that both the acquisition of panoramic pictures and the creation of panoramic views are easier. Recently, BeHere Co. announced the Portal S1 System which could capture a seamless 360-degree panoramic image with one camera shot. However, panoramic cameras are still relatively expensive and are not yet popular for VR world building.

A more popular way for producing panoramic views is to use the images taken by general-purpose off-the-shelf cameras. Several panoramic imaging systems, such as QuickTime VR (from Apple) and Real VR (from Real Space) can generate a panorama by stitching together a sequence of overlapped shots acquired with a 35mm film camera and a tripod system. Alternatively, video or digital cameras can be used. The panoramic imaging system stitches the overlapped shots on a cylinder or on a sphere and generates a 360-degree panoramic image accordingly. The advantage of this approach is that general-purpose cameras are inexpensive and widely available. However, to simplify the generation of high-quality seamless panoramas, one needs to make sure that the rotation axis, when capturing a sequence of overlapped shots, passes through the lens center of the camera.

A different method is to use a general-purpose camera together with a 180-degree fish-eye lens, which can generate a panoramic image with only two snapshots, as was done by Omniview's PhotoBubbles. Another possibility is to use a video camera to take surround views (without a tripod system), and stitch the surround views to a sphere-like polyhedron panoramic image with some authoring tools[5].

In spite of all the advantages of using panoramas to model the VR world, there is one major deficiency — lack of stereo views. Stereo perception is quite important for a truly immersive VR system. Recent improvements in computer and video technology have made stereoscopic display systems more and more prevalent[6, 7, 8, 9, 10]. The PanDC system[1] has tried to use two nearby panoramas to generate a front stereo view and a rear stereo view, each with 120-degree viewing angle. This approach can easily generate a partial stereo panorama for some restricted viewing angles, but can not generate a complete stereo panorama.

In this paper, we propose a method for automatic generation of 360-degree stereo panoramas using a sequence of stereo images acquired by two cameras. A stereo panoramic imaging system (SPISY) has been implemented by using the proposed method, and is briefly introduced in section 2. Details of the method used in the SPISY is described in sections 3 and 4. Since only one of the two cameras can have its lens center passed by the vertical rotation center when acquiring the surround images, the stitching of adjacent images can not be made seamless for the other camera not having its lens center passed by the rotation axis. In section 3, we first analyze the image disparity problem caused by the dislocation of the lens center from the vertical rotation axis, and then propose a disparity morphing algorithm to solve this problem successfully. As demonstrated in our experimental results (which can also be seen in our home page), the proposed disparity morphing algorithm can generate high-quality stereo panoramas automatically. Another major contribution of this work is to propose an aperture moderation algorithm for automatically generating a well-exposed panorama from several panoramas having different aperture settings, which is described in section 4.3.

2 The Stereo Panoramic Imaging System (SPISY)

This section briefly describes the organization of the SPISY, and states some problems which should be carefully dealt with. We shall describe our solutions for solving those problems used in the SPISY in the following sections.

[1] PanDC is a product of Orphan Technologies, and was announced in ACM SIGGRAPH'96

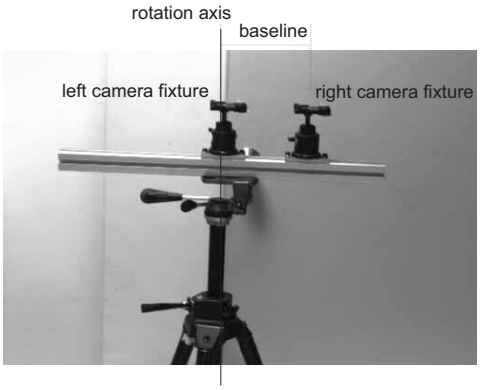

Figure 1: The camera rig used for the SPISY.

In order to generate stereo views, two cameras are used for simultaneous image acquisition in the SPISY. The camera rig used in our experiments is shown in figure 1. There are two possible setups for acquiring the stereo images. One is to have the vertical rotation axis passing through the center point between the two cameras, and the other is to have the rotation axis passing through the lens center of one camera (either left or right). The former setup can produce less distortion for each panorama, but none of two panoramas are perfect. The latter setup produces more distortion for the panorama obtained by the camera not locating at the rotation axis, but the panorama obtained by the camera locating at the rotation axis is much clearer and can be used as a reference in stitching the other panorama. In the SPISY, we let the vertical rotation axis pass through the lens center of left camera, and can use the parameters estimated from the stitching of the left-eye panorama as a reference for stitching the right-eye panorama.

In the SPISY, stereo image pairs are taken when the stereo camera set is rotated horizontally, with 15 degrees per step (or 30 degrees per step if an wider angle lens is used), where the rotation axis roughly passes through the lens center of the left camera. After a sequence of stereo image pairs are acquired, the SPISY performs a series of image processing procedures on those images to generate a stereo panorama. Figure 2 shows the block diagram of the SPISY. Detailed description of each block in figure 2 is given in section 4.

The major difficulty encountered in automatic generation of stereo panoramas is the image disparity problem caused by the dislocation of the camera when the stereo camera set is rotated. In the next section, we will analyze the disparity problem and propose a disparity morphing technique to solve it. Other details of the SPISY will be described in section 4.

3 Disparity Morphing

Since the SPISY uses two cameras with a horizontal stereo baseline to capture the stereo images while there is only one vertical rotation axis, at least one camera cannot have its lens center locating on the rotation axis. In the SPISY, we let the rotation axis pass through the lens center of the left camera, and hence, the right camera can not have its lens center locating

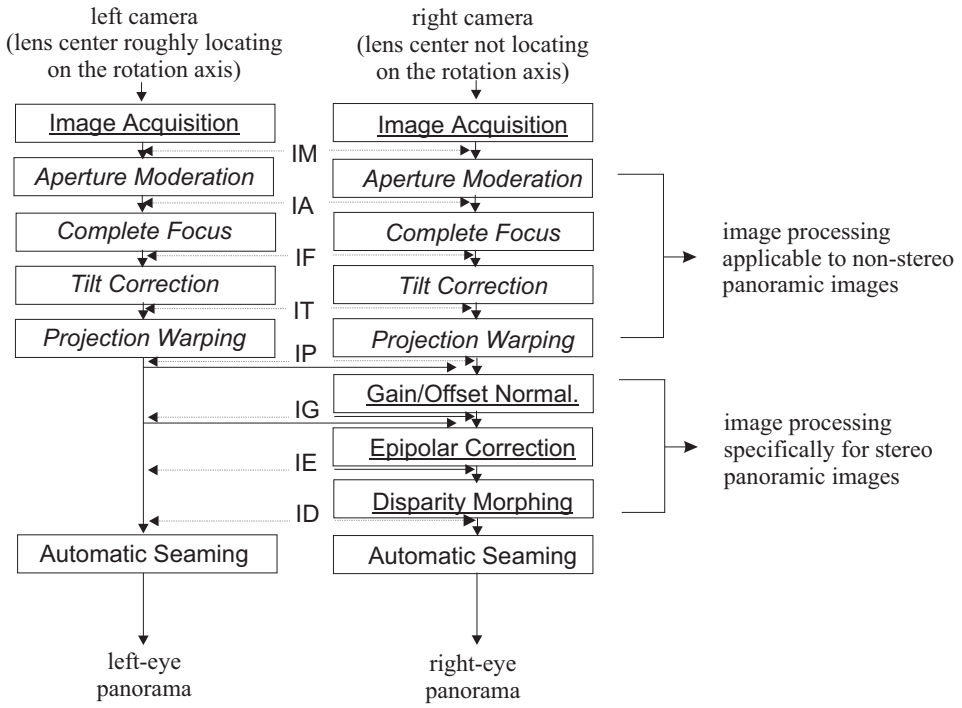

Figure 2: The block diagram of the SPISY. The underlined blocks are the procedures specifically designed for the generation of high-quality stereo panorama, and the italic blocks are the procedures which is also applicable to the generation of non-stereo panorama.

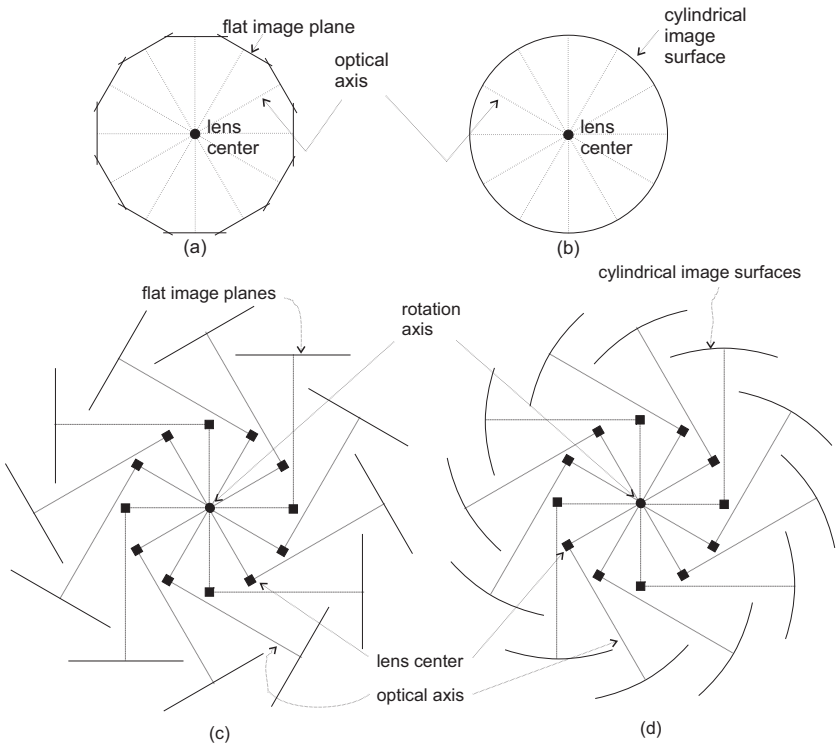

Figure 3: (a) When the rotation axis passes through the lens center of the camera (in our case, the left camera), the image planes of the surround images are rotated by 30 degrees per step. (b) After performing the projection warping on the surround images by projecting those images onto a cylindrical surface, the overlapping areas of adjacent images can match perfectly. (c) If the rotation axis does not pass through the lens center of the camera (in our case, the right camera), the image planes of the surround images will also be rotated by 30 degrees per step. (d) After performing the projection warping on surround images by projecting each image onto a cylindrical surface, the overlapped areas of adjacent images can not match perfectly due to the dislocation of the lens center caused by the rotation of the camera set.

on the rotation axis at the same time. This makes the seaming of the right panoramic images difficult to be done automatically, especially when there are some nearby objects locating on the seaming boundary.

When the rotation axis passes through the lens center of the camera, just as the camera setup recommended by QuickTime VR and the setup of left camera of the SPISY, the position of the lens center will not be changed during the images acquisition procedure. Figure 3(a) shows the lens center and the projection planes for each direction of image-taking. After performing the projection warping on the surround images by projecting those images onto a cylindrical surface, as shown in figure 3(b), the overlapping areas of adjacent images can match perfectly. All the objects appeared on the overlapping area will be of the same size and same spatial relation in the two adjacent images. Thus, panoramic imaging systems can easily stitch those surround images to generate a clear panorama.

If the rotation axis does not pass through the lens center of the camera, as the right camera

of our SPISY, the position of the lens center will vary during the acquisition process, as shown in figure 3(c). Even if we perform the projection warping by projecting those image onto the cylindrical surface, as shown in figure 3(d), neither are the projection centers (lens center) at the same position, nor do the projection surfaces constitute the same cylindrical surface. As a result, the objects appeared in the overlapping area will not be of the same size, and the spatial relationships of those objects will be inconsistent in the two adjacent images containing those objects. We must deal with such image disparity problem in order to generate a clear panorama.

Our goal is to construct two panoramas, the left-eye panorama and the right eye panorama, from a sequence of stereo images so that the left-eye view and the right-eye view can be easily generated by the method used in the widely-available panoramic player, such as QuickTime VR, Surround Video, and Real VR. It is relatively easy to stitch a sequence of overlapping images when the rotation axis passes through the lens center[11]. Therefore, there is few difficulty in constructing a seamless and clear left-eye panorama. To construct the right-eye panorama automatically, we propose a disparity morphing algorithm described below. Figure 4(a) shows two consecutive positions of the right camera, and their corresponding cylindrical image surfaces, where R_c is the intersection point of the rotation axis and the rotatable bar of tripod, and O_1 and O_2 are positions of the lens center for acquiring those two images. Let b be the length of the stereo baseline between R_c and the lens center of the right camera, θ be the rotation angle between consecutive image acquisition, and I_1 and I_2 be the adjacent images projected on two corresponding cylindrical surfaces.

Let the overlapping areas of I_1 and I_2 be IO_1 and IO_2, respectively. Next, project each overlapping area of the cylindrical images onto a corresponding flat image plane, as shown in figure 4(b). Let IF_1 and IF_2 be the overlapping images obtained by projecting IO_1 and IO_2 onto two parallel flat image planes perpendicular to the motion baseline $\vec{O_2O_1}$, respectively. The motion baseline $\vec{O_2O_1}$ can be treated as the viewing direction or the optical axis of the two images, IF_1 and IF_2. Let the intersection points of this motion baseline $\vec{O_2O_1}$ with IF_1 and IF_2 be FOE_1 and FOE_2, respectively, as shown in figures 5(a) and 5(b). That is, FOE_1 and FOE_2 are the focus of expansion (or the epipolars) of the images IF_1 and IF_2. For any 3D object point P, the 3D plane passing through O_1, O_2, and P, is referred to as the epipolar plane corresponding to P. The epipolar line corresponding to P in IF_1 is defined to be the intersection of the image plane of IF_1 with the epipolar plane corresponding to P, i.e., the line passing through FOE_1 and p_1. Similarly, the epipolar line corresponding to P in IF_2 is the line passing through FOE_2 and p_2. The image positions of p_1 in IF_1 and p_2 in IF_2 depend on the distance of the object point P. However, from epipolar geometry[12, 13], the focus of expansion is fixed and all the epipolar lines in IF_1 and IF_2 must intersect at FOE_1 and FOE_2, respectively. Suppose a camera moves from O_1 to O_2 while maintaining its optical axis at the direction of $\vec{O_2O_1}$, the image sequence we observe can be considered as a morphing sequence from IF_1 and IF_2. During the morphing, all the objects move away the focus of expansion, as can be seen in figure 5(b). Because we are approximating the curved trajectory of the local camera motion from O_1 to O_2 by a straight trajectory, there is no rotational camera motion between IF_1 and IF_2. Hence, IF_1 and IF_2 can be aligned together with respect to their focus of expansion, FOE_1 and FOE_2. The speed of each object moving along the epipolar line

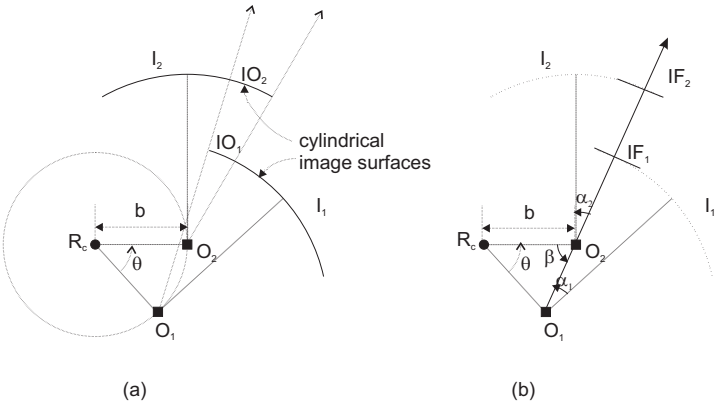

Figure 4: (a) Two consecutive position of the right camera, and the cylindrical image surfaces. (b) Projection of cylindrical images, I_1 and I_2, to flat images, IF_1 and IF_2.

depends on the distance of the object from the camera, which is unknown in general. It is well known that, as the camera moves from O_1 to O_2, a closer 3D object point will move faster away from the focus of expansion.

In order to find FOE_1 and FOE_2 for determining the moving directions of objects in IF_1 and IF_2, we must estimate α_1, the angle between the motion baseline and the optical axis for capturing I_1, and α_2, the angle between the motion baseline and the optical axis for capturing I_2, as shown in figure 4(b). It can be easily verified that

$$\alpha_1 = \alpha_2 = \frac{\theta}{2}. \tag{1}$$

Due to the inexactness of camera rotation increment, the true value of θ in equation(1) is not known. However, because the right camera should have the same rotation increment as that of the left camera, which can be estimated when constructing the left-eye panorama, we can use this information and equation (1) to determine the focus of expansion within the overlapping portion of I_1 and I_2. It should be noticed that if the optical axis of the camera is not perpendicular to the stereo baseline, then equation (1) is not valid and the method of determining epipolar geometry by estimating the fundamental matrix, which is described below, can be used. The famous eight-point algorithm can be used to estimate the fundamental matrix if more than eight corresponding points in the two overlapping images are given[12, 13, 14]. Once the fundamental matrix is known, the focus of expansion can be determined easily. The corresponding points can be selected manually. We can also use some robust algorithms to automatically select the corresponding points and determine the epipolar geometry[15]. Figures 6(a) and 6(b) show an example of the result of applying the eight-point algorithm on two adjacent right-eye images.

In the following, we shall give a simple theoretical analysis on the amount of image disparity caused by the dislocation of the lens center from the rotation axis. In figure 5, we redraw the triangle O_1O_2P of figure 5(a). Let P_0 be the point that is the projection of P onto the motion

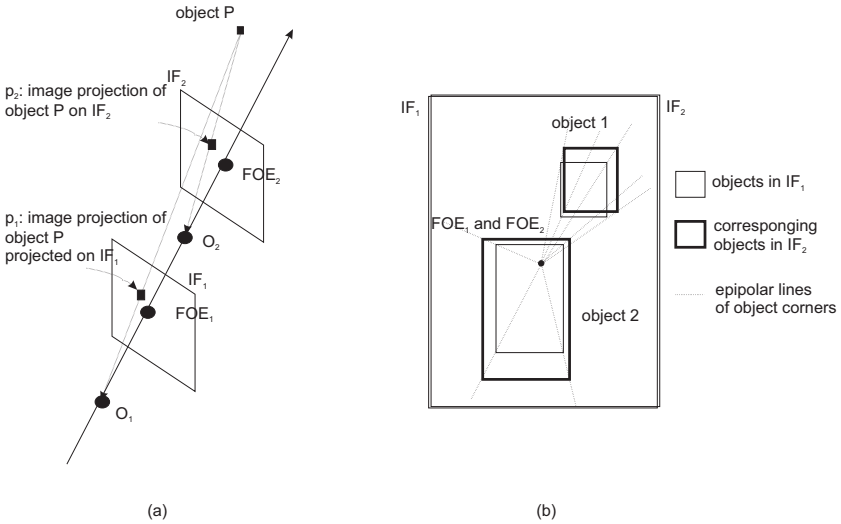

Figure 5: (a) shows the relationships of the focuses of expansion (FOE_1 and FOE_2), and the corresponding object P in the world, projected on IF_1 and on IF_2. (b) shows the focuses of expansion (FOE_1 and FOE_2), and the corresponding objects in IF_1 and IF_2 by image perspective of view.

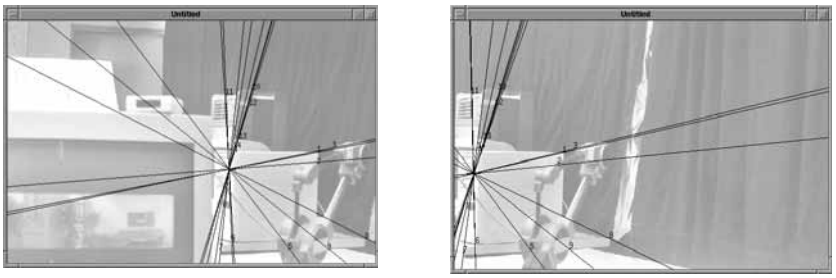

Figure 6: (a) and (b) show the focus of expansion and some epipolar lines on two adjacent right-eye images to be stitched together.

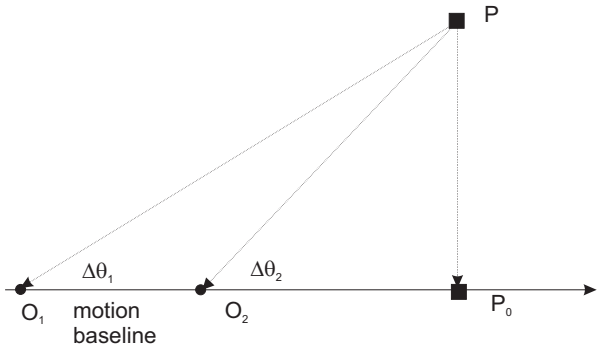

Figure 7: Redraw of the triangle O_1O_2P in figure 5(a), for analyzing the amount of image disparity caused by the dislocation of the lens center.

baseline, $\Delta\theta_1$ and $\Delta\theta_2$ be the angle of $\angle PO_1P_0$ and $\angle PO_2P_0$, respectively. For small $\Delta\theta_2$,

$$\Delta\theta_2 = \tan^{-1}\frac{P\bar{P}_0}{O_2\bar{P}_0} \approx \frac{P\bar{P}_0}{O_2\bar{P}_0}. \tag{2}$$

Also, for small $\Delta\theta_1$,

$$\Delta\theta_1 = \tan^{-1}\frac{P\bar{P}_0}{O_1\bar{O}_2 + O_2\bar{P}_0} \approx \frac{P\bar{P}_0}{O_1\bar{O}_2 + O_2\bar{P}_0} \approx \frac{O_2\bar{P}_0}{O_1\bar{O}_2 + O_2\bar{P}_0} \times \Delta\theta_2. \tag{3}$$

From equations (2) and (3), we can estimate the image disparity D:

$$D = \Delta\theta_2 - \Delta\theta_1 = \frac{O_1\bar{O}_2}{O_1\bar{O}_2 + O_2\bar{P}} \times \Delta\theta_2. \tag{4}$$

For example, if $O_1\bar{O}_2$ is 4 cm, $O_2\bar{P}_0$ is 200 cm, and $\Delta\theta_2$ is 6 degrees, then the disparity distortion D is approximately equal to 0.12 degree, which corresponds to 4.8 pixels, where the pixel resolution is about 0.025 degrees per pixel. The result of the above analysis can be used for setting the search range of the disparity morphing algorithm described below.

Disparity morphing algorithm

Once the epipolar geometry between IF_1 and IF_2 is determined, we can apply any feature-based or region-based algorithm to find the correspondence between IF_1 and IF_2 along the epipolar lines. In the SPICY, we apply the adaptive early jump-out block matching technique with a proper search regions along the epipolar lines. Figures 8(a) and 8(b) show the search regions for block matching on IF_1 and IF_2, respectively.

Let $p_1(i)$ be the i-th pixel on IF_1, the $v_{p_1}(i)$ be the vector from pixel $p_1(i)$ to its corresponding pixel $cp_1(i)$ on IF_2, and s be $O_1\bar{O}_s/O_1\bar{O}_2$ as shown in figure 9(a). We can create an image IF_s at view point O_s by copying each pixel value at $p_1(i)$ to position $p_s(i)$, where $p_s(i)$ is computed by the following equation:

$$p_s(i) = p_1(i) + s \times v_{p_1}(i), \tag{5}$$

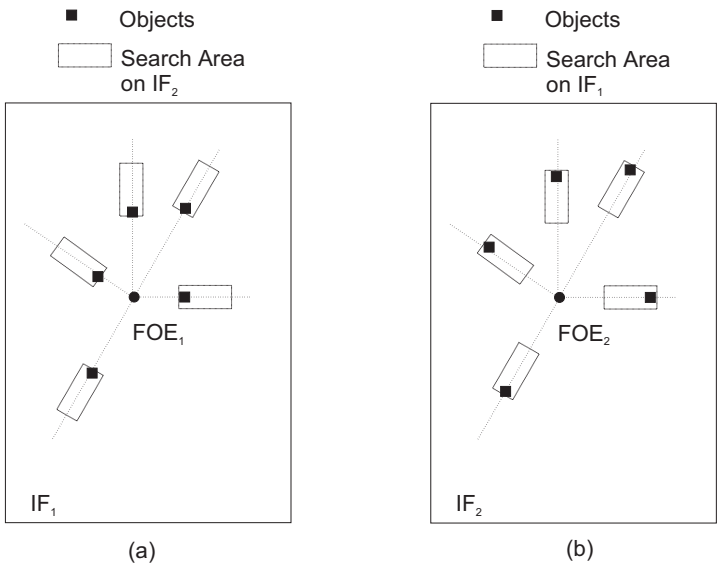

Figure 8: The search regions for block matching on (a) IF_1 and (b) IF_2, respectively.

and then interpolating the pixels not covered by any value in the above procedure.

Assume that there are K vertical scan lines in image IF_1, we can create $K-1$ images IF_s where s is from $1/K$ to $(K-1)/K$ with increment $1/K$, and then compute the disparity morphing image $IF_{morphing1}$, by letting k-th vertical line of $IF_{morphing1}$ equal to the k-th vertical line of $IF_{k/K}$. Figure 9(b) illustrates the resulted image $IF_{morphing1}$.

The same procedure with reverse direction of camera motion can be applied on IF_2 to generate the disparity morphing image $IF_{morphing2}$ as shown in figure 10. Both images $IF_{morphing1}$ and $IF_{morphing2}$ are used in image seaming and image blending described in section 4 to generate the right-eye panorama.

4 Detailed Description of the SPISY

This section describes each component the detail of the proposed SPISY shown in figure 2. In the SPISY, the only given camera parameters are the focal length, fl, and the horizontal and vertical pixel spacings, rx and ry, all of which can be roughly known from the camera and lens specification.

After both the right-eye and left-eye panoramic images are stitched together by using the procedures described in the following subsections, users can see stereo views through a pair of stereo glasses and control the viewing direction through a mouse. The turn-left, turn-right, turn-up, turn-down, zoom-in and zoom-out functions are all available in the SPISY with a natural and user-friendly way.

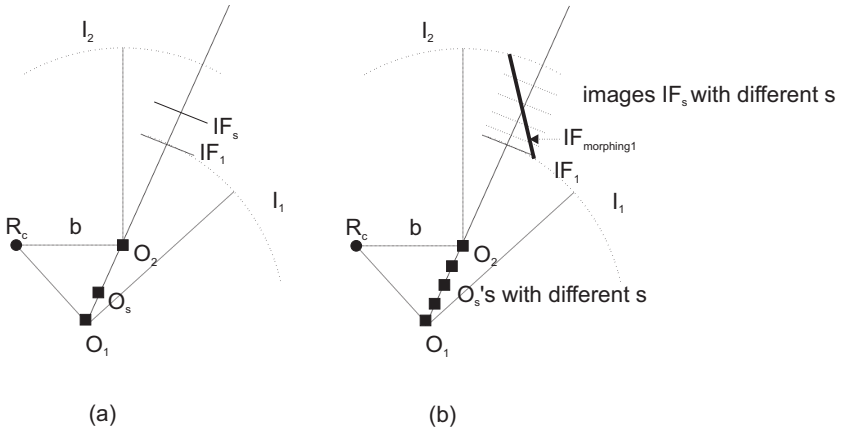

Figure 9: (a) Image IF_s at view point O_s generated from IF_1. (b) The disparity morphing image $IF_{morphing1}$.

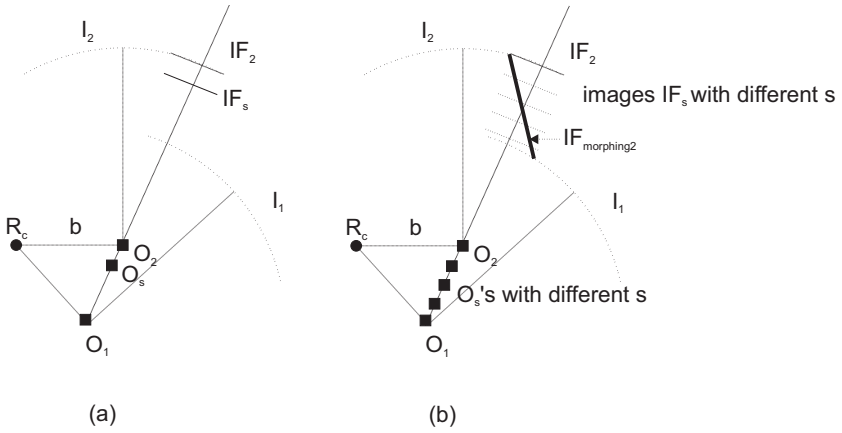

Figure 10: (a) Image IF_s at view point O_s generated from IF_2. (b) The disparity morphing image $IF_{morphing2}$.

4.1 Acquisition of a Series of Stereo Image Pairs

The first step is to capture a sequence of overlapped images. In the SPISY, the stereo images are taken with two digital cameras mounted on a rotational tripod. In our experiments, images are taken when the stereo camera set is rotated horizontally, with 15 degrees per step (or 30 degrees per step when using wide-angle lenses), where the rotating axis roughly passes through the lens center of the left camera. To generate complete-focus views using the technique described in section 4.2, seven image pairs with different focus settings are captured for each rotation step, which are indexed from 0 to 6, where the image pair with index 6 has the nearest focusing distance. In order to generate moderated-aperture views using the technique described in section 4.3, two or four images with different aperture settings are captured for each focus setting. After the stereo images acquisition process, a sequence of images $IM_{(c,a,f,r)}$ are obtained, where the variable c (camera selector) can be $left$ or $right$, the variable a (the index for aperture setting) ranges from 0 to 1 (or 3 if four aperture settings are used), the variable f (the index for the focal length) ranges from 0 to 6, and the variable r (rotation) ranges from 0 to 345 degrees, stepped by 15 degrees.

Notice that, the varying focus setting is not always required. For example, when constructing the outdoor panoramic stereo images with a wide angle lens (i.e., field of depth is relatively large in this case), only one focus setting is enough, and the procedure for generating the complete-focus images, described in the next subsection, can be omitted. The varying aperture setting is not always required, either. For example, when constructing stereo panorama with uniform lighting environment, only one aperture setting is enough, and the procedure for generating the moderated-aperture images, described in section 4.3, can be omitted.

4.2 Generation of Complete-Focus Images

For each camera, each aperture setting and each rotation step, seven images with different focal lengths are captured during the image acquisition process. The SPISY then generates the complete-focus images by selecting the correctly-focused image for each pixel[16]. This process contains two stages, the gradient calculation and the medium filtering. The gradient calculation stage uses equation (6) to calculate the gradients of each pixel.

$$G_{(c,a,f,r)}(x,y) = \sum_{n=-3}^{n=2} |IM_{(c,a,f,r)}(x+n+1,y) - IM_{(c,a,f,r)}(x+n,y)| + |IM_{(c,a,f,r)}(x,y+n+1) - IM_{(c,a,f,r)}(x,y+n)|, \quad (6)$$

where (x, y) is the x and y coordinates of the pixel in the image.

For each pixel, the index for the focal length, f, is selected if its corresponding gradient, $G_{(c,a,f,r)}(x,y)$, is maximum among all focal lengths. The selection function is defined as follows:

$$F_{(c,a,r)}(x,y) = Arg\left\{\max_{f} G_{(c,a,f,r)}(x,y)\right\}. \quad (7)$$

Due to the blurring effect caused by out-of-focus, the result of equation (7) is noisy especially on the boundaries of the objects with different distances. Therefore, a medium filter defined in equation (8) is used to remove the noises near object boundaries.

$$F'_{(c,a,r)}(x,y) = medium_{-5\le i\le 5,-5\le j\le 5} F_{(c,a,r)}(x+i,y+j). \tag{8}$$

Finally, a complete-focus image is generated by using the following equation:

$$IC_{(c,a,r)}(x,y) = IM_{(c,F'_{(c,a,r)}(x,y),r)}(x,y). \tag{9}$$

Figure 15 shows an example of a complete-focus image obtained by using the above process.

4.3 Aperture Moderation

For each camera, each rotation step and each focus setting, several images with different aperture settings are captured during the image acquisition process. The number of images required depends on the lighting variation of environment. In most outdoor panoramas, one proper aperture setting is enough. But for some indoor or outdoor panoramas with highly varying lighting, we need to capture two or more images with different aperture settings to make every things at least clear within one of those images. In the latter case, we need to integrate those images into a balanced-exposure image where all objects can be clearly seen simultaneously.

The SPISY generates a balanced-exposure image by using an aperture moderation process which computes a weighted sum of all the image values observed with different aperture settings for each pixel. This process contains two stages, the index selection stage and the weighted fusion stage.

The index selection stage describes the weights for each pixel in each image, and then selects the aperture index with maximal weight for each pixel:

$$Index_{(c,r)}(x,y) = Arg\{\max_a Weight_{(c,a,r)}(x,y)\}, \tag{10}$$

where $Weight_{(c,a,r)}(x,y)$ is a weighting defined below:

$$Weight_{(c,a,r)}(x,y) = \frac{var_{(c,a,r)}(x,y) + 100}{0.001 + \left|(mean_{(c,a,r)}(x,y) - 128\right|}, \tag{11}$$

where $var_{(c,a,r)}(x,y)$ and $mean_{(c,a,r)}(x,y)$ are the variation and the mean value of pixel $IC_{(c,a,r)}(x,y)$, within a 3 by 3 window, respectively.

The numerator of equation (11) implies that we want to select the aperture index that gives higher variation at the pixel, because the over-exposed and under-exposed images usually have low intensity variation. Since the intensity of the over-exposed or under-exposed pixel usually differs more from the middle intensity value (128), the denominator makes the process tend to select the aperture index that gives an intensity value nearer to the middle intensity value.

In highly textured area, the aperture index chosen by equation (10) usually corresponds to the image having higher contrast for the pixel (x,y). Within low textured area, however, the result of index selection stage is unstable. Therefore, we need the weighted fusion stage, described below, to generate a smooth balanced-exposure image.

For each pixel, let $IndexWeight_{(c,a,r)}(x,y)$ be the number of pixels having $Index_{(c,r)}(x,y)$ equal to $Arg\{\max_a Weight_{(c,a,r)}(x,y)\}$ within the window of size $MeanWindow$ by

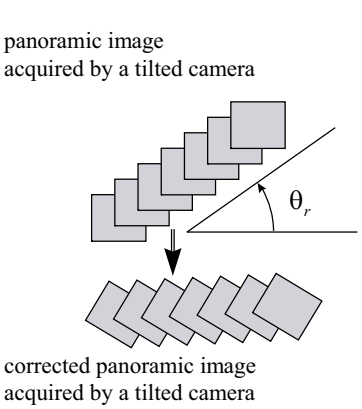

Figure 11: The diagram shows the tilting problem, where the upper part is the panoramic image generated by a tilted camera without correction, and the lower part is the corrected panoramic image.

$MeanWindow$ and centered at (x, y). Then, the balanced-exposure images are generated by using the following equation:

$$IA_{(c,r)}(x,y) = \sum_{a} IndexWeight_{(c,a,r)}(x,y) \times IC_{(c,a,r)}(x,y) \tag{12}$$

Larger $MeanWindow$ will generate smoother balanced-exposure images, but will require more time to compute. In our experiments, $MeanWindow$ is set to 101 where the size of one image shot is 1152×764. Figure 16 shows an example of the experimental results using this aperture-moderation process.

4.4 Tilt Correction

Let the camera coordinate system (CCS) be the coordinate system associated with the camera (or more precisely, the left camera), where its x and y axes are aligned with the horizontal and vertical axes of the image plane, respectively, and its z axis is pointing toward the optical axis of the camera. Let the world coordinate system (WCS) be the coordinate system associated with the real world to be pictured, where its y-axis is the rotation axis with respect to which the stereo camera set is rotating when taking a sequence of overlapped stereo image pairs. Since the SPISY does not require a well-calibrated tripod system, it does not assume the knowledge of the angle between the y-axis of the CCS and the y-axis of the WCS; ideally, the angle should be zero, but it is usually not the case in practice.

Let the angle between the y-axes of the two coordinate systems be θ_r. If we seam the panoramic images without correction, the resulted panoramic image will tilt for θ_r degrees, as shown in figure 11. Therefore, we have developed the procedure described below to correct this tilting effect in order to produce better-quality panoramic images.

The function of the tilt correction is as follows.

$$IT_{(c,r)}(x,y) = IA_{(c,r)}(\sqrt{x^2+y^2}\cos(\theta_r + \arctan\frac{y}{x}), \sqrt{x^2+y^2}\sin(\theta_r + \arctan\frac{y}{x})) \tag{13}$$

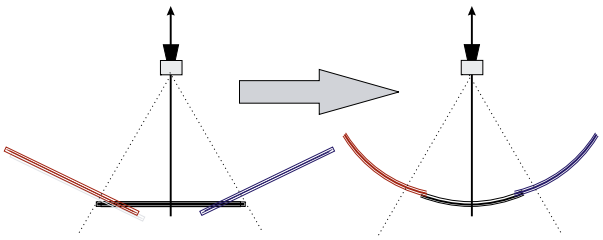

Figure 12: The diagram shows the effect of the projection warping, where the left part shows the image plane before projection warping and the right part shows the result of the projection warping.

Realization of tilt correction

In equation (13), the SPISY needs to know the angle θ_r to correct the tilt distortion. However, since there is no such information in the given camera parameters, the SPISY has to find θ_r by itself. There are two stages in the tilt correction procedure of the SPISY.

The first stage is a registration process. Before correcting the tilting effect, we first uses the hierarchical adaptive early jump-out block-matching algorithm described in subsection 4.9 to register (or pre-seam) the panoramic images IC produced by the complete-focus procedure. However, in this stage we do not have to actually seam the panoramic images together. What we need from the registration process here is the seaming offset, $(\Delta x, \Delta y)$, for each pair of adjacent images. By summing up the $(\Delta x, \Delta y)$'s of all consecutive image pairs we can get the overall offset $(\Delta X, \Delta Y)$. The angle θ_r is then obtained by the following equation.

$$\theta_r = \arctan \frac{Y}{X} \tag{14}$$

The second stage of the tilt-correction procedure is to correct the tilted images by using equation (13), where the θ_r in (13) is obtained by (14).

4.5 Projection Warping

The projection warping procedure corrects the distortion caused by perspective projection. Since the projection plane of an ordinary camera is a plate, the distances from each pixels to the lens center are different. The distance from the pixels near the image center to the lens center is shorter than that from the pixels near the image boundary to the lens center. Thus, an object projected to the center portion of the image will look smaller than that projected to the boundary portion of the image[4].

In order to correct the distortion of the camera projection and eliminate the seaming discontinuity, the following projection warping is performed (refer to figure 12). Here, the projection warping for the left and right images are performed independently.

$$IP_{(c,r)}(x, y) = IT_{(c,r)}(x, \frac{y}{\cos(\arctan \frac{x \times rx}{fl})}) \tag{15}$$

4.6 Gain/Offset Normalization

Since the sequence of stereo images are captured by two different digital cameras, the luminance and chrominance of the stereo images corresponding to the same object may be different, due to the different characteristics of these two cameras. In order to give more comfortable stereo viewing and to correct epipolar-line inconsistency, we should perform some equalization between these two panoramic stereo images. In the SPISY, we perform a global gain/offset normalization.

Let $E_{c,rgb}$ and $V_{c,rgb}$ be the mean and standard deviation of each of the RGB components based on all pixels in the projection-warped image, respectively, where the variable rgb can be RED, $GREEN$ or $BLUE$. The following equation is the function of the gain/offset normalization.

$$IG_{(c,r,rgb)}(x,y) = (IP_{(c,r,rgb)}(x,y) - E_{c,rgb}) \times \frac{V_{left,rgb}}{V_{c,rgb}} + E_{left,rgb}, \tag{16}$$

where $IP_{(c,r,rgb)}(x,y)$ is the rgb component of the image pixel $IP_{(c,r)}(x,y)$. Let $IG_{(c,r)}(x,y)$ be the output color images of the gain/offset normalization process. After the gain/offset normalization process, the mean and standard deviation of the stereo image pair will be the same. This is a preliminary step for the epipolar correction process described in the next subsection.

4.7 Epipolar Correction

Because the tripod used in the SPISY is allowed to be not very precise, the left and right cameras may not be on the same horizontal plane, and hence the epipolar lines of the two images may not correspond to the same scan line. For example, if the left camera is higher than the right camera, the object projected on the left image will be lower than that on the right image. Figure 13 illustrates the epipolar line problem. Figure 13(a) shows a tilted tripod, on top of which the two cameras are not on the same horizontal plane. Figures 13(b) and 13(c) are the images captured by left and right cameras, respectively. The car in the left image is lower than that in the right image. If the epipolar-line inconsistency problem is not fixed, human beings can not comfortably match the stereo images to form a stereo view.

In order to solve the epipolar-line inconsistency problem, we must first detect the corresponding epipolar lines in the stereo images. Assume the slopes of the left and right epipolar lines are M_{left} and M_{right}, respectively, and the vertical offset of the left and right epipolar lines are O_{left} and O_{right}, respectively. The epipolar correction procedure used in the SPISY adjusts the panoramic images of the right camera to match that of the left camera. The function of the epipolar correction is as follows.

$$\begin{aligned} IE_{(c,r)}(x,y) = IG_{(c,r)}(& \sqrt{x^2+y^2}\cos(M_c - M_{left} + \arctan\tfrac{y}{x}), \\ & \sqrt{x^2+y^2}\sin(M_c - M_{left} + \arctan\tfrac{y}{x}) + O_c - O_{left}), \end{aligned} \tag{17}$$

where M_c is equal to M_{left} or M_{right} for left or right images, respectively, and O_c is equal to O_{left} or O_{right} for left or right images, respectively.

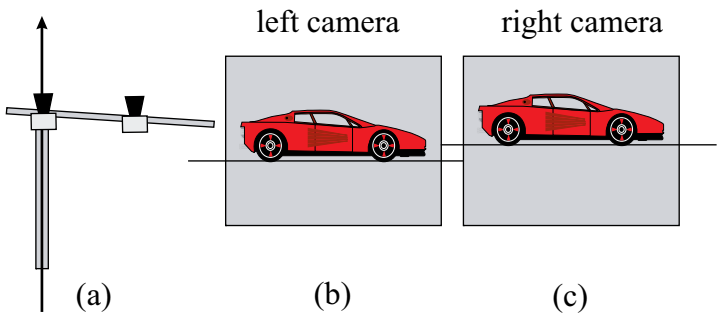

Figure 13: The diagram shows the result of the stereo images without epipolar correction. (a) shows a tilted tripod, where the two cameras are not on the same horizontal plane. (b) and (c) show the images captured by left and right cameras, respectively.

Realization of epipolar correction

It is not trivial to estimate M_{left}, M_{right}, O_{left} and O_{right} from any given panoramic stereo images. Fortunately, after the tilt correction process, we can assume M_{left} is equal to M_{right}, and the only difference of left and right epipolar line is the vertical position. The vertical differences of the epipolar lines pairs for each stereo images pairs are calculated first in the epipolar correction procedure before determine O_{left} and O_{right} for the 360° panoramic stereo images. The block matching technique can be applied to find the vertical offset of two epipolar lines in each stereo images pair. For each pair of stereo images, the template for block matching is chosen to be at the center of the right image, with its size of 1/3 by 1/3 of the image size, and the search range for horizontal and vertical directions is of $\pm 1/3$ of the image width and $\pm 1/14$ of the image height, respectively.

Let EO_r be the vertical difference of the epipolar lines for stereo image pair at camera orientation r. To eliminate the effect of noise, we take the medium value, EO_m, of the EO_r sequences as the global vertical difference for all image pairs. Finally, the function of the epipolar correction can be rewritten as follows.

$$IE_{(left,r)}(x, y) = IG_{(left,r)}(x, y) \tag{18}$$

$$IE_{(right,r)}(x, y) = IG_{(right,r)}(x, y + EO_m) \tag{19}$$

4.8 Disparity Morphing

Section 3 has described the disparity morphing technique used to solve this disparity problem. The resulted image of the disparity morphing algorithm contains two parts. One is the non-overlapped area, which is directly copied from the epipolar-corrected image. The other one is the overlapped area, which is obtained from $IF_{morphing1}$ and $IF_{morphing2}$ by using the disparity morphing technique described in section 3. Our disparity morphing technique together with the image blending algorithm described in section 4.9 is really a data-driven (adaptive) image interpolation method which composes the panorama by morphing two adjacent images along the epipolar lines according to "the result of image matching" (i.e. disparity) that reveals 3D

range information.

4.9 Automatic Image Seaming

Finally, all the warped images should be stitched together to generate the left-eye and right-eye panoramas. Since the camera position and orientation associated with each image is not assumed to be known exactly, the SPISY adopts a hierarchical adaptive early jump-out block matching algorithm to solve the image registration problem. Once the image registration problem is solved, the task of image seaming can be completed by using an image blending algorithm.

Brief review of the adaptive early jump-out technique

In [17], we has proposed an adaptive early jump-out technique for fast motion estimation and template matching. Early in 1972, Barnea and Silverman [18] introduced a class of sequential similarity detection algorithms for expediting the similarity detection between two structured data sets. Their contribution was to propose a monotonically-increasing threshold sequence algorithm where a threshold sequence could be defined such that if, at any accumulation stage in the computation of the MSE or the MAE, the partial result was greater than the corresponding threshold in the sequence, one could jump out of the similarity test. Recently, Cooper et al. applied this sequential algorithm to the dissimilarity test in corner detection, and called it the Early Jump-Out (EJO) technique [19].

The calculation of mean absolute error (MAE) can be written into the following form:

$$MAE(\Delta x, \Delta y) = \frac{1}{n^2} \sum_{i=0}^{n \times n-1} \left| S_c(i) - S_{r(\Delta x, \Delta y)}(i) \right| \tag{20}$$

where $S_c(i)$ represents the value of the i-th pixel in the current image block and $S_{r(\Delta x, \Delta y)}(i)$ represents the value of the i-th pixel in the reference image block with displacement $(\Delta x, \Delta y)$.

Let AE_j be the accumulated sum of absolute errors at step j. That is,

$$AE_j = \sum_{i=0}^{j} \left| S_c(i) - S_{r(\Delta x, \Delta y)}(i) \right|. \tag{21}$$

Let us define the EJO threshold sequence EJS_j, such that if the accumulated error AE_j is greater than EJS_j, the matching process is terminated and a pre-defined large error value is returned. An important issue here is how to determine the EJO threshold sequence EJS_j. It turned out that the threshold sequences determined either by the mathematical early-jump-high models[18, 19] or by some training methods[20] were unsuitable for motion estimation. An adaptive algorithm for determining the EJO threshold sequence was then proposed in [17]. The adaptive technique has the capability to learn the EJO threshold sequence on-line, and can calculate the motion vectors or the block matching offsets two orders faster than the original full range search algorithm.

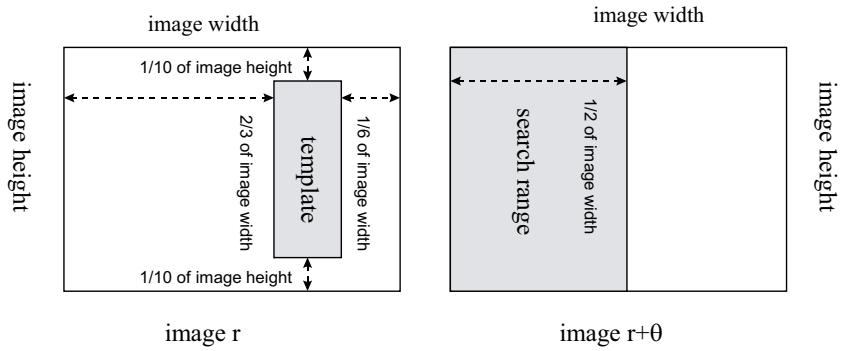

Figure 14: Illustration of the search template and the search area used in the image seaming procedure.

Hierarchical adaptive EJO block matching for panoramic image registration

In this paper, we use block matching to find the seaming offsets for each pair of adjacent images. The adaptive EJO technique[17] is suitable for block matching in the seaming procedure. As shown in Figure 14, the SPISY use a template which has the size of 1/6 of the original image width by 4/5 of the original image height, and is locating at 1/6 of the original image width to the right boundary and 1/10 of the image height to the upper boundary. The search range is chosen to be the left half image area of image $(r + \theta)$ to be registered.

The search range used here is much larger than that used in motion estimation for video compression. Hence, a hierarchical search strategy is proposed to further speedup the searching process. The search step size for both the horizontal and vertical directions is 16 pixels for the first level, and then the search step size is reduced to be 1/2 of the previous one for each following level until the step size is equal to one pixel. The full search area is used for the first level, then the search area is reduced to be 1/2 of the previous one centered at the best match position of the previous level for each following level. The purpose of using the hierarchical search strategy here is not only to reduce the number of matching but also to obtain a good (i.e., tighter) early jump-out threshold sequence as early as possible. By using the hierarchical search strategy, the leaning speed of the early jump-out threshold sequence is much faster, which can then save the matching time significantly.

Image Blending

After the seaming offsets are calculated by using the hierarchical adaptive early jump-out block matching, the adjacent images is seamed with a blending function to make seamless panoramic stereo images. If a pixel of a panoramic image is covered by only one view, the pixel value is equal to the corresponding pixel value of that view. If a pixel of a panoramic image is covered by two adjacent views, the pixel value is the weighted sum of the corresponding pixels of the two adjacent views, where the weight is the offset between the pixel and the left or right boundary of the corresponding image.

(a) (b) (c)

Figure 15: (a) and (b) show two images with different focus settings. (c) shows the complete-focus image.

5 Experimental Results

Several pictures are shown in this section to illustrate the processing and the results of the SPISY. Figures 15(a) and 15(b) show two of the original images captured by the cameras with different focal distances. Since the distances of the objects in the images differ a lot, neither of the two images contain clear images of all objects. For example, in Figure 15(a), the bookshelf is clear but the nine-dot calibration plate is blurred. On the other hand, the bookshelf in Figure 15(b) is blurred while the calibration plate is clearly imaged. Figure 15(c) shows a complete-focus image generated by the SPISY. All objects in the complete-focus image are clear and sharp for subsequent processing.

Figures 16(a) and 16(b) show two of the complete-focus images with different aperture setting. Due to the wide illumination range of the scene, we can not find a proper aperture setting to provide an image where all objects have good contrast. Figure 16(c) is the resulted image of the index selection stage of the aperture moderation procedure, which is quite noisy. Figure 16(d) shows the resulted image of the weighted fusion stage, where all objects are well-exposed.

The processes of applying a series of warping procedures are illustrated in Figures 17–19. Figure 17 shows the result of tilt-correction and projection-warping. Figures 17(a) and 17(b) are two adjacent images of "lobby 1" acquired by the right camera locating at position 1 (r degrees) and position 2 ($r + \theta$ degrees). Notice that the doorknob is not at the same scan line in this image pair. After applying both the tilt-correction and projection-warping procedures, the images of the doorknob and chairs match more accurately, as shown in figures 17(c) and 17(d).

Figure 18 illustrates the effect of the epipolar-correction procedure. Figures 18(a) and 18(b) show a pair of stereo images obtained after applying both the tilt-correction and projection-warping procedures. Figures 18(c) and 18(d) show the stereo images obtained by applying the epipolar-correction procedure to the images shown in Figures 18(a) and 18(b). After the epipolar-correction procedure, the epipolar-line inconsistency between the left-eye and the right-eye panoramic images is largely eliminated.

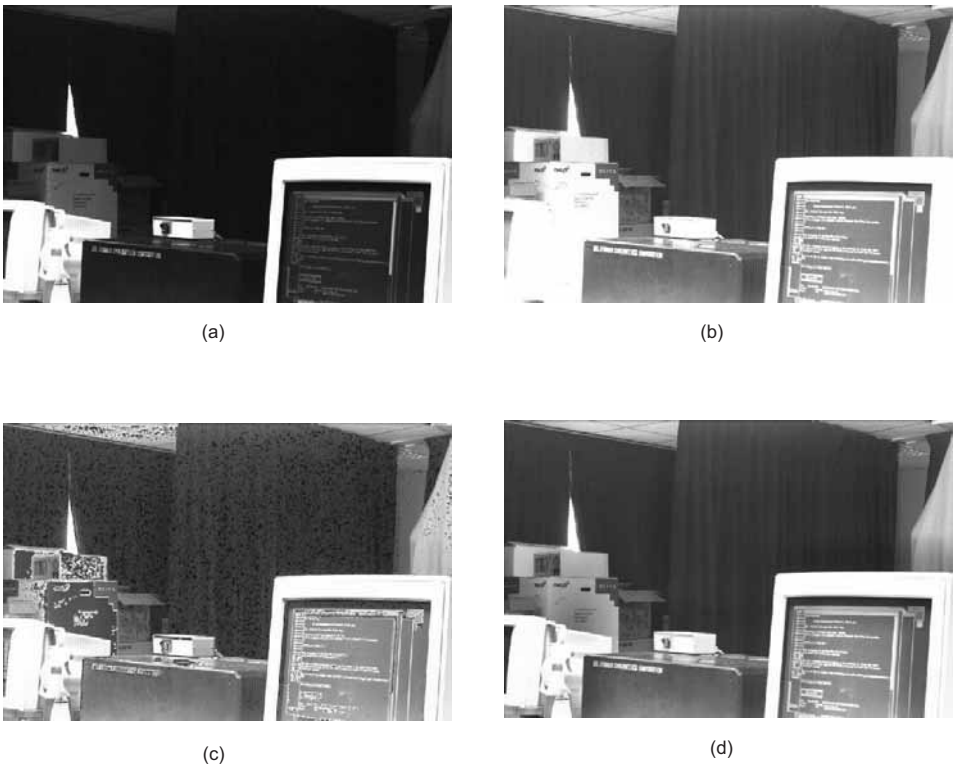

Figure 16: (a) and (b) show two images with different aperture settings. (c) show the resulted image of the index selection stage. (d) show the output image of the aperture moderation procedure.

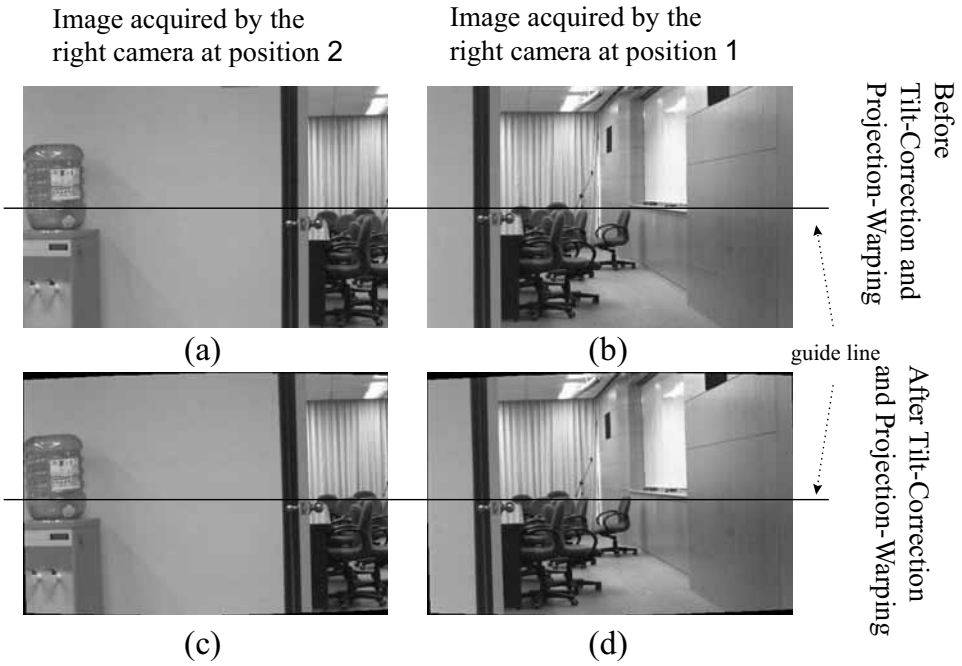

Figure 17: (a) and (b) show the images acquired by the right camera at position 2 and position 1, respectively. (c) and (d) show the result of applying both the tilt-correction and projection-warping procedures to (a) and (b). As one can see by following the guide line in the overlapping area, the images of the doorknob and chairs in (c) and (d) match more accurately than those in (a) and (b).

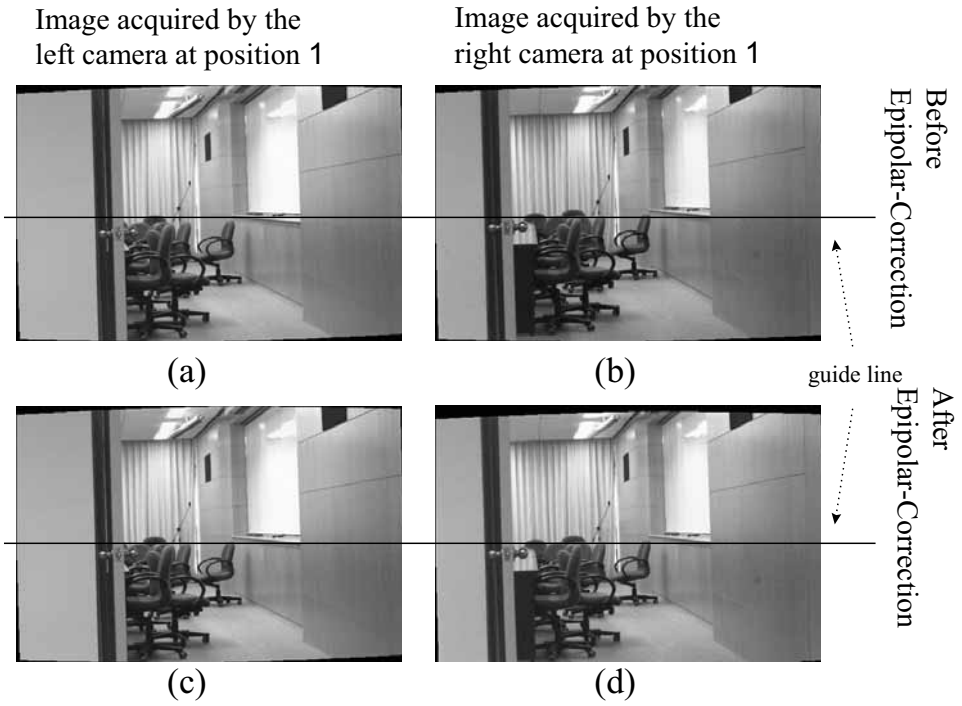

Figure 18: (a) and (b) show the images acquired by the left camera and right camera at position 1, respectively. (c) and (d) show the result of applying the epipolar-correction procedure to (a) and (b). As one can see by following the guide line in the overlapping area, the images of the doorknob and chairs in (c) and (d) match more accurately than those in (a) and (b).

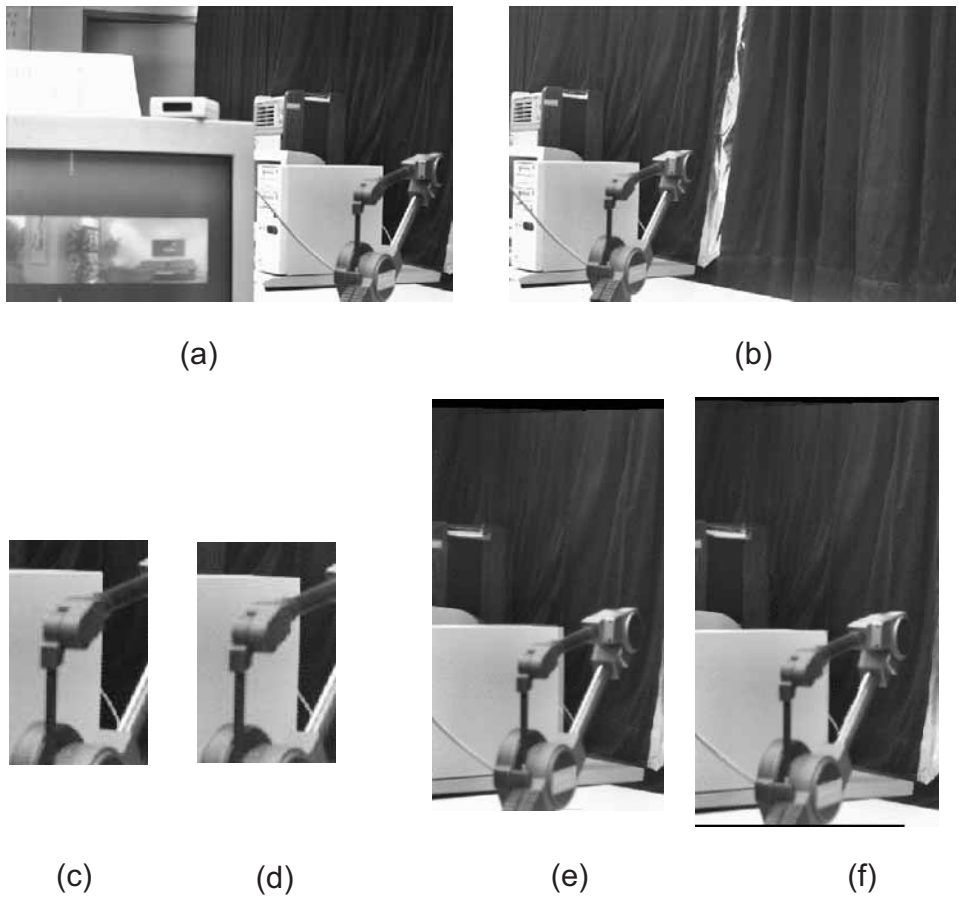

Figure 19: (a) and (b) show the images acquired by the right camera at position 2 and position 1, respectively. (c) and (d) are the zoom-in images of (a) and (b) which show the image disparities of between (a) and (b). (e) is the overlapping portion of the resulted panorama without using disparity morphing. Even with image blending, the object boundaries in image (e) are blurred. (f) is the overlapping portion of the resulted panorama using disparity morphing, where all object boundaries are clear.

Figure 19 displays the effect of the disparity-warping procedure. Figures 19(a) and 19(b) show two adjacent image acquired by the right camera. Figures 19(c) and 19(d) are the zoom-in images of 19(a) and 19(b) to show the different object disparities of objects within 19(a) and (b). Figure 19(e) is the overlapping portion of the resulted panorama without using disparity morphing. Even with image blending, the object boundaries in image (e) are blurred. Figure 19(f) is the overlapping portion of the resulted panorama using disparity morphing, where all object boundaries are clear.

Figure 20 shows four pairs of stereo panoramas automatically stitched together by the SPISY, and Figure 21 shows a picture where a user is interactively viewing the panoramic stereo images generated by the SPISY.

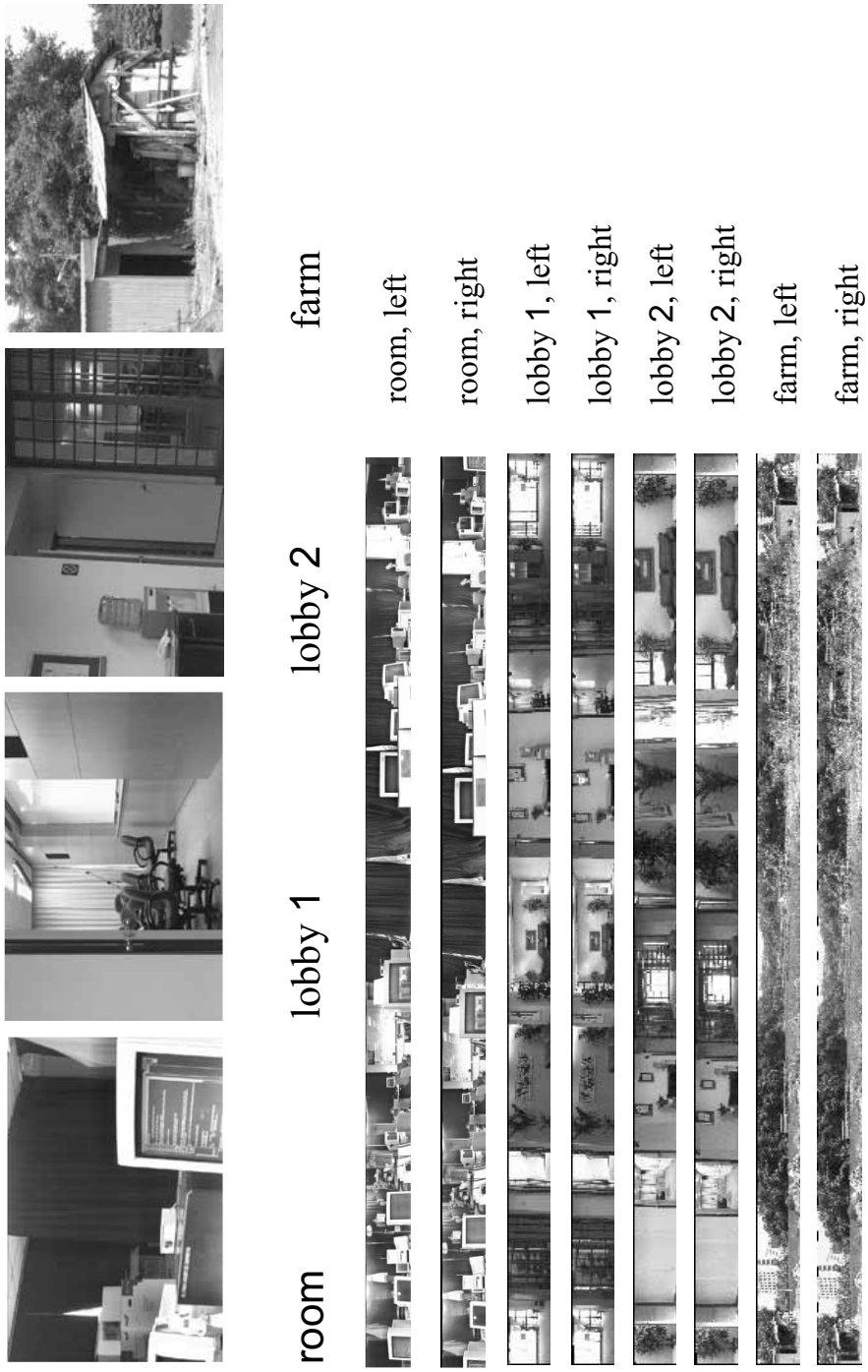

Figure 20: This figure shows four examples of the stereo panoramas generated by the SPISY: "room", "lobby 1", "lobby 2", and "farm". Each of the four images on the left side shows a snapshot of the original images. The eight strips of images on the right are the stereo panoramas obtained by the SPISY.

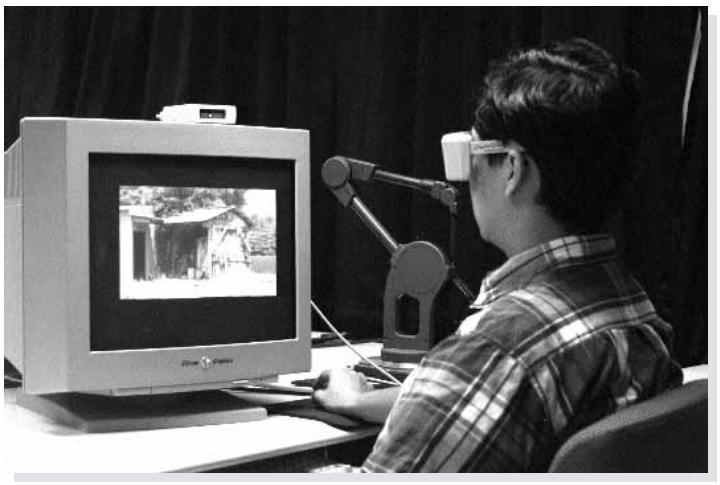

Figure 21: A user is interactively viewing the panoramic stereo images generated by the SPISY.

6 Conclusions

A new stereo panorama imaging system (SPISY) has been presented in this paper. The SPISY can automatically compose a pair of panorama from a sequence of stereo images and provide the viewer full 360-degree panoramic stereo view of an environment interactively. We have shown that the major difficulty of generating a stereo panorama comes from the image disparity between the two adjacent images to be stitched together, which is caused by the inherent restriction that only one of the two cameras can have its lens center passed by the rotation axis. We then analyzed the image disparity problem caused by the dislocation of the lens center from the vertical rotation axis, and proposed a disparity morphing algorithm to solve it. Another major contribution of this work is to propose an aperture moderation algorithm for automatically generating a well-exposed panorama from several panoramas having different aperture settings.

In summary, the SPISY consists of the following nine different modules described in Section 4:
(1) Acquisition of a sequence of stereo images — It does not require a specially-designed or carefully-calibrated tripod system.
(2) Generation of complete-focus images — This module provides complete-focus views everywhere.
(3) Aperture Moderation — This module provides well-exposed views everywhere.
(4) Tilt correction — This module rotates the image to compensate the undesired effect caused by the side-tilting of the rotation axis or of the camera, and is a pre-requisite step for the subsequent steps, such as projection-warping and epipolar-correction. A pre-seaming (or registration) procedure is included in this module.
(5) Projection warping — This module is widely-used in the existing panoramic imaging systems (see [4]) and is especially important when using wide-angle lens.
(6) Gain/Offset normalization — This module compensates the effect caused by different characteristics of the two cameras, which can provide more comfortable stereo viewing and is also

a pre-requisite step for the epipolar-correction module which requires block-matching between the stereo image pair.
(7) Epipolar correction — This module compensates the effect caused by the tilting or the vertical offset between the two cameras, and can provide more comfortable stereo viewing.
(8) Disparity morphing — This module corrects the distortion of image disparity caused by the dislocation of the lens center from the rotation axis. It can provide smoother stitching and more realistic stereo viewing in general. A pre-seaming (or registration) procedure is also included in this module.
(9) Automatic image seaming — This module first determines the seaming offsets for each pair of adjacent images (an image-registration problem), and then stitches together all the registered images into one right-eye panorama and one left-eye 360-degree panorama by using an image blending function. To solve the image-registration problem efficiently and robustly, we proposed a hierarchical adaptive EJO block-matching algorithm which is especially useful for panoramic image registration.

There are still several unsolved problems in developing a panoramic imaging system. For example, moving objects can cause problems for most panoramic imaging systems. This problem can be partially solved by taking multiple images at the same viewpoint, and then integrating them together appropriately. Another important but difficult research topic is on eliminating the feeling of discontinuity caused by "view hopping" in a panoramic imaging system. We are currently working on this problem.

References

[1] G. Miller, E. Hoffert, S.E. Chen, E. Patterson, D. Blackketter, S. Rubin, S. A. Applin, D. Yim, and J. Hanan. The virtual museum: interactive 3D navigation of a multimedia database. *The Journal of Visualization and Computer Animation*, (3):183–197, 1992.

[2] S. E. Chen and L. Williams. View interpolation for image synthesis. *Computer Graphics (Proc. SIGGRAPH'93)*, pages 279–288, 1993.

[3] W.K. Tsao and M. Ouhyoung. An alternative approach of rendering high quality images for virtual environments using scanned images. In *HDTV '95*, pages 7B–1–7B–8, Taipei, Taiwan, ROC, 1995.

[4] S. E. Chen. QuickTime VR—an image-based approach to virtual environment navigation. *SIGGRAPH Computer Graphics Proceedings, Annual Confernece Series*, pages 29–38, 1995.

[5] W.-K. Tsao, J.-J. Su, B.-Y. Chen, and M. Ouhyoung. Photo VR: A system of rendering high quality images for virtual environments using sphere-like polyhedral environment maps. In *Proceeding of Second Workshop on Real-Time and Media Systems, RAMS'96*, pages 397–403, Taipei, Taiwan, ROC, July 1996.

[6] J. Owczarczyk and B. Owczarczyk. Evalution of true 3D display systems for visualizing medical volume data. *The Visual Computer*, 6(4):219–226, August 1990.

[7] D. F. McAllister. 3-D displays. *Byte*, pages 183–188, May 1992.

[8] D. F. McAllister. *Stereo Computer Graphics and Other True 3D Technologies*. Princeton, NJ: Princeton Univ. Press, 1993.

[9] L. F. Hodges and S. W. McWhorter. Stereoscopic display for design visualization. *Sig. Proc.: Image Comm.*, 4(1):3–13, November 1991.

[10] J. Hsu, Zygmunt Pizlo, D. M. Chelberg, C. F. Babbs, and E. J. Delp. Issues in the design of studies to test the effectiveness of stereo imaging. *IEEE Trans. on Systems, Man, and Cybernetics—Part A: Systems and Humans*, 26(6), November 1996.

[11] S. E. Chen. QuickTime VR — an image-based approach to virtual environment navigation. *SIG GRAPH '95*, pages 29–38, 1995.

[12] Q.-T. Luong and O. D. Faugeras. The fundamental matrix: Theory, algorithms, and stability analysis. *Inter. J. of Computer Vision*, 17(1):43–75, January 1996.

[13] Z. Zhang. Determining the epipolar geometry and its uncertainty: A review. INRIA-report, July 1996.

[14] R. I. Hartley. In defence of the 8-point algorithm. In *Fifth Inter. Conf. on Computer Vision*, pages 1064–1070, Massachusetts Institute of Technology, Cambridge, Massachusetts, June 1995.

[15] R. Deriche, Z. Zhang, Q.-T. Luong, and O. Faugeras. Robust recovery of the epipolar geometry for an uncalibrated stereo rig. In *Computer Vision—ECCV'94*, pages 567–576, 1994.

[16] Ens J. E. *An investigation of methods for determining depth from focus*. PhD thesis, The University of British Columbia, July 1990.

[17] H.-C. Huang and Y.-P. Hung. Adaptive early jump-out technique for fast motion estimation on video coding. In *Proceeding of ICPR*, pages 864–868, Vienna, August 1996.

[18] D. I. Barnea and H. F. Silverman. A class of algorithm for fast digital image registration. *IEEE Trans. Comput.*, 21:179–186, 1972.

[19] J. Cooper, S. Venktesh, and L. Kitchen. Early jump-out corner detectors. *IEEE Trans. on Patt. Anal. Mach. Intell.*, 15:823–828, 1993.

[20] A.-T. Tsao, Y.-P. Hung, C.-S. Fuh, and H.-Y. M. Liao. On learning the threshold sequence for the early jump-out template matching. In *TAAI Artificial Intelligence Workshop*, pages 186–193, Taipei, Taiwan, ROC, September 1995.