Panoramic Stereo Imaging System with Automatic Disparity Warping and Seaming

Ho-Chao Huang and Yi-Ping Hung

Institute of Information Science, Academia Sinica, Taipei, Taiwan, ROC

Running head: Panoramic Stereo Imaging System

Correspondence should be sent to:

Professor Yi-Ping Hung

Institute of Information Science

Academia Sinica

Nankang, Taipei, 115

Taiwan

Tel: 886-2-7883799 ext. 1718

Fax: 886-2-7824814 (Attn: Dr. Hung)

Email: hung@iis.sinica.edu.tw

Abstract

Two commonly used approaches for building a virtual reality (VR) world are the model-based approach and the image-based approach. Recently, the image-based approach has received much attention for its advantages of being easier to build a VR model and of being able to provide photo-realistic views. However, traditional image-based VR systems can not produce the stereo views that can give the users the feeling of 3D depth. In this paper, we present a panoramic stereo imaging (PSI) system which can produce stereo panoramas for image-based VR systems. This PSI system is referred to as the PSI-II system, which is an improved system of our previous experimental PSI-I system. The PSI-I system uses a well-calibrated tripod system to acquire a series of stereo image pairs, while the PSI-II system does not require the use of a well-calibrated tripod system and can automatically generate a stereo-pair of panoramic images by using a novel disparity warping technique and a hierarchical seaming algorithm. Our PSI-II system can automatically correct the epipolar-line inconsistency of the stereo images pairs and the image-disparity caused by the dislocation of the camera's lens center in the image acquisition process. Our experiments have shown that the proposed method can easily provide realistic 360° panoramic views for image-based VR systems.

List of Symbols

1. G, F, F'
2. IF, IT, IP, IG, IE, ID, IM
3. POI, POI'
4. n, x, y, s_x, s_y
5. OFF, OFF1, PI, PO, CA
6. MAE, MSE
7. fl, dx, dy
8. r, Δr, θ_r, ϕ, ψ, Δx, Δy

1 Introduction

Virtual reality (VR) has become one of the hottest commercial and research topics[1, 2]. When developing a VR application, building a VR world has been known to be a very time-consuming work. Two commonly-used approach for building a VR world are the image-based approach[3, 4, 5] (eg., Quick-Time VR, Surround Video, RealVR, and IPIX, etc.) and the model-based approach[6, 7] (eg., AutoCAD, 3D Studio, etc.). The model-based approach first constructs the 3-D models of the real world objects and then generates the VR images by rendering the 3-D models. The model-based approach provides good interactivity, but building 3-D models with the model-based approach is a tedious work. On the other hand, the image-based approach builds panoramic views by seaming photographs of the real world directly, and enjoys the advantage of easy and fast VR world building and of having photo-realistic views. However, the image-based approach has some drawbacks in VR world building. For example, it is hard to produce stereo images and hard to travel with arbitrary paths within the virtual reality world.

We have implemented an experimental Panoramic Stereo Imaging system, called the PSI-I system[8], to solve the problem of lacking stereo views in image-based VR systems. In the PSI-I system, the images are acquired with a stereo camera set mounted on a computer-controllable binocular head, i.e., the IIS head[9] shown in figure 1. The well calibrated IIS head provides accurate camera parameters for image warping and seaming. Also, the PSI-I system described in [8] can provide complete-focus stereo views, by using an auto-focusing technique to select the correctly-focused image for each pixel.

However, the IIS head is a complicated experimental equipment, and is not yet available for most people. For practical consideration, we have developed a new Panoramic Stereo Imaging system, the PSI-II system, which can use off-the-shelf cameras without calibration. Since the camera parameters are not assumed to be available in the PSI-II system, a novel automatic seaming algorithm based on image characteristics has been proposed. Furthermore, our PSI-II system can automatically correct the epipolar-line inconsistency of the stereo image pair and the image-disparity caused by the dislocation of the camera's lens center in the image acquisition process. Compared to the traditional image-based VR systems, the PSI-II system is the first system that provides the following five features simultaneously:

1. It can provide 360-degree panoramic stereo views.
2. It does not require special camera rigs, and can tolerate inexactness of camera configuration.
3. It can provide complete-focus views.
4. It can correct the epipolar-line inconsistency and the image-disparity automatically.
5. It can automatically generate a pair of panoramic stereo images by seaming a series of overlapping images .

The organization of this paper is as follows. Section 2 briefly reviews the panoramic imaging system.

Figure 1: The IIS head.

Section 3 describes the proposed PSI-II system in detail. Some experimental results are shown in section 4 and conclusions are given in section 5.

2 Preliminary Review

In recent years there has been increased interest in image-based rendering systems. In [10], the authors proposed the plenoptic function which provides a consistent framework for image-based rendering system. Various image-based methods, such as view morphing[11], are characterized by the different ways they implement the plenoptic function. In [12], the authors described a new light field representation, the light slab, for storing all the radiance values in free space, which was useful for many applications requiring interaction with 3D scenes.

In many image-based rendering systems, such as the one presented by McMillan and Bishop[10], creating panoramic images is a basic and important component. 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 with one shot. The advantage of using panoramic cameras is mainly that both the acquisition of panoramic pictures and the creation of panoramic views are easy. However, panoramic cameras are quite expensive and not popular for VR world builders.

In the past few years, a popular way of creating panoramic images is to use general-purpose off-the-shelf cameras. Many panoramic imaging systems, such as the famous QuickTime VR, the RealVR/PhotoVista and the PanoVR, can create a panoramic image by stitching together a series of overlapping shots acquired with a 35mm film camera and a customized tripod rig. Alternatively, video or digital cameras can be used. Those panoramic imaging systems can stitch the overlapping 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, but the major difficulty is that the panoramic imaging systems must have the capabilities of first warping the input images appropriately and then stitching the images at proper positions of the cylinder or of the sphere. Also, the VR world builder needs to make sure that the rotation

axis, when capturing a series of overlapping shots, approximately passes through the lens center of the camera.

Another 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 shots, as done by IPIX (http://www.ipix.com/). Some companies (e.g., http://www.toyota.com/) have used this technique to advertize their products. In [4], the authors use a video camera to take the surround views, and then stitch the surround views to a sphere-like polyhedron panoramic image with their authoring tool. In all of the methods described above, none provide 360-degree panoramic *stereo* images.

3 Panoramic Stereo Imaging System

This section describes the details of the proposed PSI-II system. In order to generate stereo views, two cameras are used for simultaneous image acquisition. Figure 2 shows the block diagram of the PSI-II system. In the PSI-II system, the only camera parameters that may be needed are the focal length, fl, and the horizontal and vertical pixel spacings, dx and dy.

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 position and orientation through a mouse. The turn-left, turn-right, turn-up, turn-down, zoom-in and zoom-out functions are all available in our PSI-II system in a natural and user-friendly way.

3.1 Acquisition of a Series of Stereo Image Pairs

The first step of image-based VR world building is to capture a series of overlapping images. With our PSI-II system, the stereo images are taken with two digital cameras mounted on a rotational tripod. Figure 3 illustrates the rotation of the stereo camera set.

In our experiments, images are taken when rotating the stereo camera set horizontally, with Δr degrees per step (in our experiments, Δr is set to be 15 because the field of view of the camera used is about 30 degrees), 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 3.2, n image pairs with different focus settings are captured for each rotation step, which are indexed from 0 to $n-1$, where the image pair with index $n-1$ has the nearest focusing distance. In our experiments, n is usually equal to 1 or 2 for taking the outdoor and most of the indoor images. When taking the images of our laboratory, n is selected to be 7 because there are many objects in different distances in our laboratory. After the stereo images acquisition process, a set of images $IM_{(c,f,r)}$ are obtained, where the variable c (camera selector) can be $left$ or $right$, the variable f (the index for the focal length) ranges from 0 to $n-1$, and the variable r (rotation) ranges from 0 to 360 degrees, stepped by Δr degrees.

Notice that, the varying focus setting is not always required. For example, when constructing the outdoor

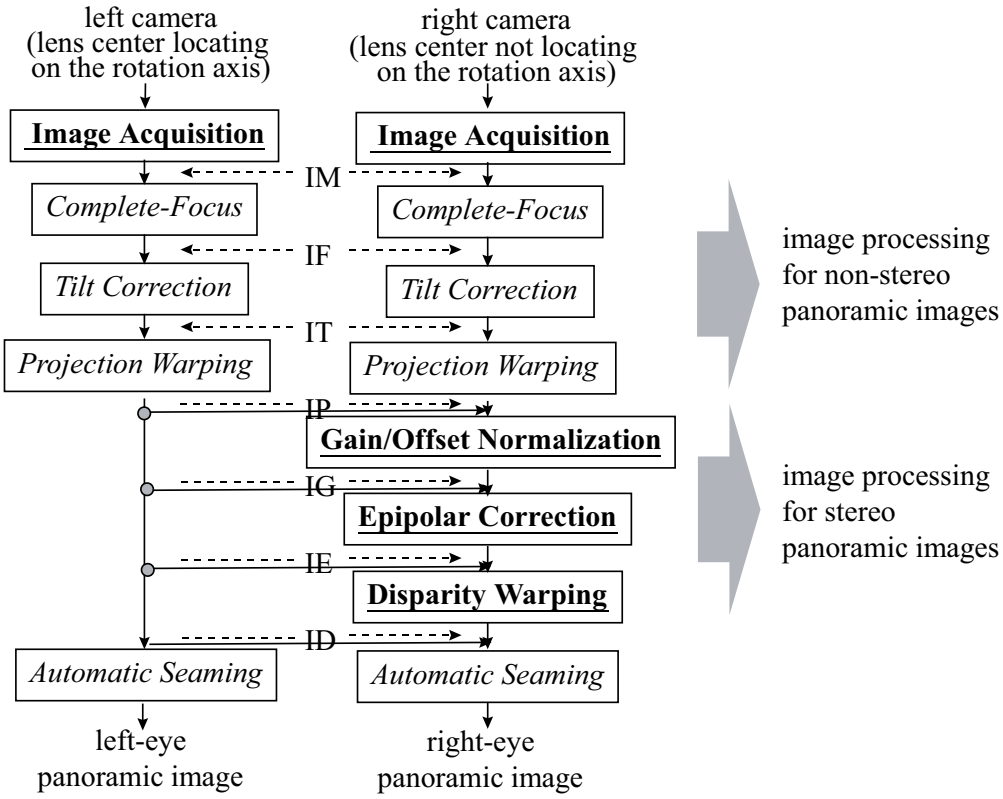


Figure 2: The block diagram of the panoramic stereo imaging system. The underlined blocks are the procedures related to the stereo feature, and the italic blocks are the procedures which are not specifically related to the stereo feature.

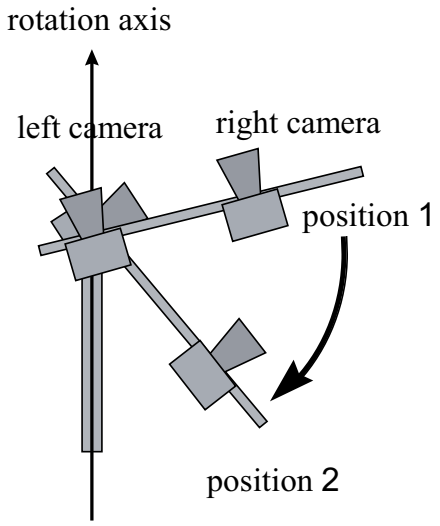


Figure 3: Illustration of the rotation of the stereo camera set.

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 (i.e., $n = 1$), and the procedure for generating the complete-focus images, described in the next subsection, can be bypassed.

3.2 Generation of Complete-Focus Images

For each camera and each rotation step, n images with different focal lengths are captured during the image acquisition process. Our PSI-II system then generates the complete-focus images by selecting the correctly-focused image for each pixel. This process contains two stages, the contrast calculation and the median filtering. The contrast calculation searches for the images (from the n available images of different focal lengths) with the highest contrast (or spatial frequency) in two 7×1 regions (one horizontal and one vertical) about the given pixel. The first step of the contrast calculation stage uses Eq. (1) to calculate the contrast of each pixel.

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

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

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

$$F_{(c,r)}(x,y) = Arg\left\{\max_f G_{(c,f,r)}(x,y)\right\}. \quad (2)$$

Due to the blurring effect caused by out-of-focus, the result of Eq. (2) is noisy especially on the boundaries of the objects with different distances. Therefore, a median filter defined in Eq. (3) is used to remove the noises near object boundaries.

$$F'_{(c,r)}(x,y) = median_{-5\le i\le 5,-5\le j\le 5}(F_{(c,r)}(x+i,y+j)). \quad (3)$$

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

$$IF_{(c,r)}(x,y) = IM_{(c,F'_{(c,r)}(x,y),r)}(x,y). \quad (4)$$

3.3 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

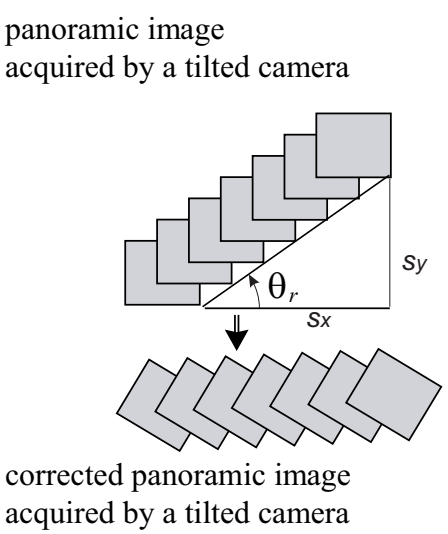


Figure 4: The diagram illustrates the tilting problem, where the upper part is the panoramic image generated by a tilted camera without correction, and the lower part is the corrected panoramic image.

y-axis is the rotation axis with respect to which the stereo camera set is rotating when taking a sequence of overlapping stereo image pairs. Since our PSI-II system 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 4. 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.

$$IT_{(c,r)}(x,y) = IF_{(c,r)}(\sqrt{x^2+y^2}\cos(\theta_r + \tan^{-1}\frac{y}{x}), \sqrt{x^2+y^2}\sin(\theta_r + \tan^{-1}\frac{y}{x})) \quad (5)$$

Equation (5) rotates images $IF_{(c,r)}$ by θ_r, the tilting angle, to generate the tilt-corrected images, $IT_{(c,r)}$, where θ_r is equal to $tan^{-1}(s_y/s_y)$ as shown in figure 4.

Realization of tilt correction

In Eq. 5, the PSI-II system needs to know the angle θ_r to correct the tilt distortion. However, since there is no such information in the given camera parameters, our PSI-II system has to estimate θ_r from the images. Therefore, the tilt correction procedure of the PSI-II system contains two stages. The first stage estimates the tilt angle θ_r and the second stage corrects the tilting effect by using the estimated θ_r.

The first stage is a registration process. Before correcting the tilting effect, we first uses the hierarchical

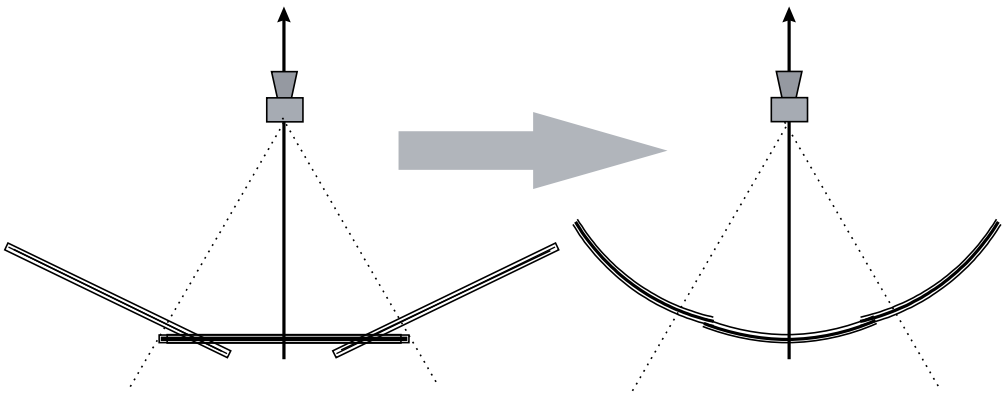


Figure 5: The diagram shows the effect of the projection warping, where the left part shows the image planes before projection warping and the right part shows the result of the projection warping.

adaptive early jump-out block-matching algorithm described in subsection 3.8 to register the panoramic images IF 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 consecutive images. By summing up the $(\Delta x, \Delta y)$'s of all the consecutive image pairs we can obtain the overall offset (s_x, s_y). The angle θ_r is then estimated by the following equation:

$$\theta_r = \tan^{-1} \frac{s_y}{s_x}. \tag{6}$$

The second stage of the tilt-correction procedure is to correct the tilted images by using Eq. (5), where the θ_r in Eq. (5) is obtained by using Eq. (6).

3.4 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 pixel to the lens center are different. The 3D distance from the lens center to a pixel (on the image plane) near the image center is shorter than that from the lens center to a pixel near the image boundary. Thus, an object projected onto the central portion of the image will look smaller than that projected to a portion near the image boundary[3].

In order to correct the distortion caused by perspective projection and to eliminate the seaming discontinuity, the following projection warping is performed (refer to figure 5). Here, the projection warping for the left and right images are performed independently by using Eq. (7):

$$IP_{(c,r)}(x, y) = IT_{(c,r)}(\frac{\tan(x \times dx/fl)}{dx/fl}, y/\cos(x \times dx/fl)). \tag{7}$$

Figure 6 shows the relationship between the image coordinates of IT and IP, where IT is the tilt-corrected image on a *planar* image plane and IP is the projection-warped image on a *cylindrical* image plane. Suppose

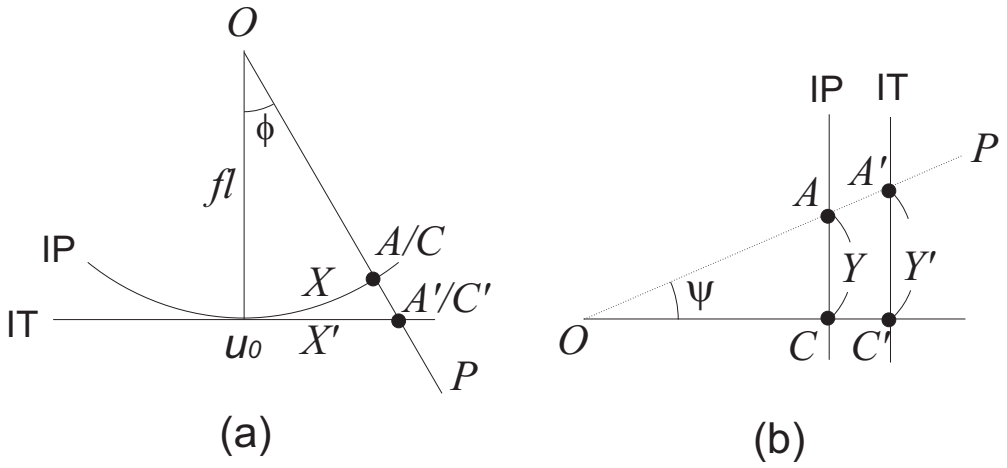


Figure 6: Suppose an object point P is projected onto a planar image plane IT and a cylindrical image plane IP at positions A' and A, respectively. (a) The relationship between the x-components of pixels A' and A. (b) The relationship between the y-components of A' and A.

an object point P is projected onto IT and IP at positions A' and A, respectively. Figure 6(a) illustrates the relationship between the x-components of A' and A. Let x' and x be the horizontal index values of A' and A on images IT and IP, respectively. Then, from figure 6(a), we have $X' = x' \times dx = fl \times \tan\phi$ and $X = x \times d\phi = \phi$, where $d\phi$ is the horizontal pixel size on image IP and is equal to dx/fl radian. Hence

$$x' = x \times \frac{\tan\phi}{\phi} = \frac{\tan(x \times dx/fl)}{dx/fl} \tag{8}$$

because $\phi = x \times d\phi = x \times dx/fl$. Figure 6(b) illustrates the relationship between the y-components of A' and A. Let y' and y be the vertical index values of A' and A on images IT and IP, respectively. Then, from figure 6(b), we have $Y = y \times dy = OA = fl/\cos\theta$ and $Y' = y' \times dy = OA' = OC'/\cos\psi = (fl/\cos\phi)/\cos\psi = fl/\cos\phi\cos\psi$. Hence,

$$y' = y/\cos\phi = y/\cos(x \times dx/fl). \tag{9}$$

From equations (8) and (9), we then have equation (7).

3.5 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. To give more comfortable stereo viewing and to correct epipolar-line inconsistency later, we should perform some equalization between these two panoramic stereo images. In the PSI-II system, we perform a global gain/offset normalization.

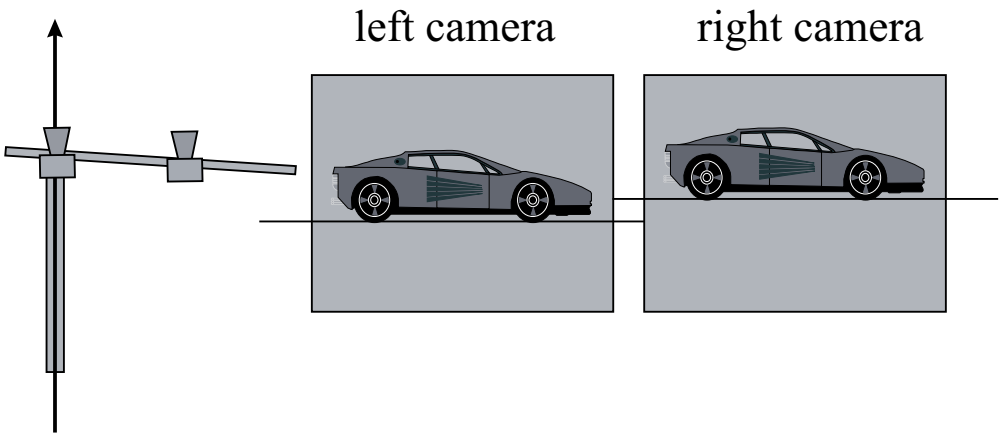


Figure 7: Illustration of a pair of stereo images without epipolar correction. (a) shows a tilted tripod, where the two cameras are not on the same horizontal plane. (b) and (c) show the images captured by left and right cameras, respectively.

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 images, respectively, where the variable rgb can be RED, $GREEN$ or $BLUE$. The following equation is the function of the gain/offset normalization.

$$IG_{(c,r,rgb)}(x,y) = (IP_{(c,r,rgb)}(x,y) - E_{c,rgb}) \times \frac{V_{left,rgb}}{V_{c,rgb}} + E_{left,rgb}, \tag{10}$$

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.

3.6 Epipolar Correction

Because the tripod used in our PSI-II system 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 left and right 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 projected on the right image. Figure 7 illustrates the epipolar-line inconsistency problem. Figure 7(a) shows a tilted tripod, on top of which the two cameras are not on the same horizontal plane. Figures 7(b) and 7(c) illustrate the images captured by the 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, users can not comfortably match the stereo images to form a stereo view.

To solve the epipolar-line inconsistency problem, we must first determine the corresponding epipolar lines in the stereo images. However, it is not a trivial problem to determine the epipolar geometry correctly and reliably[14, 15]. It is fortunate that in our case, after the tilt correction process, one can assume the epipolar

lines in the stereo images are horizontal, that is because the stereo images are roughly in parallel. Thus, the only difference between the left and right epipolar-line is the vertical position. The vertical differences of the epipolar-line pairs for each stereo images pair are calculated first in the epipolar correction procedure. 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 being 1/3 by 1/3 of the image size, and the search range for horizontal and vertical directions is $\pm 1/3$ 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 median 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.

$$IE_{(left,r)}(x, y) = IG_{(left,r)}(x, y) \tag{11}$$

$$IE_{(right,r)}(x, y) = IG_{(right,r)}(x, y + EO_m) \tag{12}$$

3.7 Disparity Warping

Since the PSI-II system uses two cameras to capture the stereo images simultaneously while there is only one rotation axis, at least one camera cannot have its lens center locating on the rotation axis. In the PSI-II system, 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 on the rotation axis at the same time. This will make the seaming of the right panoramic images difficult to be done automatically, especially when there are some nearby objects locating on the seaming boundary. As illustrated in figure 8(c), when the right camera is at position 1, object point 1 and object point 2 are on the same projection line. But when at position 2, object point 1 and object point 2 will be projected to different positions on the right image. That is, if object point 1 and object point 2 are not of the same distance, the relationship of object point 1 and object point 2 on the right image at camera position 1 will be different from that at camera position 2. If this issue is not handled with care, the image-disparity may either blur the boundaries of the seamed panoramic images or make the seamed boundaries look discontinuous.

Let PO_i be the position (x_i, y_i, z_i) of object point i in the world coordinate system, CA_r be the position of the camera at orientation r in the world coordinate system, and $PI_{(i,r)}$ be the image coordinates of the object point i when the camera is at orientation r. To perform disparity warping, we need to find $PI_{(i,r)}$ for each i and r, which is a function of PO_i and CA_r, i.e.,

$$PI_{(i,r)} = f(PO_i, CA_r, fl, dx, dy). \tag{13}$$

Because $PI_{(i,r)}$ is a function of PO_i and CA_r, as shown in Eq. (13), it seems that the PSI-II system has to find PO_i and CA_r first. Unfortunately, PO_i is not known unless 3D structure of the environment is known,

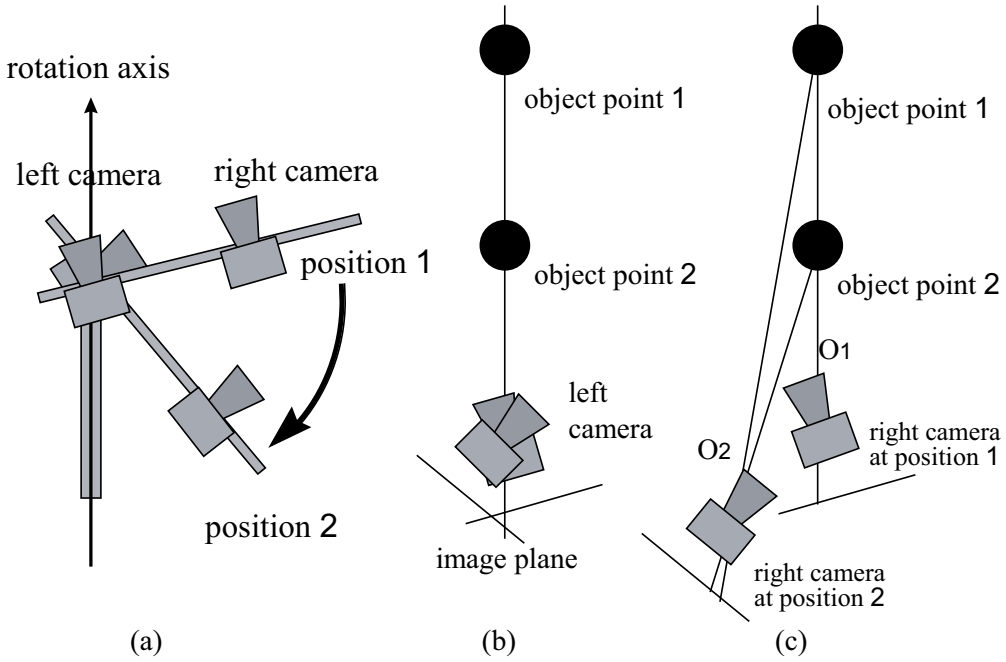


Figure 8: This figure shows the image-disparity caused by dislocating the lens center from the rotation axis. (a) shows the camera configuration used in the PSI-II system. (b) shows that there is no image-disparity for the left images, because the lens center of the left camera is located on the rotation axis. (c) illustrates how the image-disparity occurs for the right images. Because the lens center of the right camera is not on the rotation axis, the objects with different distances will result in different amount of image-disparity.

and CA_r is hard to be estimated accurately without careful calibration. That is, it is difficult to compute $PI_{(i,r)}$ by using Eq. (13) directly. In the PSI-II system, an approximation using image processing techniques is adopted to solve the disparity warping problem, as described below.

Realization of disparity warping

The disparity-warping procedure used in the PSI-II system contains two stages. The first stage is a registration process. Similar to the tilt-correction procedure, the consecutive images are registered by using the hierarchical adaptive early jump-out algorithm described in subsection 3.8. With this registration process, the global seaming offset $(\Delta x_r, \Delta y_r)$ for each pair of consecutive images can be obtained.

The second stage first estimates the individual disparity offset for each object point in the overlapping area of the two consecutive images. As shown in figure 9, a pair of consecutive images is pre-matched with the global seaming offset $(\Delta x_r, \Delta y_r)$ computed for this pair in the first stage. The block-matching technique is then applied to find the disparity offsets $POI_{(i,r,r+\Delta r)}$ for each object point. In order to warp the images smoothly, we then compute the warping offsets POI' by interpolating those disparity offsets based on the distances from the object points to the right and left horizontal boundaries of the overlapping area. The following equation is used to obtain POI':

$$POI'_{(i,r,r+\Delta r)} = POI_{(i,r,r+\Delta r)} \times \frac{OFF1_{(i,r,r+\Delta r)}}{OFF1_{(i,r,r+\Delta r)} + OFF_{(i,r,r+\Delta r)}}, \tag{14}$$

where $OFF_{(i,r,r+\Delta r)}$ is the horizontal distance between the image position of the object point i in image r and the left boundary of the overlapping area in image r, and $OFF1_{(i,r,r+\Delta r)}$ is the horizontal distance between the image position of the corresponding object point i in image $r+\Delta r$ and the right boundary of the overlapping area in image r, as shown in figure 9. Finally, the output image intensity values for the overlapping area of image r are obtained by using the following equation:

$$ID_{(right,r)}((x_{i,r}, y_{i,r}) + POI'_{(i,r,r+\Delta r)}) = IE_{(right,r)}(x_{i,r}, y_{i,r}), \tag{15}$$

where $x_{i,r}$ and $y_{i,r}$ are the horizontal and vertical image coordinates of the object point i in image r, respectively. The result of equation (15) may not cover all the pixels of the overlapping area, and those uncovered pixels can be obtained by interpolation.

Similarly, we can apply the same technique to find and interpolate the image disparities of the left overlapping area on image $r+\Delta r$, which is overlapped by image r, by using the following equation:

$$POI'_{(i,r+\Delta r,r)} = POI_{(i,r+\Delta r,r)} \times \frac{OFF_{(i,r+\Delta r,r)}}{OFF_{(i,r+\Delta r,r)} + OFF1_{(i,r+\Delta r,r)}}, \tag{16}$$

and then obtain the resulted image by using the following equation:

$$ID_{(right,r+\Delta r)}((x_{i,r+\Delta r}, y_{i,r+\Delta r}) + POI'_{(i,r+\Delta r,r)}) = IE_{(right,r+\Delta r)}(x_{i,r+\Delta r}, y_{i,r+\Delta r}). \tag{17}$$

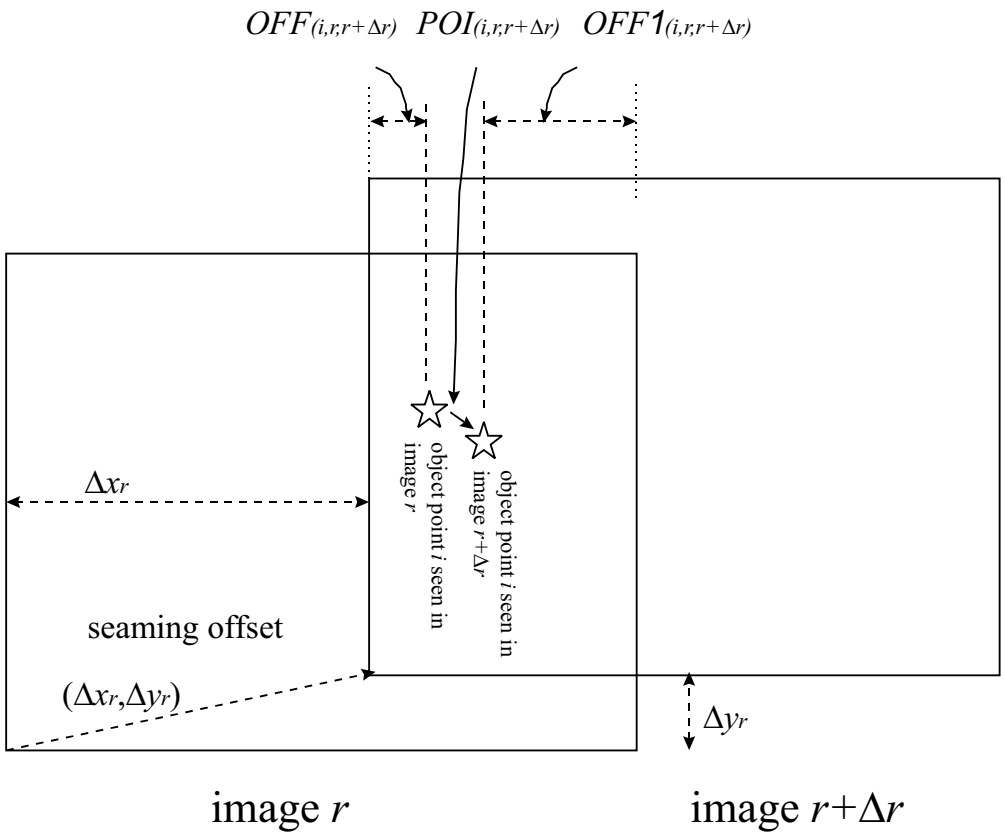


Figure 9: Illustration of the disparity warping.

The disparity warping procedure is the most time consuming procedure among all the procedures involved in generating a pair of stereo panoramas. Its complexity is order of the product of "the number of pixels within the overlapping area" and "the size of the search template" and "the size of the search area" for each pixel. In large areas with little texture, the disparity warping procedure may find incorrect disparity offsets, but those incorrect disparity offset will cause little degradation in visual quality because those areas only contain little texture. In areas with repeating texture, the disparity warping procedure may obtain incorrect disparity offsets only when the repeating period is smaller than half of the search area. In our experiments, the size of the search area is only about 10 pixels, because the distance between the lens centers of image r and image $r+\Delta r$ is quite small. Hence, only the repeating texture with very small period may find incorrect disparity offsets, which is unusual to happen. Even if it happens, it may not be noticeable due to the small repeating period.

3.8 Automatic Image Seaming

Finally, all the warped images should be stitched together to generate a pair of 360-degree panoramic stereo images. Since the camera position and orientation associated with each image are not assumed to be known exactly, the PSI-II system 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 accomplished by using an image blending algorithm.

Brief review of the adaptive early jump-out technique

In [16], we has proposed an adaptive early jump-out technique for fast motion estimation and template matching. Early in 1972, Barnea and Silverman [17] introduced a class of sequential similarity detection algorithms for expediting the similarity detection between two structured data sets. Their contribution was to propose a sequential algorithm where a monotonically-increasing 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 [18].

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

$$MAE(\Delta x, \Delta y) = \frac{1}{n^2} \sum_{i=0}^{n \times n-1} \left| S_c(i) - S_{r(\Delta x, \Delta y)}(i) \right| \tag{18}$$

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,

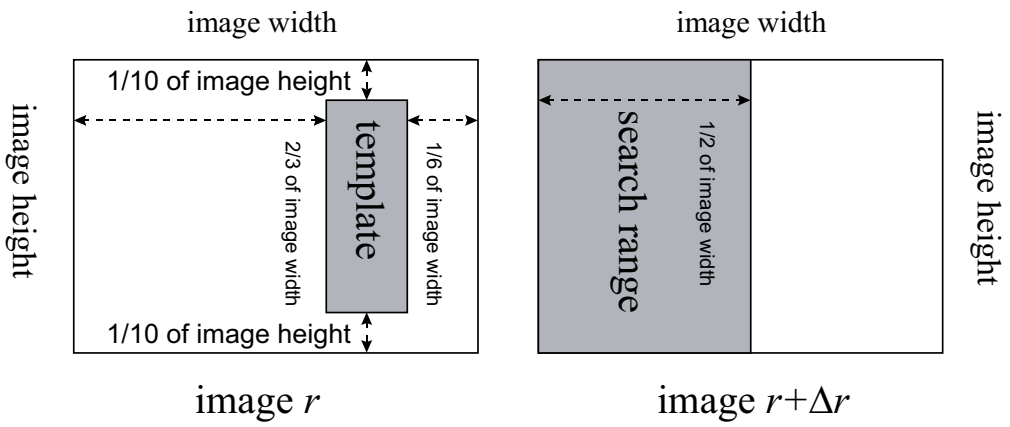


Figure 10: Illustration of the search template and the search area in the image seaming procedure.

$$AE_j = \sum_{i=0}^{j} \left| S_c(i) - S_{r(\Delta x, \Delta y)}(i) \right|. \tag{19}$$

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[17, 18] or by some training methods[19] were unsuitable for motion estimation. An adaptive algorithm for determining the EJO threshold sequence was then proposed in [16]. 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.

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 consecutive images. The adaptive EJO technique[16] is suitable for block matching in the seaming procedure. In our experiments, the template for block matching is chosen from image r (as shown in figure 10), having the size of 1/6 of the original image width by 4/5 of the original image height and 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 $(r + \Delta r)$.

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 speed up the searching process. The search step size for both the horizontal and the 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. Notice that the size of the template for block matching is the same for all levels during the hierarchical

searching procedure. That is, only the search step is changing from coarse to fine. The three-step search commonly used in motion estimation can easily be trapped in local minima. Here, we want the seaming procedure to find the global minimum. Thus, for each search level, the full search area is used. The purpose of using the hierarchical search strategy here is not to reduce the number of matching but to obtain a good (i.e., tighter) early jump-out threshold sequence as early as possible. By using the hierarchical search strategy, the learning speed of the early jump-out threshold sequence is much faster, which can then save much more matching time.

Image Blending

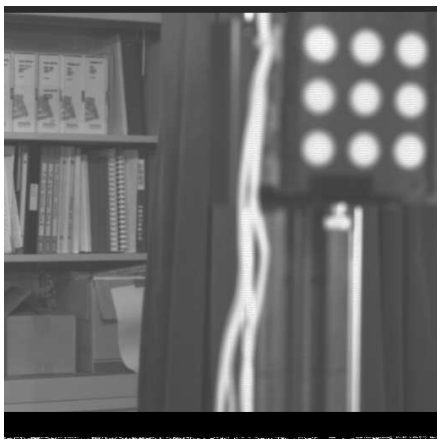
After the seaming offsets are calculated by using the hierarchical adaptive early jump-out block matching, the consecutive images is stitched together with a blending function. If a pixel of the panoramas is covered by only one view, the pixel value is equal to the corresponding pixel value. If a pixel of the panoramas is covered by two consecutive views, the pixel value is the weighted sum of the corresponding pixels of the two consecutive views, where the weight is the offset between the pixel and the right (or left) boundary of the corresponding image.

4 Experimental Results

Several pictures are shown in this section to illustrate the processing and the results of the PSI-II system. Figure 11 shows two of the original images captured by the cameras with different focal distances. Since the 3D distances of the objects in the two images differ a lot, none of the two images contain clear images of all objects. For example, in figure 11(a), the bookshelf is clear but the nine-dot calibration plate is blurred. On the other hand, the bookshelf in figure 11(b) is blurred while the calibration plate is clearly imaged. Figure 11(c) shows a complete-focus image generated by the PSI-II system. All objects in the complete-focus image are clear and sharp.

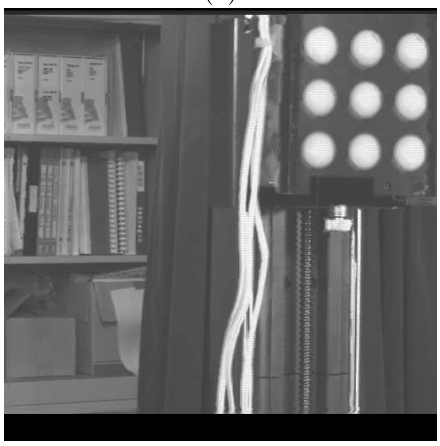
The processes of applying a series of warping procedures are illustrated by figures 12–14. Figure 12 shows the result of tilt-correction and projection-warping. Figures 12(a) and 12(b) are two consecutive images of "lobby 1" acquired by the right camera locating at position 1 (r degrees) and position 2 ($r + \Delta r$ degrees). Notice that the doorknob is not on the same scan line in this image pair. After applying both the tilt-correction and projection-warping procedures, the image locations of the doorknob and chairs match better, as shown in figures 12(c) and 12(d).

Figure 13 illustrates the effect of the epipolar-correction procedure. Figures 13(a) and 13(b) show a pair of stereo images obtained after applying both the tilt-correction and projection-warping procedures. Figures 13(c) and 13(d) show the stereo images obtained by applying the epipolar-correction procedure to the images shown in figures 13(a) and 13(b). After the epipolar-correction procedure, the epipolar-line inconsistency between the left-eye and the right-eye panoramic images is largely eliminated.



(a)

(b)



(c)

Figure 11: (a) and (b) Show two images with different focus settings. (c) Shows the complete-focus image obtained with the PSI-II system.

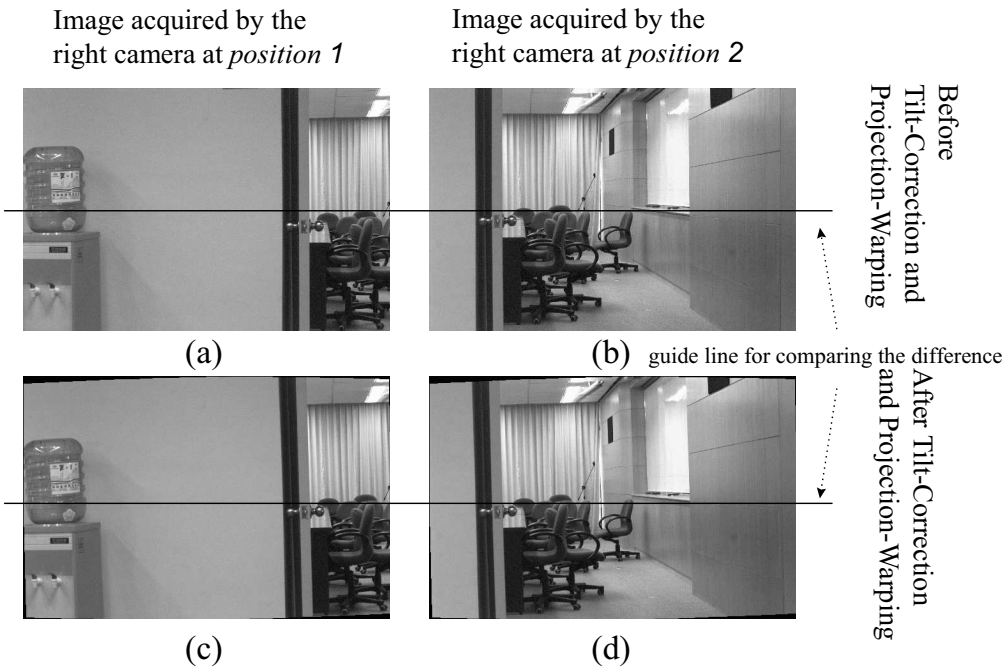


Figure 12: (a) and (b) show the images acquired by the right camera at *position 1* and *position 2*, respectively. (c) and (d) show the result of applying both the tilt-correction and projection-warping procedures to (a) and (b). As one can see by following the guide line in the overlapping area, the images of the doorknob and chairs in (c) and (d) match better than those in (a) and (b).

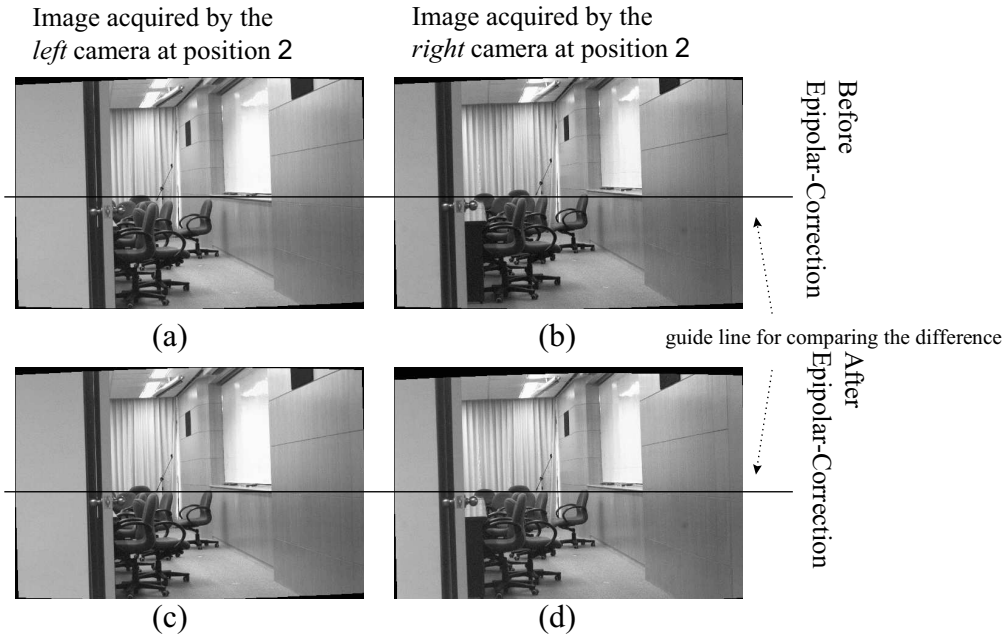


Figure 13: (a) and (b) show the images acquired by the *left* camera and the *right* camera at position 2, respectively. (c) and (d) show the result of applying the epipolar-correction procedure to (a) and (b). As one can see by following the guide line in the overlapping area, the images of the doorknob and chairs in (c) and (d) match better than those in (a) and (b).

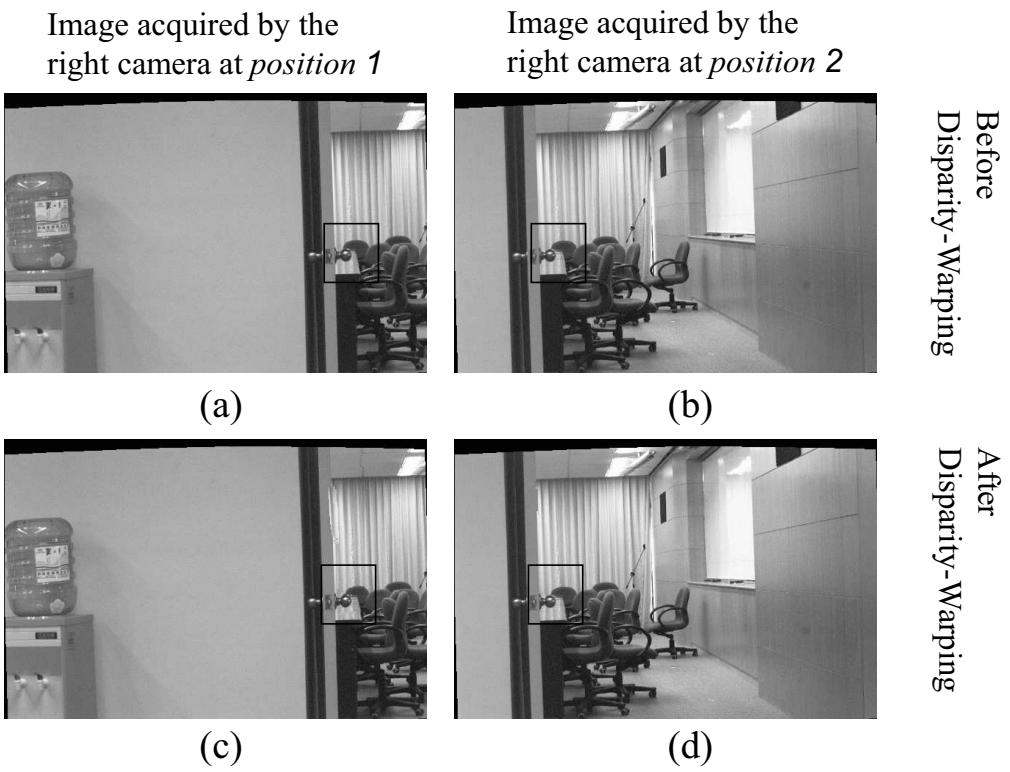


Figure 14: (a) and (b) show the images acquired by the right camera at *position 1* and *position 2*, respectively. Notice that these two images are the results obtained after applying the epipolar-correction procedure to the two consecutive images shown in figures 12(c) and 12(d). (c) and (d) show the result of applying the disparity-warping procedure to (a) and (b). As one can see by examining the area marked by the square box, the images of the doorknob and the desk in (c) and (d) match better than those in (a) and (b).

Figure 14 displays the effect of the disparity-warping procedure. Figures 14(a) and 14(b) show the images obtained after applying the epipolar-correction procedure to the two consecutive images shown in figures 12(c) and 12(d). Here, figure 14(b) is exactly the same as figure 13(d). Notice that the area of the desk behind the door seen in figure 14(a) is different from that seen in figure 14(b). To avoid discontinuity (or double blurring) in the seamed image, we applied the disparity-warping procedure described in the previous section and obtained the two images shown in figures 14(c) and 14(d).

Figure 15 shows three pairs of panoramic stereo images automatically stitched together by the PSI-II system. for visual evaluation, figure 16 shows a pair of stereo images which is cut from a stereo panorama produced by the PSI-II system. More experimental results can be found in `http://smart.iis.sinica.edu.tw/~jet/panoramas/iis.html`.

Figure 15: This figure shows three examples of the panoramic stereo images generated by the PSI-II system: "lobby 1", "lobby 2", and "farm". Each of the left three images shows a snapshot of the original images. The six strips of images on the right are the panoramic stereo images obtained after automatic warping and seaming.

Figure 16: A pair of stereo images which is a part of a stereo panorama.

5 Conclusions

This paper has presented a new panoramic stereo imaging system—the PSI-II system. This system can automatically generate a pair of stereo panoramas from a series of overlapping stereo images and provide the viewer full 360° panoramic stereo views of an environment interactively. In summary, the PSI-II system consists of the following eight different modules described in Section 3, and all the procedures are fully automatic except for the acquisition procedure:

(1) Acquisition of a series of overlapping 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 when required.

(3) 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 registration procedure similar to the one described in section 3.8 is included in this module.

(4) Projection warping — This module is widely-used in the existing panoramic imaging systems (see [3]) and is especially important when using wide-angle lens.

(5) Gain/Offset normalization — This module compensates the effect caused by different characteristics of the two cameras, which can lead to more comfortable stereo viewing and is also a pre-requisite step for the epipolar-correction module that requires block-matching between the stereo image pair.

(6) Epipolar correction — This module compensates the effect caused by the vertical offset between the two cameras, and can also result in more comfortable stereo viewing.

(7) Disparity warping — This module corrects the image-disparity caused by the dislocation of the lens center from the rotation axis. It can result in smoother stitching and more realistic stereo viewing in general. A registration procedure similar to the one described in section 3.8 is also included in this module.

(8) Automatic image seaming — This module first determines the seaming offsets for each pair of consecutive images (an image-registration problem), and then stitches together all the registered images into one right-eye

and one left-eye 360-degree panorama by using an image blending function. To solve the image registration problem efficiently and robustly, we have proposed a hierarchical adaptive EJO block-matching algorithm which is especially useful for the registration of panoramic images.

There are still several unsolved problems in developing an image-based VR 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 problem is the feeling of discontinuity caused by "view hopping" from one panoramic viewing node to another panoramic viewing node in a panoramic imaging system. We are currently extending the disparity warping technique to solve this problem, and have obtained some good results (which will be described in another paper).

Acknowledgments

The authors would like to thank Dr. Tse Cheng at Advanced Technology Center, Industrial Technology Research Institute, for many useful discussions. Also, we very appreciated anonymous reviewers' comments.

References

[1] J. M. Moshell. Three views of virtual reality: Virtual environments in the U. S. military. *IEEE Computer*, 26(2):81–82, February 1993.

[2] G. B. Newby. Virtual reality. *Annual Review of Information Science and Technology*, 28, 1993. Medford, N,J,: Learned Information.

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

[4] 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.

[5] R. Szeliski and H.-Y. Shum. Creating full view panoramic image mosaics and environment maps. *Proceedings SIGGRAPH '97*, pages 251–258, 1997.

[6] C. H. Séquin and R. W. Bukowski. Interactive virtual building environment. In *Proc. of PacificGraphics'95*, pages 159–179, 1995.

[7] Y.-W. Lei. *The SpaceWalker Walkthrough system for unrestricted three-dimensional polygon environments.* PhD thesis, National Taiwan University, Taipei, Taiwan, ROC, 1996.

[8] H.-C. Huang Y.-P. Hung and S.-W. Shih. Panoramic stereo imaging with complete-focus views for virtual reality. In *Proceeding of Second Workshop on Real-Time and Media Systems, RAMS'96*, pages 405–410, Taipei, Taiwan, ROC, July 1996.

[9] S.-W. Shih, Y.-P. Hung, and W.-S. Lin. Calibration of an active binocular head. to appear in *IEEE Trans. on Systems, Man, and Cybernetics,* Vol. 28A, No. 3, May 1998 (ftp://140.109.23.242/pub/TechRpt/TR-IIS-96-007.ps.gz).

[10] L. McMillan and G. Bishop. Plenoptic modeling: An image-based rendering system. *Proceedings SIGGRAPH '95*, pages 39–46, 1995.

[11] S. M. Seitz and C. R. Dyer. View morphing. *Proceedings SIGGRAPH '96*, pages 21–30, 1996.

[12] M. Levoy and P. Hanrahan. Light field rendering. *Proceedings SIGGRAPH '96*, pages 31–42, 1996.

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

[14] Z. Zhang. Determining the epipolar geometry and its uncertainty: A review. INRIA-report, No. 2927, July 1996.

[15] 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.

[16] H.-C. Huang and Y.-P. Hung. Adaptive early jump-out technique for fast motion estimation on video coding. To appear in Graphical Models and Image Processing, November, 1997.

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

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

[19] 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.