Authoring and Visualizing Stereo Panoramic Images with Independent Objects

Johnson Chen

Computer Science Department and CITR,
The University of Auckland, Auckland, New Zealand
yche158@ec.auckland.ac.nz

Shou-Kang Wei

Department of Information and Computer Science,
Keio University, Yokohama, Japan
shoukang@ozawa.ics.keio.ac.jp

Reinhard Klette

Computer Science Department and CITR,
The University of Auckland, Auckland, New Zealand
r.klette@auckland.ac.nz

Abstract

Object embedded stereo panoramic images have received increasing attention recently due to their numerous applications in computer vision. This article explores an approach to embed object images, both dynamic and static, into background panoramas. The approach involves three steps; namely, image acquisition, post-acquisition, and image visualization. A GUI tool is developed for image registering, merging, and pose recovering. A panoramic player is also developed for visualizing the resultant anaglyphic panorama.

Keywords: panoramic image, dynamic object, static object, camera pose, anaglyphic image

1   Introduction

This paper discusses embedding of objects into panoramic images so that a viewer can visualize and manipulate the objects directly in the panoramas without noticing any artificial process applied on them. This is of potential use for interactive 3D games, website interaction, virtual reality, or multimedia.

There are two common approaches to create object embedded panoramic images; the image-based approach and the model-based approach. The model-based approach allows a user to look at the virtual world from an arbitrary viewing direction but its rendering speed depends on the complexity of the objects. So it is not realistic to visualize a sophisticated panorama without a special graphics accelerator. The image-based approach can generate photo-realistic scenes; however, it is difficult to allow the user to view objects from different viewpoints.

C.S. Chen et al. [1] introduce a method to composite 3D graphic objects into a panorama. The 3D objects are generated with a model-based approach, while the panorama is created by an image-based approach. It gives great flexibility to manipulate objects; however, complexity constraints of modeling the objects limit their uses.

Y.P. Hung et al. [4] propose a purely image-based approach to insert object movies into panoramas so that objects and the panorama maintain geometric consistency without reconstructing the 3D geometric models of the objects. This approach is complexity independent.

S.K. Wei et al. [7] introduce an image-based method to embed objects into stereo panoramas which keeps both geometric and photometric consistency between the panorama and the inserted objects. The objects (e.g. an auto-spin globe) have independent behaviours, playing simultaneously and consistently with background panoramas.

All the above approaches deal with static objects, which do not change their physical position in the panorama. We are also interested in showing dynamic objects, (e.g. a moving car), animated static objects, (e.g. a tree swinging in the wind), such animations can significantly vivify the virtual world and have many potential uses. This article explores an approach to embed the dynamic and static objects into background panoramas seamlessly, such that their animations can be visualized realistically.

The rest of the article is organized as follows: Section 2 describes the geometric transformation. Section 3 discusses the fundamentals of stereo panoramic images. Experiments are shown in Section 4. In Section 5, we summarize the project and address future works.

2   Geometric Transform

There are three main types of panoramas: cylindrical, spherical, and cubical. Cylindrical panoramas (Fig. 1) are known to be easy to construct and are considered "sufficient"[1] for many applications [3].
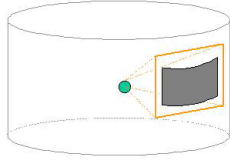

Figure 1: Cylindrical image carrier.

Since the object images which we take are planar, they need to be warped into cylindrical images when embedded into a panorama. When visualizing the panorama, the visible part of the panorama is reprojected onto a planar surface. Therefore, two kinds of 2D transformations are involved, planar to cylindrical and cylindrical to planar. Figure 2 shows the mapping between a cylindrical image and a planar image. The plane image has an xy coordinate system and the cylindrical image has an $x'y'$ coordinate system. We assume that the planar image and cylindrical image have the same width and height and both coordinate systems are measured in pixel units.

2.1   Planar to Cylindrical Image

Any point $p = (x, y)$ in a planar image can be mapped onto a point $p' = (x', y')$ in a cylindrical image by applying Eqs. (1) and (2),

$$x' = \frac{w}{2} - f\,\frac{\frac{w}{2} - x}{\sqrt{f^2 + \left(\frac{w}{2} - x\right)^2}} \tag{1}$$

$$y' = \frac{h}{2} - f\,\frac{\frac{h}{2} - y}{\sqrt{f^2 + \left(\frac{w}{2} - x\right)^2}} \tag{2}$$

where w is the width, h is the height of an image, and f is the focal length.

This transformation is used to warp object images into cylindrical images.

<hr>

[1] Here "sufficient" means that we do not have any obvious degradation of cylindrical panoramas compared with spherical or cubical panoramas.
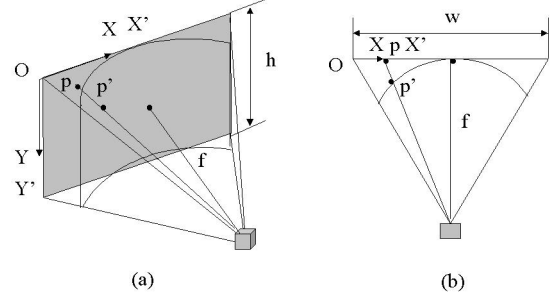

Figure 2: Mapping between cylindrical and planar image (a) side view, (b) top view.

2.2   Cylindrical to Planar Image

Any point $p' = (x', y')$ in a cylindrical image can be mapped onto a point $p = (x, y)$ in a planar image following Eqs. (3) and (4) (see Fig. 2),

$$x = \frac{w}{2} - f\,\frac{\frac{w}{2} - x'}{\sqrt{f^2 - \left(\frac{w}{2} - x'\right)^2}} \tag{3}$$

$$y = \frac{h}{2} - \frac{\frac{h}{2} - y'}{f}\sqrt{f^2 + \left(\frac{w}{2} - x'\right)^2} \tag{4}$$

where w is the width, h is the height of an image, and f is the focal length.

This transformation is mainly used for visualizing a stereo panorama.

2.3   Object Image to Panorama

Any point $p' = (x', y')$ in an object image with an $x'y'$ coordinate system can be mapped onto a point $p = (i, j)$ in a panorama with an ij coordinate system (see Fig. 3), using Eqs. (5) and (6)

$$i = x' + i' \tag{5}$$

$$j = y' + j' \tag{6}$$

where i', j' are horizontal and vertical distances between the original points (in xy coordinates) and the transformed points in ij coordinates.

3   Stereo Panoramic Imaging

3.1   Anaglyphic Image and Panorama

Color anaglyphic images are an inexpensive and efficient way to visualize stereo images in comparison to other approaches, such as shutter stereoscopes or mirror stereoscopes. To view an anaglyphic image, a viewer requires 3D eyeglasses, which have two colour
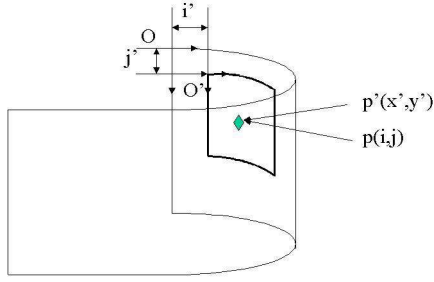
Figure 3: Mapping from object image to background panorama.

filters, for example, 3D eyeglasses may have a red filter on the left eye and a green or blue filter on the right eye allowing only the specified colour(s) to pass into the respective eyes. The depth information can then be recognized so that a 3D image is perceived. This technique is chosen because it is easy to implement, only requiring a pair of 3D eyeglasses.

To make anaglyphic images, the following steps are necessary [7]:

- Create a new image A, which has the same size with the left and right image L and R.

- Copy the green and blue channel from R into A.

- Copy the red channel from L into A and displace it horizontally for perceiving it at a proper distance.

- Cut off the non-overlapping areas in A.

Conventionally, panoramas are made by taking many photographic or video images to cover the whole viewing space we are interested in, then align and composite them using image mosaicing or stitching techniques [6]. Here, we take the image sequence for a small increment of an angular interval, then fetch the central columns of each image and place them side-by-side to form a panorama. This method has an outstanding advantage of bypassing the stitching process and reducing photometric inconsistency errors.

3.2 Image Acquisition and Camera Model

Since anaglyphic images are composed by image pairs, they could be acquired by a binocular camera model (see Fig. 4). Basic rules of setting up the camera positions are as follows:

- The optical axes are parallel and oriented towards the scene object.

- Image pairs are coplanar.

- Rows in the left and right image are collinear.

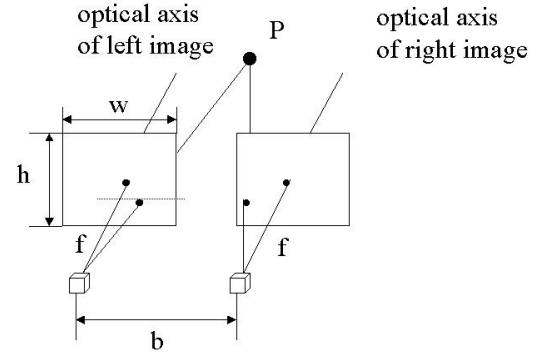- The spacing between the left and right camera approximates the human eyes distance.

Figure 4: Binocular camera model.

The distance between the focal points of the left image L and the right image R, which is denoted by b in Fig. 4, is the base distance. Both the left and right images are located in the same image plane, the distance between the image plane and the focal point is the focal length f.

After setting up the cameras, we can capture stereo image pairs for a background panorama and object images as follows:

- For a background panorama, we obtain the images while the cameras are rotating with a small angular increment.

- For dynamic objects, we fix all camera settings for obtaining the background panorama and capture a video clip while rotating cameras to trace a moving object, one at a time.

- For static objects, we fix all camera settings for obtaining the background panorama and direct the cameras to the static object of interest and take a video clip, one at a time.

If columns near the image centre line are taken to form panoramas, errors occur in both x and y directions. This is because the images are planar, while the panorama is cylindrical (Fig.5). Let w and w' be the number of columns in a planar image and a cylindrical image, respectively; let h and h' be the number of rows in a planar image and a cylindrical image, respectively, then $w' < w$ and $h' < h$. However, if horizontal error $e_h = (w - w')$ and vertical error $e_v = (h - h')$ are both small, e.g. less than 1 pixel, the resulting panorama will not be affected. Therefore, the angular interval increment, α, needs to be obtained in such a way that errors can be controlled within an acceptable range.

To calculate the angular interval increment α, the first step is to determine the errors in the x direction. Let
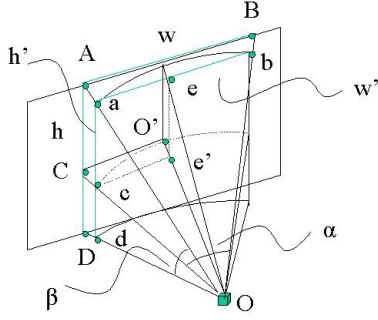
Figure 5: Errors in x, y direction from a planar to a cylindrical image.

$w = AB = 2f \tan\left(\frac{\alpha}{2}\right), w' = ab = 2f \sin\left(\frac{\alpha}{2}\right)$, the horizontal error between the planar and cylindrical image is

$$e_h = AB - ab = 2f(\tan\frac{\alpha}{2} - \sin\frac{\alpha}{2}) \tag{7}$$

Figure 6 shows the horizontal error e_h under different α values and focal lengths. The area under line 1 in Fig. 6 indicates that horizontal errors are less than 1 pixel. According to the figure, the horizontal error e_h will increase when α or f increase. However, if $\alpha < 9°$, the horizontal error will be less than 1 pixel for a range of focal lengths.
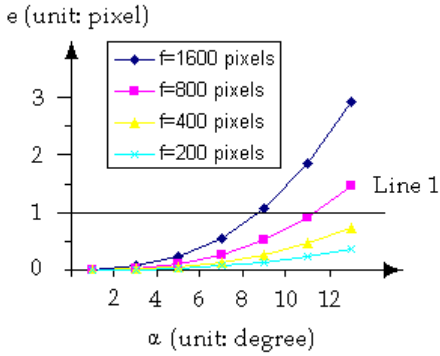

Figure 6: Horizontal errors with different focal length and α.

Next, the errors in the y direction need to be considered. Let

$$
\begin{aligned}
h &= AD = 2f \tan\frac{\beta}{2} \\
h' &= ad = 2ac = \frac{cO}{CO}h = \frac{f}{\sqrt{CO'^2 + OO'^2}}h \\
&= \frac{f}{\sqrt{f^2 \tan^2\frac{\alpha}{2} + f^2}} = \frac{1}{\sqrt{\tan^2\frac{\alpha}{2} + 1}}h \\
&= \cos\frac{\alpha}{2}h
\end{aligned}
$$

β is the vertical field of view, and e_v is the vertical error between the planar image and the cylindrical image

$$e_v = 2f(1 - \cos\frac{\alpha}{2})\tan\frac{\beta}{2} \tag{8}$$

Figure 7 shows the relationship between the vertical error e_v and α with $\beta = 50°$. The area under line 1 indicates that vertical errors are less than 1 pixel. If $\alpha < 3°$, the vertical error will be under 1 pixel for a range of focal lengths.
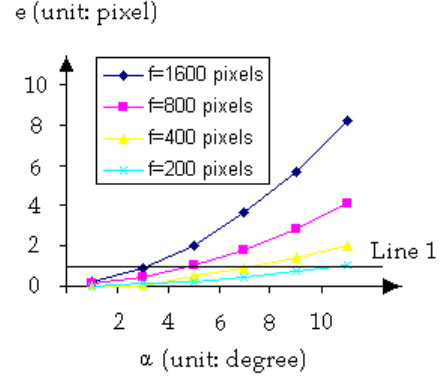

Figure 7: Vertical errors with different focal length and α.

To summarize, when $\alpha < 3°$, both vertical and horizontal errors are believed to be acceptable for most applications.

Since the binocular camera model needs two cameras sitting side by side, it is impossible to have both nodal points of the cameras passing through a single rotation centre. This yields errors because a 3D point of successive images may no longer have the same y value. However, as discussed in [7], this error has less effect on the resulting panorama as long as the closest object is far enough from the cameras. Two camera models (Fig. 8) can be used for image acquisition. We adopt the model A as it reduces errors (see experiments results in [7]). However, a theoretical proof is left for further studies.
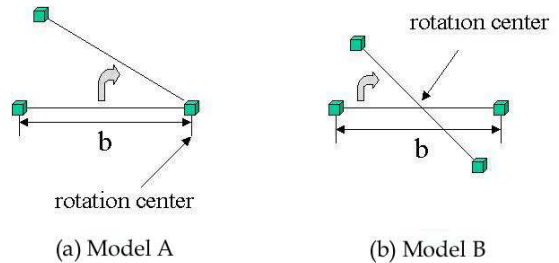

(a) Model A   (b) Model B

Figure 8: Top view of a binocular camera model: (a) rotation centre at right camera (b) rotation centre between both cameras.

3.3 Camera Focal Length

The only unknown in Eqs. (1-4) is the focal length f. For a digital camera or a video camera, the focal length as well as the field of view is provided by the camera specification. It is trivial to obtain the focal length in pixels or millimetres. However, the effective focal length in pixels is not provided for normal cameras. In this case, camera calibration is required. A simple approach [5] is proposed to obtain an accurate focal length for a panoramic image. This is a method different from conventional means of camera calibration. S.B. Kang et al. state that "the relative compositing length[2] error due to camera focal length error is disproportionately much less (i.e., in terms of percentages) than the relative focal length error." The algorithm is as follows: Let the initial estimation of focal length be f_0, use f_0 to generate a panorama and get the compositing length L_0, which leads to a new focal length $f_1 = L_0/2\pi$. Use f_1 to determine the new compositing length L_1, then obtain a new focal length $f_2 = L_1/2\pi$. Repeat the process until $|f_n - f_{n-1}| \leq \varepsilon$, where ε is any small number. f_n is regarded as the final estimated focal length. It is proved that the convergence of f_n toward f_{true} is exponential with a sub-pixel accuracy (when $\varepsilon < 1$ pixel). Hence, performance is guaranteed.

3.4 Camera Pose Recovery

Camera pose recovery plays a very important role in determining where to place object images in a panorama. The pose estimation is dependent on the non-moving object both in the panorama and the object images. So automatic pose estimation is not trivial, since we cannot guarantee the non-moving object is always in front of the moving objects. Therefore, we expect to find a mechanism to perform pose recovery manually without being tedious. We define a search region both in the object images and the panorama; a match between object images and panorama is limited to these search regions. This mechanism is efficient for a static object, since we only define search regions in the background panorama and the first frame of the static object. We can then obtain the camera's pose parameters. However, camera poses for a dynamic object are not trivial; it will be tedious to specify the search regions for each frame of a dynamic object video clip, especially if it is long. Our algorithm for dynamic object pose recovery is as follows:

1. Divide the dynamic object video clip into several segments. Each segment has at least one identical non-moving object that can be visible all the time both in the panorama and object images. Moreover, the direction of object movement should remain the same within the segment.

2. For each segment, specify search regions in the panorama, and the first and last object image frame.

3. Determine location and size of the search regions of the rest of the frames using linear interpolation, and adjust the search regions manually if necessary.

4. Obtain pose parameters by searching pre-defined search regions.

5. Repeat steps 1 to 4 for all images in the object video clip.

3.5 Panoramic Player

An interactive stereo panorama player is developed to navigate and visualize objects embedded into a panorama with a pair of 3D eyeglasses. The basic functions for the player include 360 degree horizontal and less than 180 degree vertical panning, and zooming (in/out). The player loads both the background panorama and the visible object images, composites the background panorama and the object images on-the-fly according to user commands, and the recovered pose parameters, then reprojects it onto a planar viewport.

4 Experiments

We used two EVI-D30 digital cameras to capture images. The focal length of the cameras is 64.8 mm, the horizontal and vertical field of view are 36.6° and 48.8°, respectively, the image size 288×384 pixel. Therefore the focal length is $f = 435$ in pixels. As indicated by Figs. 6 and 7, we choose to capture images at 2.5 degree intervals which is adequate for generating a panorama with both the vertical and horizontal errors less than 1 pixel. The resulting background stereo panorama is shown in Fig. 9.

When we estimated the camera pose, the background panorama[3] and the inserted object image are converted to binary edge images. We then specified two search regions both in the background panorama and the object images. The camera pose parameters, which are used to embed the object into the panorama later, can be obtained by searching the pre-defined search regions. Figure 10 shows the process of a static object pose recovery. Figure 11 illustrates the example of an object-embedded panorama (composed by the right image sequences and an object image).

---

[2]Here the compositing length is the width of the resultant panorama.

[3]This panorama is not the stereo panorama, it is the panorama which is superimposed by right image sequence

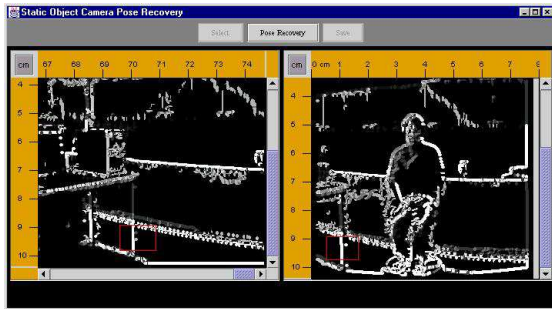Figure 9: Panorama of computer vision lab at CITR.
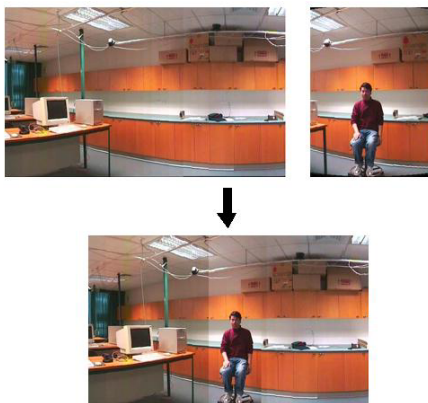

Figure 10: Camera pose recovery.

Figure 11: Part of an oject-embedded panorama.

5 Conclusions and Future Works

We have developed authoring and rendering tools for object-embedded panoramic image visualization. They can be categorized as follows: the panoramic image generator which composites the acquired images to form a stereo background panorama, the stereo image maker which makes stereo object images, the pose recovery which finds the geometric transformation parameters, and the panoramic player which allows a user to visualize the object-embedded panorama. Each part of the software can be extended and the panoramic player plays in real-time.

Two major problems need to be addressed in future. One is the photometric consistency; the other is object manipulation in the interactive panoramic player. As we take the same scene for the panorama and the object images, all photographic parameters, such as the light, color and texture, are kept consistent in both the panorama and the object images. This significantly reduces the difficulty to keep photometric consistency. However, the photometric problem still causes visible edges between the embedded objects and the panorama. The edges can be removed or reduced using "linear distribution of intensity difference" [2] or similar techniques. Manipulable objects are always attractive, such as a door that can be opened or closed; or moving cars that can be stopped according to user commands. Both topics are subject to further research.

References

[1] Chen, C.S., Hsieh, W.T, "Composition of 3D graphic objects and panorama", *Proc. ICAT 2000,* pp 207-214 (2000).

[2] Chen, C.Y., "Image stitching - comparisions and new techniques", CITR-TR-30, The University of Auckland, New Zealand (1998).

[3] Chen, S.E., "QuickTime VR - an image-based approach to virtual environment navigation", *Proc. SIGGRAPH'95,* pp 29-38 (1995).

[4] Huang, Y.P., Chen, C.S., Tang, G.Y., Tsai, Y.P., Huang, C.F., Lin, S.W., Yu, C.H., "A purely image-based approach to augmenting panoramas with object movies", *Proc. Inte Workshop on Computer Graphics and Virtual Reality,* pp 80-88 (2001).

[5] Kang, S.B., Weiss. R, "Characterization of errors in compositing cylindrical panoramic images", *Panoramic Vision: Sensors, Theory, and Applications (R. Benosman and S.B. Kang, editors),* pp 205-226, Springer, New York (2000).

[6] Szeliski, R., Shum, H.Y., "Creating full view panoramic image mosaics and environment maps", *Proc. SIGGRAPH'97,* pp 251-258 (1997).

[7] Wei, S.K., Huang, Y.F., Klette, R., "Color anaglyphs for panorama visualizations", CITR-TR-19, The University of Auckland, New Zealand (1998).