Proceedings of International Conference on Image
Processing and Character Recognition

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
Email: jet@iis.sinica.edu.tw, hung@iis.sinica.edu.tw

Abstract

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 hierarchical seaming algorithm. Our PSI-II system can automatically correct the epipolar lines inconsistency of the stereo images pairs and the disparity distortion of the images caused by the dislocation of the camera's lens center in the image acquisition process. With the stereo and automatic correction features, the proposed method can easily provide more realistic 360° *panoramic views than other image-based VR systems can.*

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 is 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, and Omniview PhotoBubbles, 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 reconstructs the VR images by rendering the 3-D models. The model-based approach provides good interactivity, but building 3-D models with model-based approach is a tedious work. On the other hand, the image-based approach builds panoramic VR views by seaming photographs of the real world directly, and enjoys the advantage of easy and fast VR world building and 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. For 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]. The well calibrated IIS head provides accurate camera parameters for image warping and seaming.

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 available for the new PSI-II system anymore, 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 disparity distortion of the images 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 has the following five new features: (1) It can provide 360-degree panoramic stereo views. (2) It can provide complete-focus views. (3) It can correct the disparity distortion and the epipolar-line inconsistency automatically. (4) It can automatically seam a series of overlapped images to generate a pair of panoramic stereo images. (5) It does not require special camera rigs, and can tolerate inexactness of camera configuration.

2 Panoramic Stereo Imaging System

This section describes the detail of the proposed PSI-II system. In order to generate stereo views, two cameras are used for simultaneous image acquisition. Figure 1 shows the block diagram of the PSI-II system. In our PSI-II system the only given camera parameters are the focal length, fl, and the horizontal and vertical pixel spacing, rx and ry.

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 with a natural and user-friendly way.

2.1 Acquisition of a series of Stereo Image Pairs

The first step of image-based VR world building is to capture a series of overlapped images. With our PSI-II sys-

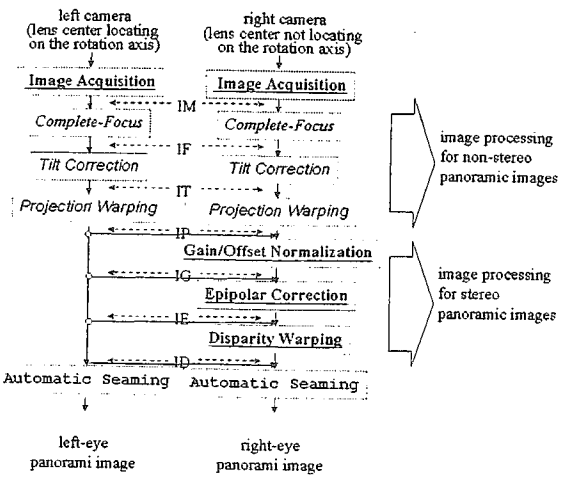


Figure 1: The block diagram of the panoramic stereo imaging system. The underlined blocks are the procedures related to the stereo feature, and the italic blocks are the procedures which are not specifically related to the stereo feature. The automatic seaming procedure uses the adaptive early jump-out technique to speedup the block matching operation in image registration.

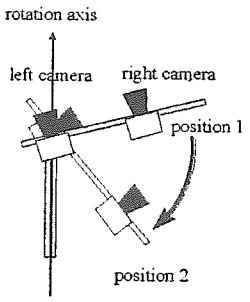


Figure 2: Illustration of the rotation of the stereo camera set.

tem, the stereo images are taken with two digital cameras mounted on a rotational tripod. Figure 2 illustrates the rotation of the stereo camera set.

In our experiments, images are taken when rotating the stereo camera set horizontally, with 15 degrees per step, 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 2.2, 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. After the stereo images acquisition process, a sequence 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 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.

2.2 Generation of Complete-Focus Images

For each camera and each rotation step, seven 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[10]. This process contains two stages, the gradient calculation and the medium filtering. The gradient calculation stage uses equation (1) to calculate the gradients 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 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, $G_{(c,f,r)}(x,y)$, is 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 equation (2) is noisy especially on the boundaries of the objects with different distances. Therefore, a medium filter defined in equation (3) is used to remove the noises near object boundaries.

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

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

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

2.3 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 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 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 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 3. Therefore, we have developed

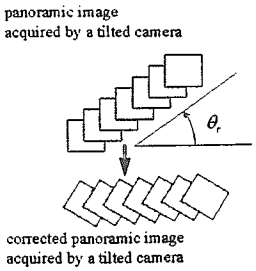


Figure 3: The diagram shows the tilting problem, where the upper part is the panoramic image generated by a tilted camera without correction, and the lower part is the corrected panoramic image.

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.

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

Realization of tilt correction

In equation 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 find θ_r by itself. There are two stages in the tilt correction procedure of the PSI-II system.

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 2.8 to register 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 consecutive 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.

$$\theta_r = \arctan\frac{Y}{X} \quad (6)$$

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

2.4 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 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].

Figure 4: 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.

In order to correct the distortion of the camera projection and eliminate the seaming discontinuity, the following projection warping is performed (refer to figure 4). Here, the projection warping for the left and right images are performed independently.

$$IP_{(c,r)}(x,y) = IT_{(c,r)}(x, \frac{y}{\cos(\arctan\frac{x \times rx}{fl})}) \quad (7)$$

2.5 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. In order to give more comfortable stereo viewing to correct epipolar-line inconsistency, we should perform some equalization between these two panoramic stereo images. In our PSI-II system, 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.

$$IG_{(c,r,rgb)}(x,y) = (IP_{(c,r,rgb)}(x,y) - E_{c,rgb}) \times \frac{V_{left,rgb}}{V_{c,rgb}} + E_{left,rgb}. \quad (8)$$

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.

2.6 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 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 5 illustrates the epipolar line problem. Figure 5(a) shows a tilted tripod, on top of which the two cameras are not on the same horizontal plane. Figure 5(b) and 5(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

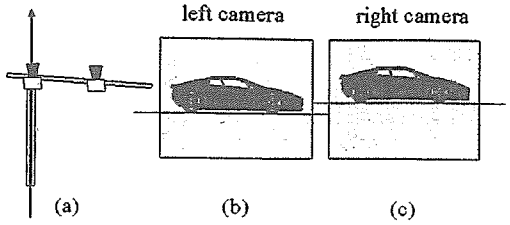


Figure 5: The diagram shows the result of the stereo images without epipolar correction. (a) shows a tilted tripod, and two cameras are not on the same horizontal plane. (b) and (c) show the images captured by left and right cameras, respectively.

inconsistency problem is not fixed, human beings can not comfortably match the stereo images.

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 PSI-II system adjusts the panoramic images of the right camera to match that of the left camera. The function of the epipolar correction is as follows.

$$IE_{(c,r)}(x,y) = IG_{(c,r)}(\sqrt{x^2+y^2}\cos(M_c - M_{left} + \arctan\frac{y}{x}), \sqrt{x^2+y^2}\sin(M_c - M_{left} + \arctan\frac{y}{x}) + O_c - O_{left}), \quad (9)$$

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.

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.

$$IE_{(left,r)}(x,y) = IG_{(left,r)}(x,y) \quad (10)$$

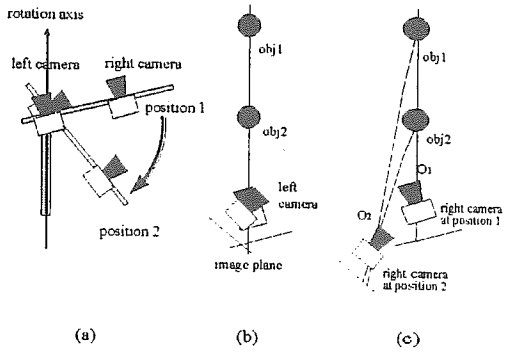


Figure 6: This figure shows the effect of disparity distortion caused by dislocating the lens center from the rotation axis. Figure (a) shows the camera configuration used in the PSI-II system. Figure (b) shows there is no disparity distortion for the left images, because the lens center of the left camera is on the rotation axis. Figure (c) illustrates how the disparity distortion 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 produce different amount of disparity distortion.

$$IE_{(right,r)}(x,y) = IG_{(right,r)}(x, y + EO_m) \quad (11)$$

2.7 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 at 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. This results in the distortion of stereo disparity, and also 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. As illustrated in figure 6, when the right camera is at position 1, object 1 and object 2 are on the same projection line. But when at position 2, object 1 and object 2 will be projected to different positions on the image. That is, if object 1 and object 2 are not of the same distance, the relationship of object 1 and object 2 on the right image at position 1 will be different from that at camera position 2. If this issue is not handled with care, the disparity distortion will occasionally either blur the boundaries of the seamed panoramic images or make the seaming boundaries look discontinuous.

Let PO_i be the position (x_i, y_i, z_i) of object 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 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.

$$PI_{(i,r)} = f(PO_i, CA_r, fl, rx, ry) \quad (12)$$

Realization of disparity warping

Because $PI_{(i,r)}$ is a function of PO_i and CA_r, as shown in equation (12), the PSI-II system should find PO_i and CA_r first. Unfortunately, PO_i is not known unless 3D structure of the environment is known, and CA_r is hard to be estimated accurately without careful calibration. That is, it is difficult to compute $PI_{(i,r)}$ by using equation (12) directly. In the PSI-II system, an approximation approach using image processing techniques is adopted to solve the disparity warping problem. The disparity warping contains two stages.

The first stage is a registration process. Similar to the tilt correction procedure, the panoramic images are registered by using the hierarchical adaptive early jump-out algorithm described in subsection 2.8. With this registration process, the global seaming offset $(\Delta x_r, \Delta y_r)$ for each consecutive image pairs can be obtained.

The second stage first estimates the individual disparity offset for each object point in the overlapped area of a consecutive image pair. As shown in figure 7, 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 pixel by pixel to find the disparity offset $POI_{(i,r)}$ for each object point (seen as a pixel). In order to warp the images smoothly, we then compute the warping offsets POI' for both images of the same object by using the following two equations.

$$POI'_{(i,r)} = POI_{(i,r)} \times \frac{OFF_{(i,r)}}{OFF_{(i,r)}+OFF_{(i,r+\Delta r)}}, \quad (13)$$

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

where Δr is the difference of the rotation angle between two consecutive images (in our experiment, Δr is chosen to be 15 degrees), $OFF_{(i,r)}$ is the offset between the position of the object i in image r and the left boundary of the overlapped area in image r, and $OFF_{(i,r+\Delta r)}$ is the offset between the position of the object i in image $r+\Delta r$ and the right boundary of the overlapped area in image $r+\Delta r$, as shown in figure 7. Finally, the output image values for the overlapping area in the two consecutive images are obtained by using the following two equations:

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

$$ID_{(right,r+\Delta r)}((x_{i,r+\Delta r}, y_{i,r+\Delta r}) + POI'_{(i,r+\Delta r)}) = IE_{(right,r+\Delta r)}(x_{i,r+\Delta r}, y_{i,r+\Delta r}), \quad (16)$$

where $x_{i,r}$ and $y_{i,r}$ is the horizontal and vertical image coordinates of object i in image r, respectively.

The result of the equations (15) and (16) may not cover all the pixels of the overlapped area, and those uncovered pixels can be obtained by interpolation.

2.8 Automatic Image Seaming

Finally, all the warped images should be stitched together to generate the panoramic stereo images. Since the camera position and orientation associated with each image is 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 completed by using an image blending algorithm.

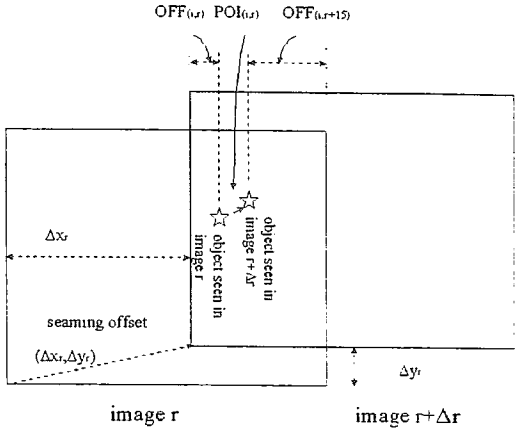


Figure 7: The figure to illustrate the disparity warping of the PSI-II system.

Brief review of the adaptive early jump-out technique

In [11], we has proposed an adaptive early jump-out technique for fast motion estimation and template matching. Early in 1972, Barnea and Silverman [12] 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 [13].

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

$$MAE(\Delta x, \Delta y) = \frac{1}{n^2} \sum_{i=0}^{n\times n-1} \left| S_c(i) - S_{r(\Delta x,\Delta y)}(i) \right| \quad (17)$$

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,

$$AE_j = \sum_{i=0}^{j} \left| S_c(i) - S_{r(\Delta x,\Delta y)}(i) \right| \quad (18)$$

Let us define the EJO threshold sequence EJS_j, such that if the accumulated error AE_j is greater than EJS_j,

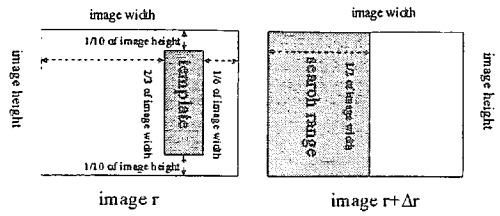


Figure 8: The figure to illustrate the search template and the search area of the images seaming procedure.

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 either the mathematical early-jump-high models[12, 13] or by some training methods[14] were unsuitable for motion estimation. An adaptive algorithm for determining EJO threshold sequence was then proposed in [11]. 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.

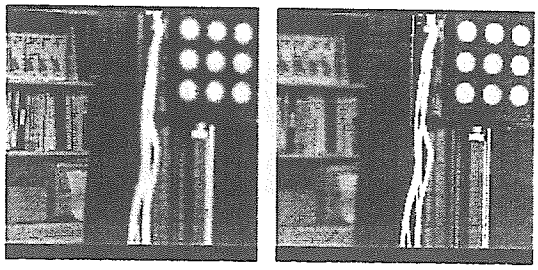
Hierarchical adaptive EJO block matching for panoramic images registration

In this paper, we use block matching to find the seaming offsets for each pair of consecutive images. The adaptive EJO technique[11] is suitable for block matching in the seaming procedure. As shown in Figure 8, the template for block matching is in image r, and 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 $r + \Delta r$ 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 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 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 consecutive 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 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 left or right boundary of the corresponding image.



(a) (b)

Figure 9: Two images with different focus settings.

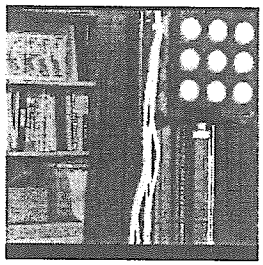


Figure 10: A complete-focus image.

3 Experimental Results

Several pictures are shown in this section to illustrate the processing and the results of the PSI-II system. Figure 9 shows two of the original images captured by the cameras with different focal distances. Since the 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 9(a), the bookshelf is clear but the nine-dot calibration plate is blurred. On the other hand, the bookshelf in Figure 9(b) is blurred while the calibration plate is clearly imaged. Figure 10 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 11–13. Figure 11 shows the result of tilt correction and projection warping. Figures 11(a) and 11(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 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,

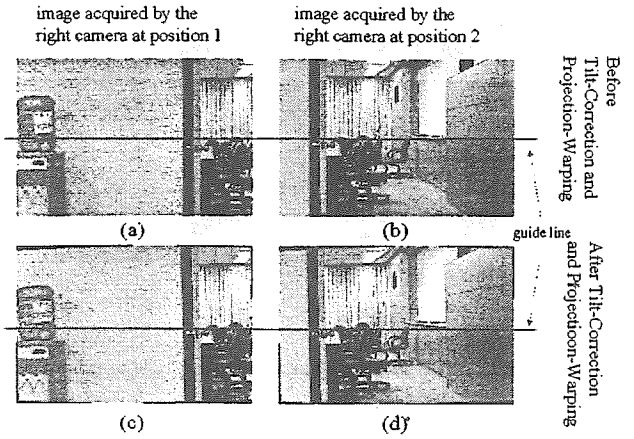


Figure 11: (a) and (b) show the images acquired by the right camera at position 1 and position 2, respectively. (c) and (d) show the result of applying both the tilt-correction and projection-warping procedures to (a) and (b). As one can see by following the guide line in the overlapping area, the images of the doorknob and chairs in (c) and (d) match more accurately than those in (a) and (b).

as shown in figures 11(c) and 11(d).

Figure 12 illustrates the effect of the epipolar-correction procedure. Figures 12(a) and 12(b) show a pair of stereo images obtained after tilt-correction and projection-warping procedures. Figures 12(c) and 12(d) show the stereo images obtained by applying the epipolar-correction procedure to the images shown in Figures 12(a) and 12(b). After the epipolar-correction procedure, the epipolar-line inconsistency between the left-eye and the right-eye panoramic images is largely eliminated.

Figure 13 displays the effect of the disparity-warping procedure. Figures 13(a) and 13(b) show the images obtained after applying the epipolar-correction procedure to the two consecutive images shown in Figures 11(c) and 11(d). Here, Figure 13(b) is exactly the same as Figure 12(d). Notice that the area of the desk behind the door seen in Figure 13(a) is different from that seen in Figure 13(b). To avoid discontinuity in the seamed image, we applied the disparity-warping procedure described in the previous subsection and obtained the two images shown in Figure 13(c) and 13(d).

Figure 14 shows three pairs of panoramic stereo images automatically stitched together by the PSI-II system, and Figure 15 shows a picture where a user is interactively viewing the panoramic stereo images generated by the PSI-II system.

4 Conclusions

A new panoramic stereo imaging system has been presented in this paper. This PSI-II system can automatically generate a pair of panoramic stereo images from a series of images and provide the viewer 360° panoramic stereo view of an environment interactively. In summary, the PSI-II system consists of the following eight different modules

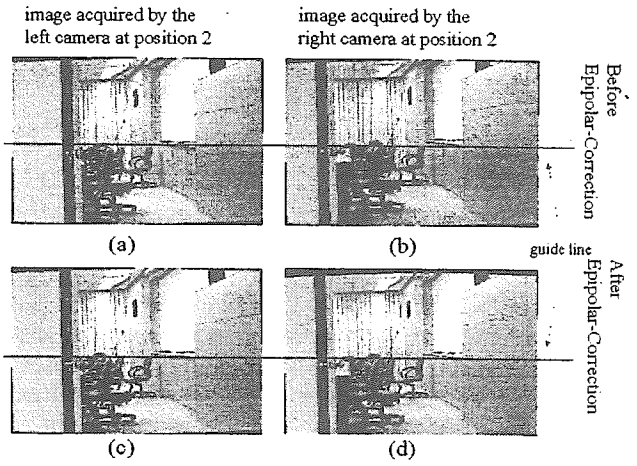


Figure 12: (a) and (b) show the images acquired by the left camera and right camera at position 2, respectively. (c) and (d) show the result of applying the epipolar-correction procedure to (a) and (b). As one can see by following the guide line in the overlapping area, the images of the doorknob and chairs in (c) and (d) match more accurately than those in (a) and (b).

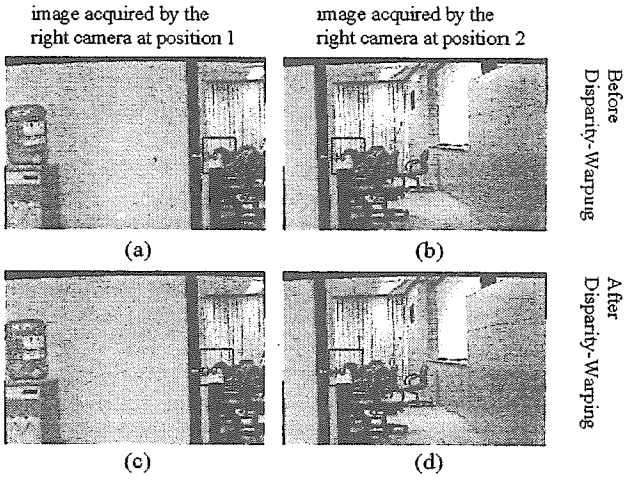


Figure 13: (a) and (b) show the images acquired by the right camera at position 1 and position 2, respectively. Notice that these two images are the results obtained after applying the epipolar-correction procedure to the two consecutive images shown in Figures 11(c) and 11(d). (c) and (d) show the result of applying the disparity warping procedure to (a) and (b). As one can see by examining the area marked by the box, the images of the doorknob and the desk in (c) and (d) match better than those in (a) and (b).

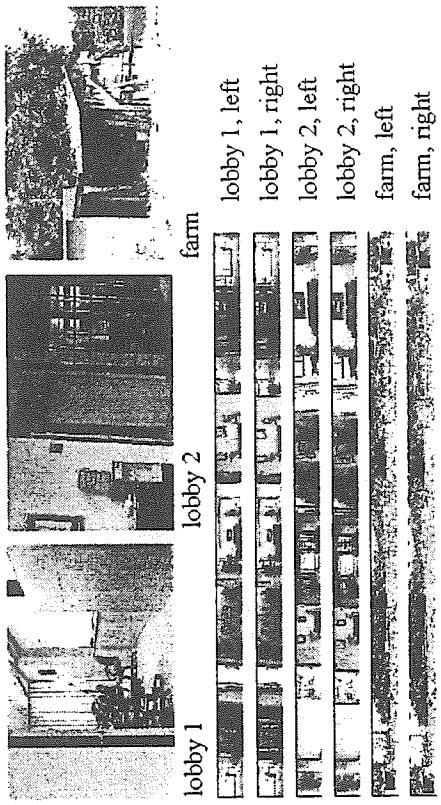


Figure 14: This figure shows three examples of the panoramic stereo images generated by the PSI-II system: "lobby 1", "lobby 2", and "farm". Each of the left three images shows a snapshot of the original image. The six strips of images on the right are the panoramic stereo images obtained after automatic warping and seaming.

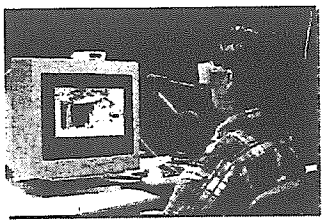


Figure 15: The photograph of a user interactively viewing the panoramic stereo images generated by the PSI-II system.

described in subsections 2.1–2.8: Acquisition of a series of stereo image pairs, generation of complete-focus images tilt correction, projection warping, gain/offset normalization, epipolar correction, disparity warping, and automatic image seaming.

There are still several unsolved problems in developing an image0based VR system. For example, moving objects can cause problems for most panoramic imaging systems. However, this problem can be partially solved by taking multiple images at same viewpoint, and then integrating them together appropriately. The feeling of discontinuity caused by "view hopping" is another important but difficult problem for the panoramic imaging systems. We are currently working on this problem.

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

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

[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. In *Proceeding of CGVIP'96*. pages 219–226, Taichung, Taiwan, ROC, 1996. TR-IIS-96-007 (ftp://140.109.23.242/pub/TechRpt/TR-IIS-96-007.ps.gz).

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

[11] 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.

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

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

[14] 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.