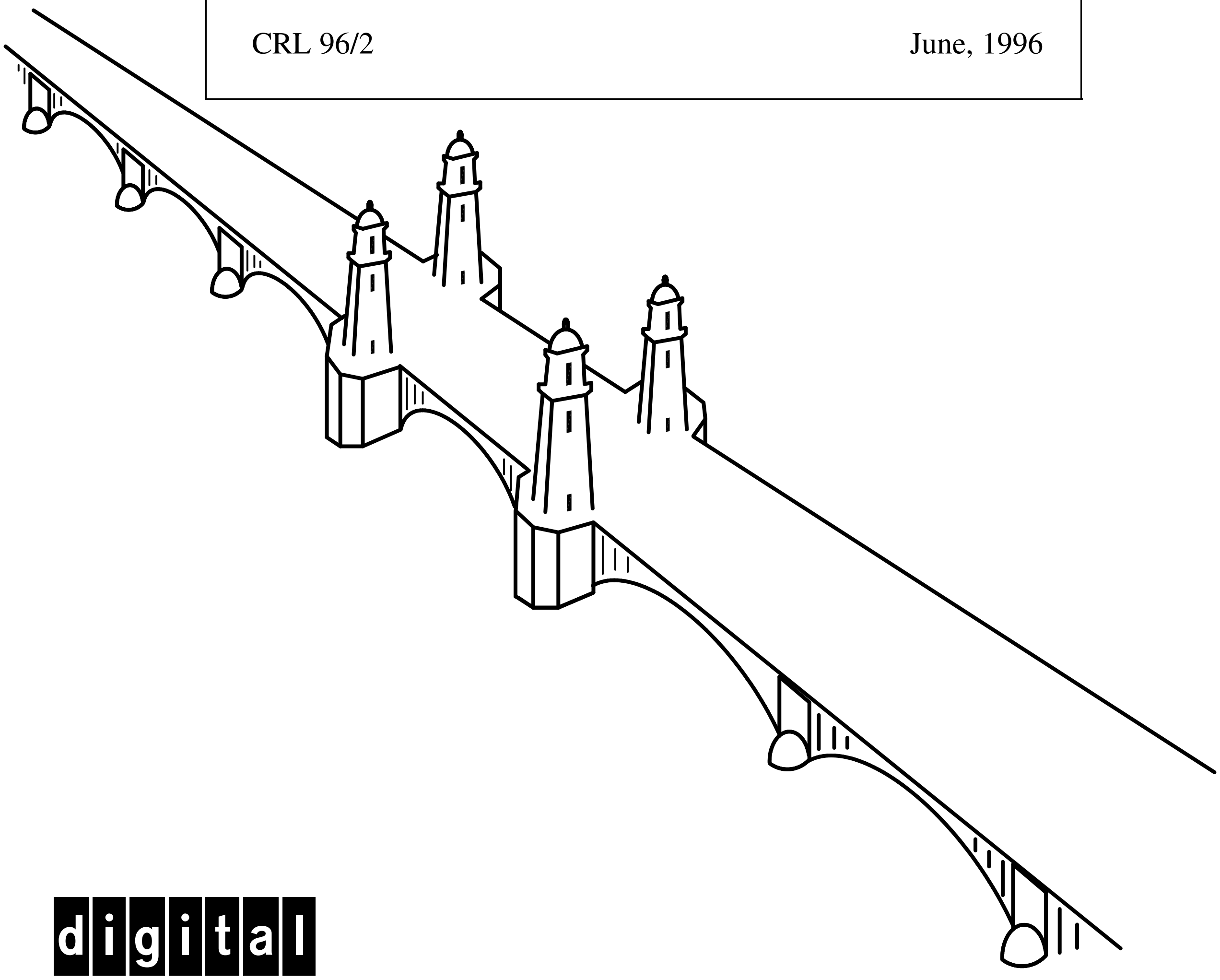
Characterization of Errors in Compositing Panoramic Images

Sing Bing Kang and Richard Weiss

Digital Equipment Corporation
Cambridge Research Lab

CRL 96/2 June, 1996



CAMBRIDGE RESEARCH LABORATORY
Technical Report Series

Digital Equipment Corporation has four research facilities: the Network Systems Laboratory, the Systems Research Center, and the Western Research Laboratory, all in Palo Alto, California; and the Cambridge Research Laboratory, in Cambridge, Massachusetts.

The Cambridge laboratory became operational in 1988 and is located at One Kendall Square, near MIT. CRL engages in computing research to extend the state of the computing art in areas likely to be important to Digital and its customers in future years. CRL's main focus is applications technology; that is, the creation of knowledge and tools useful for the preparation of important classes of applications.

CRL Technical Reports can be ordered by electronic mail. To receive instructions, send a message to one of the following addresses, with the word **help** in the Subject line:

On Digital's EASYnet:	CRL::TECHREPORTS
On the Internet:	techreports@crl.dec.com

This work may not be copied or reproduced for any commercial purpose. Permission to copy without payment is granted for non-profit educational and research purposes provided all such copies include a notice that such copying is by permission of the Cambridge Research Lab of Digital Equipment Corporation, an acknowledgment of the authors to the work, and all applicable portions of the copyright notice.

The Digital logo is a trademark of Digital Equipment Corporation.

digital™

Cambridge Research Laboratory
One Kendall Square
Cambridge, Massachusetts 02139

Characterization of Errors in Compositing Panoramic Images

Sing Bing Kang and Richard Weiss[1]

Digital Equipment Corporation

Cambridge Research Lab

CRL 96/2 June, 1996

Abstract

A panoramic image has 360° horizontal field of view, and it can provide the viewer the impression of being immersed in the scene to a certain extent. A panorama is created by first taking a sequence of images while rotating the camera about a vertical axis. These images are then projected onto a cylindrical surface before being seamlessly composited. The cross-sectional circumference of the cylindrical panorama is called the *compositing length*.

This work characterizes the error in compositing panoramic images due to errors in intrinsic camera parameters. The intrinsic camera parameters that are considered are the camera focal length and the radial distortion coefficient. We show that the error in the compositing length is more sensitive to the error in the camera focal length. Especially important is the discovery that the relative error in compositing length is always smaller than the relative error in the focal length. This means that the error in focal length can be corrected by iteratively using the composited length to compute a new and more correct focal length. This *compositing approach to camera calibration* has the advantages of not requiring both feature detection and separate prior calibration.

Keywords: 3-D scene modeling, image compositing, camera calibration.

©Digital Equipment Corporation 1996. All rights reserved.

---

[1]Computer and Information Science Dept., University of Massachusetts at Amherst, Amherst, MA 01003

Contents

List of Figures

1 Introduction

A panoramic image has a 360° horizontal field of view. Panoramic images of scenes have interesting applications in computer vision and visualization, since they are able to provide the viewer the impression of being immersed in the scene to a certain extent. For example, Apple's QuickTime VR^{TM} [Chen, 1995] product uses panoramas for scene visualization. In computer vision, applying the stereo algorithm on multiple panoramas allows the entire 3D scene data points to be extracted [Kang and Szeliski, 1995] and subsequently modeled [Kang *et al.*, 1995].

A panoramic image is produced by following a series of steps: First, a sequence of images is taken while rotating a camera about a vertical axis that passes through the camera optical center. Each image in the sequence is then projected onto a cylindrical surface whose cross-sectional radius is an initially estimated focal length (see Figure 1). The panoramic image is subsequently created by determining the relative displacements between adjacent images in the sequence and compositing the displaced sequence of images. The length of the panoramic image is termed the *compositing length*.

1.1 Analyzing the error in compositing length

In this technical report, we describe the effect of errors in intrinsic camera parameters on the compositing length. We are not aware of any prior work in this specific area. In particular, we consider the focal length and the radial distortion coefficient. An important consequence of this analysis is that a much better estimate of the camera focal length can be calculated from the current compositing length. Hence by iterating the process of projecting onto a cylindrical surface (whose cross-sectional radius is the latest estimation of the camera focal length) and compositing the new sequences, we quickly arrive at the camera focal length within a specified error tolerance. We show later that the convergence towards the true focal length is exponential. This method constitutes a simple means of calibrating a camera.

1.2 Camera calibration

Conventional means of camera calibration use a calibration or control pattern (e.g., points [Szeliski and Kang, 1994; Tsai, 1987; Weng *et al.*, 1992], lines [Beardsley and Murray, 1992; Caprile and Torre, 1990; Wang and Tsai, 1991]), or take advantage of feature structural constraints (e.g., roundness of spheres [Penna, 1991; Stein, 1995], straightness of lines [Brown, 1971; Stein, 1995]). We propose a method for featureless camera calibration based on an iterative scheme of projecting rectilinear images to cylindrical images and then compositing. In addition to determining the camera focal length, this technique results in a panoramic image (i.e., with 360° horizontal field of view) that is both physically correct (ignoring radial distortion effects) and seamlessly blended.

The basis of the compositing approach to camera calibration is the discovery that the relative compositing length error due to camera focal length error is disproportionately much less (i.e., in terms of percentages) than the relative focal length error. When a planar image is projected to a cylinder as in the compositing process, mis-estimation of the radius of the cylinder will produce an erroneous warping of the image that will affect the length of the final composite image. However, it turns out that near the center of the overlap between successive images, the amount of combined warping for both images is minimal and so is the effect of the mis-estimation. The result of this is that the percent error in length of the composite image is less than the percent error in the focal length. This makes it possible to use the composite length to recover a better estimate of the camera focal length.

The proposed technique has the advantage of not having to know the camera focal length when a panorama is to be generated from a sequence of images. This is in contrast to Apple's QuickTime VR^{TM} [Chen, 1995], which we believe that a reasonably good estimate of camera focal length is required *a priori*. This is also the case for McMillan and Bishop's method of creating panoramas [McMillan and Bishop, 1995]. Their method of estimating the camera focal length necessitates small panning rotations and relies on translation estimates near the image centers.

The method of calibration that is closest to ours is that of Stein's [Stein, 1995], in which features are tracked throughout the image sequence taken while the camera is rotated a full 360°. While this technique results in accurate camera parameters, it still requires feature detection and tracking. Our technique directly uses the given image sequence of the scene to determine camera focal length without relying on specific tracked features.

1.3 Motivation and outline

The motivation for generating panoramic images is to directly recover 3-D scene data points over a wide field of view using stereo [Kang and Szeliski, 1995] for subsequent modeling and photorealistic rendering [Kang *et al.*, 1995]. Traditional approaches to recovering 3-D data of a wide scene is

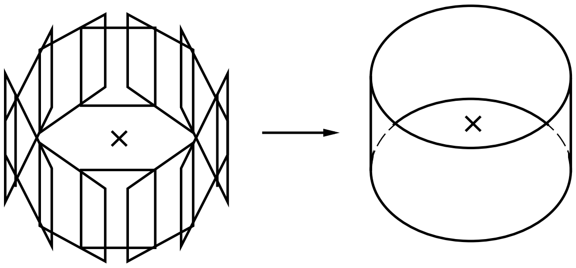


Figure 1: Compositing multiple rotated camera views into a panorama. The '$\times$' marks indicate the locations of the camera optical and rotation center.

to take stereo snapshots of the scene at various poses and then merge these 3-D stereo depth maps. This is not only computationally intensive, but the resulting merged depth maps may be subject to merging errors, especially if the relative poses between depth maps are not known exactly. The 3-D data may also have to be resampled before merging, which adds additional complexity and potential sources of errors.

The outline of this technical report is as follows: Section 2 reviews how a panoramic image is produced from a set of images. This is followed by section 3 which gives a detailed analysis of the compositing error due error in the camera focal length in Section 3. A consequence of this analysis is the iterative compositing approach to camera calibration. Section 4 looks at the effect of misestimating the radial distortion coefficient on the panoramic compositing length. We then describe the effect of errors in both focal length and radial distortion coefficient on the reconstructed 3-D data in section 5 before summarizing in section 6.

2 Generating a panoramic image

A panoramic image is created by compositing a series of rotated camera image images, as shown in Figure 1. In order to create this panoramic image, we first have to ensure that the camera is rotating about an axis passing through its optical center, i.e., we must eliminate motion parallax when panning the camera around. To achieve this, we manually adjust the position of camera relative to an X-Y precision stage (mounted on the tripod) such that the motion parallax effect disappears when the camera is rotated back and forth about the vertical axis [Stein, 1995].

In previous work [Kang and Szeliski, 1995], the camera is first calibrated to extract its intrinsic parameters, namely κ, the radial distortion coefficient, and f, the camera focal length. This is accomplished by taking snapshots of a calibration dot grid pattern at known spacings and using the iterative least squares algorithm described in [Szeliski and Kang, 1994]. As a result of our analysis reported here, this calibration step can be skipped if the radial distortion coefficient is insignificant.

A panoramic image is created using the following steps [Kang and Szeliski, 1995]:

1. Capture a sequence of rotated camera views about a vertical axis passing through the camera optical center;
2. Undistort the sequence to correct for κ;
3. Warp the undistorted (rectilinear) sequence to produce a corresponding cylindrical-based image sequence whose cross-sectional radius is equal to the camera focal length f; and finally
4. Composite the sequence of images [Szeliski, 1994].

The compositing technique comprises two steps: rough alignment using phase correlation, and iterative local refinement to minimize overlap intensity difference between successive images. In both steps, the translation is assumed to be in one direction only, namely in the x-direction (since the cylindrical-based images have been "flattened" or unrolled). This is a perfectly legitimate assumption, since camera motion has been constrained to rotate about a vertical axis during image sequence capture. If the estimated focal length is exact, then the error in the composited length is due to the digitization and image resampling effects, the limit in the number of iterations during local matching, and computer truncation or rounding off effects.

The rough alignment step has been made more robust by adding the iterative step of checking the displacement corresponding to the peak—if the intensity RMS error in matching the overlap regions is high, the peak is tagged as false and the next highest peak is chosen instead.

3 Compositing errors due to misestimation of focal length

Compositing errors occur as a result of using a wrong value of the camera focal length in converting the rectilinear images to cylindrical-based images prior to compositing. If the correct focal length used, say f_{true}, then the expected length of the composited panoramic image[1] is

$$L = 2\pi f_{true} \tag{1}$$

If the focal length f used is incorrect, then the mapped cylindrical images are no longer physically correct. The compositing step will attempt to minimize the error in the overlap region of successive images, but there is still a net error in compositing length L.

Since each column in a rectilinear image is projected to another column in the cylindrical image and translation is constrained to be along the x-direction, it suffices to consider only a scanline in our analysis. We assume for simplicity that the images are "fully textured" so that matching between corresponding pixels is unambiguous. We also assume, for ease of analysis, that the amount of camera rotation between successive frames is the same throughout the sequence (this need not be so in practice). In our analysis, the net translation is computed by minimizing the sum of squares of the pixel displacements from their matching locations. In other words, even after translation with interpolation, the pixels in the second image will not match the pixels in the first image at the same location. Each pixel will match one at a displaced location. The translation which minimizes the sum of their squares is the one that results in zero average displacement.

3.1 Derivation

In order to model the displacement of each pixel u_i in the second cylindrical image, we map it back to t_i in the image plane, find the corresponding pixel s_i in the first image based on the actual rotation α, and map that back to v_i in the cylindrical image. This is illustrated in Fig. 2, where I_1 and I_2 are the first and second cylindrical images, respectively. $I_{1,true}$ is the true cylindrical first image while α is the amount of actual camera rotation between successive frames, i.e., $2\pi/N$, N being the number of images in the sequence. Recall that the cylindrical images are formed by warping the images into a cylindrical surface whose cross-sectional radius is the estimated focal length. The mappings are given by the following equations:

$$t_i = f \tan\left(\frac{u_i}{f}\right)$$

[1]In creating the panoramic image, the order of compositing is I_1, I_2, ..., I_{N-1}, I_N, I_1, where I_k is the kth image in the sequence and N is the number of images in the sequence. The compositing length L is actually the displacement of the first frame I_1 relative to its original location.

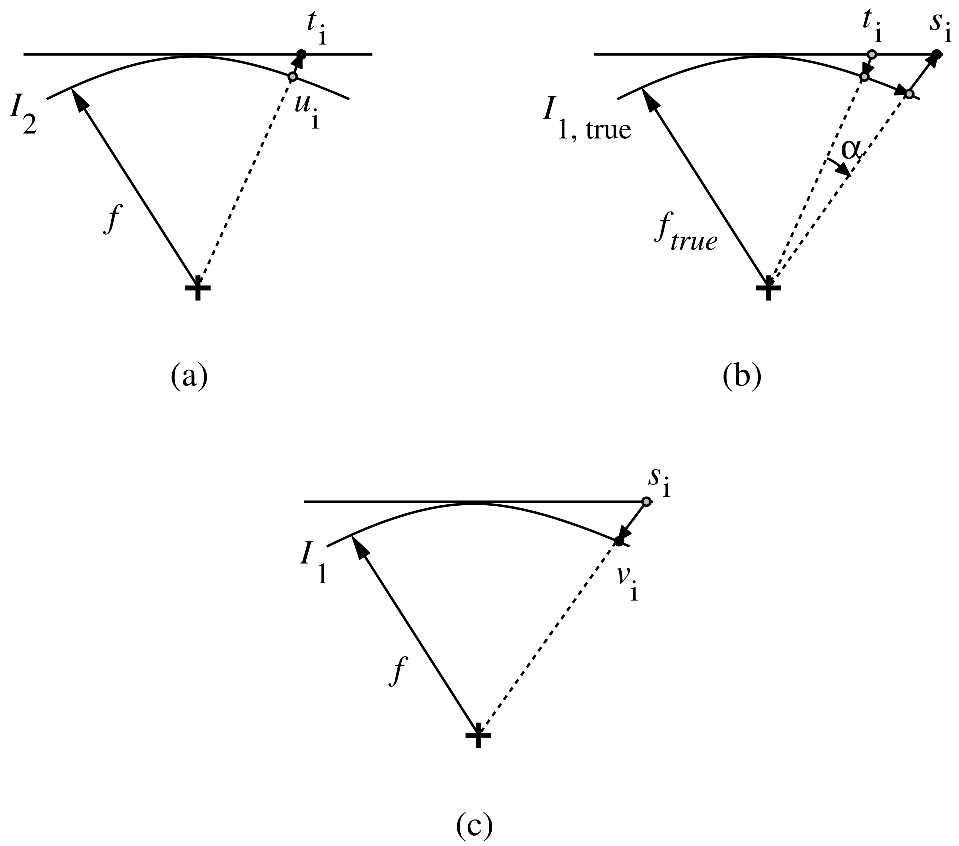


Figure 2: Effect of inexact focal length: (a) actual mapping from second cylindrical image; (b) theoretical displacement; (c) actual mapping to first cylindrical image. See text.

$$\begin{aligned} s_i &= f_{true} \tan\left(\tan^{-1}\left(\frac{t_i}{f_{true}}\right) + \alpha\right) \\ v_i &= f \tan^{-1}\left(\frac{s_i}{f}\right) \end{aligned} \tag{2}$$

As before, $\alpha = 2\pi/N$ and f_{true} is the correct focal length while f is the estimated focal length used.

Using Mathematica$^{\mathrm{TM}}$ [Wolfram, 1991], we find that, up to the third order in u_i and α,

$$\begin{aligned} v_i(u_i) &= u_i + \frac{f^2 f_{true}^2 + f^2 u_i^2 - f_{true}^2 u_i^2}{f^2 f_{true}} \alpha \\ &+ \frac{(f - f_{true})(f + f_{true})(3f^2 f_{true}^2 u_i + 3f^2 u_i^3 - 5 f_{true}^2 u_i^3)}{3 f^4 f_{true}^2} \alpha^2 \end{aligned}$$

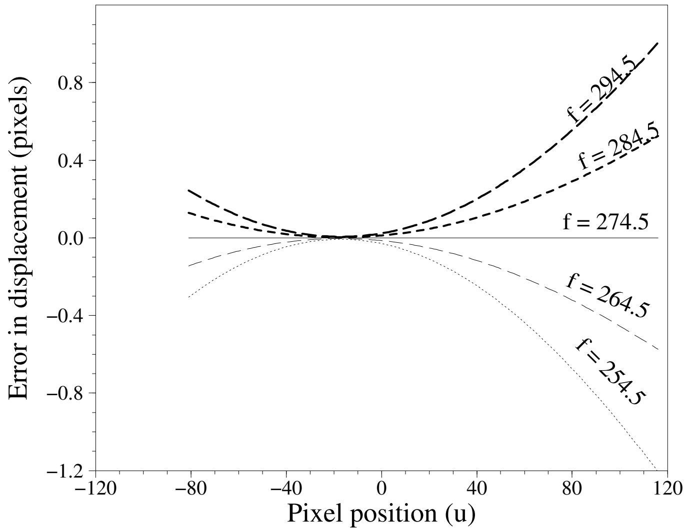


Figure 3: Graph of error in displacement vs. pixel location for varying estimated focal length f. $f_{true} = 274.5$, $N = 50$, and $l = 232$.

$$+ \quad \frac{(f - f_{true})(f + f_{true})(f^2 f_{true}^2 + 4f^2 u_i^2 - 6 f_{true}^2 u_i^2)}{3 f^4 f_{true}} \alpha^3 \tag{3}$$

The displacement between two successive frames at u_i is $d_i(u_i) = v_i(u_i) - u_i$; the plot of the variation of $d_i(u_i)$ versus u_i for $f_{true} = 274.5$ and $N = 50$ for different values of misestimated values of f is shown in Fig. 3. It is interesting to note that the minimum displacement due to focal length error occurs near the *center* of the overlap. The change in the displacement error distribution due to the amount of overlap (changing number of frames N) is shown in Fig. 4. As can be observed, as N increases, the amount of overlap increases, and interestingly enough, the overall error decreases. If all are kept constant except for the image length l, the error distribution remains the same save for u_1 and u_2, the two end pixel locations of the overlap area. They shrink to decrease the amount of horizontal overlap as l decreases. The mean displacement between two successive frames (see Fig. 5) is

$$\bar{D} = \frac{\sum_{u_i=u_1}^{u_2} (v_i(u_i) - u_i)}{u_2 - u_1 + 1} \tag{4}$$

Note that if the interframe displacement is equal throughout the sequence and that the length of each image is l, then $u_2 = l/2$ and $u_1 = 2\pi f_{true}/N - l/2$. The *mean* total displacement, i.e., the

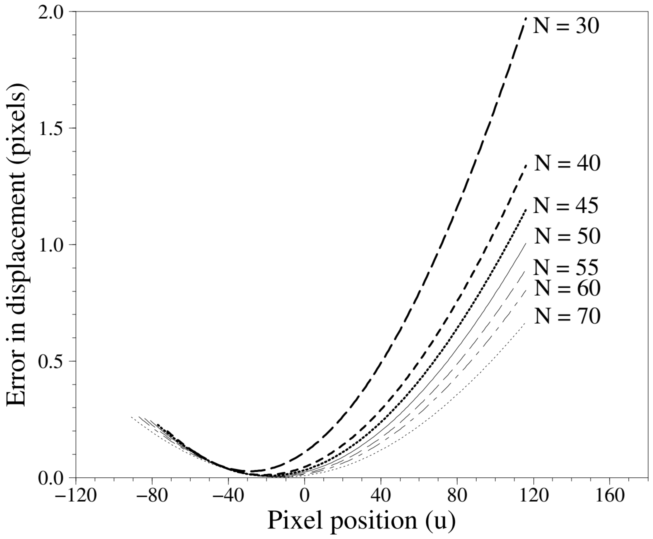


Figure 4: Graph of error in displacement vs. pixel location for varying number of frames N. $f_{true} = 274.5$, $f = 294.5$ and $l = 232$.

composite length, is given by

$$L = N\bar{D} \tag{5}$$

Suppose we reestimate the focal length from the composite length, namely $f' = L/(2\pi)$. The question is: Is f' a better estimate of the focal length than f? It turns out that for initial estimates of f close to the true value, the answer can be shown to be yes. To see this, if $f \approx f_{true}$, then from (3), (4) and (5), we get

$$\begin{aligned} f' = \frac{L}{2\pi} &\approx \frac{N}{2\pi}\left\{\left[f_{true} + \left(1 - \frac{f_{true}^2}{f^2}\right)\frac{\overline{u^2}}{f_{true}}\right]\alpha\right. \\ &+ \left[\left(1 - \frac{f_{true}^2}{f^2}\right)\overline{u} - \frac{1}{3}\left(1 - \frac{f_{true}^2}{f^2}\right)\left(5\frac{f_{true}^2}{f^2} - 3\right)\frac{\overline{u^3}}{f_{true}^2}\right]\alpha^2 \\ &+ \left.\left[\frac{1}{3}f_{true}\left(1 - \frac{f_{true}^2}{f^2}\right) - 2\left(1 - \frac{f_{true}^2}{f^2}\right)\left(\frac{f_{true}^2}{f^2} - \frac{2}{3}\right)\frac{\overline{u^2}}{f_{true}}\right]\alpha^3\right\} \\ &\approx f_{true} + \left(1 - \frac{f_{true}^2}{f^2}\right)\frac{\overline{u^2}}{f_{true}} \end{aligned}$$

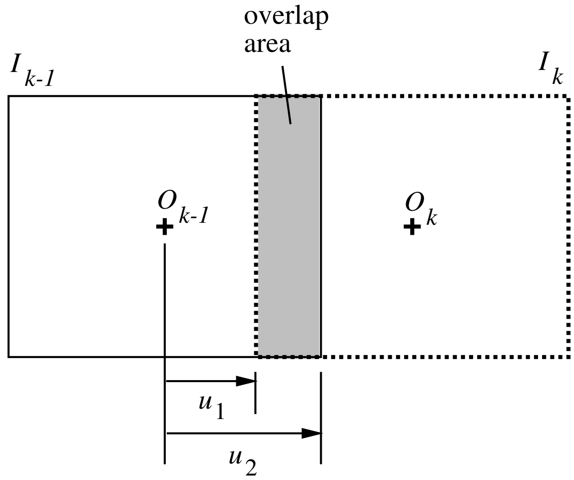


Figure 5: Overlap between successive images I_{k-1} and I_k, with centers at O_{k-1} and O_k respectively.

$$
\begin{aligned}
&+ \left[\left(1-\frac{f_{true}^2}{f^2}\right)\overline{u} - \frac{1}{3}\left(1-\frac{f_{true}^2}{f^2}\right)\left(5\frac{f_{true}^2}{f^2}-3\right)\frac{\overline{u^3}}{f_{true}^2}\right]\alpha \\
&+ \left[\frac{1}{3}f_{true}\left(1-\frac{f_{true}^2}{f^2}\right) - 2\left(1-\frac{f_{true}^2}{f^2}\right)\left(\frac{f_{true}^2}{f^2}-\frac{2}{3}\right)\frac{\overline{u^2}}{f_{true}}\right]\alpha^2 \qquad (6)
\end{aligned}
$$

noting that $\alpha = 2\pi/N$, and where

$$
\begin{aligned}
\overline{u} &= \frac{\sum_{i=u_1}^{u_2} i}{u_2 - u_1 + 1} \\
&= \frac{1}{2}(u_2 + u_1) = \frac{\pi f_{true}}{N}, \qquad (7)
\end{aligned}
$$

$$
\begin{aligned}
\overline{u^2} &= \frac{\sum_{i=u_1}^{u_2} i^2}{u_2 - u_1 + 1} \\
&= \frac{1}{3}\left(u_2^2 + u_1 u_2 + u_1^2 + \frac{1}{2}(u_2 - u_1)\right) = \frac{1}{3}\left((u_2+u_1)^2 - u_2 u_1 + \frac{1}{2}(u_2 - u_1)\right) \\
&= \frac{1}{3}\left[\frac{4\pi^2 f_{true}^2}{N^2} - \left(\frac{2\pi f_{true}}{N} - \frac{l}{2}\right)\frac{l}{2} + \frac{\pi f_{true}}{N} - \frac{l}{2}\right] \qquad (8)
\end{aligned}
$$

and

$$
\overline{u^3} = \frac{\sum_{i=u_1}^{u_2} i^3}{u_2 - u_1 + 1}
$$

$$
\begin{aligned}
&= \frac{1}{4}(u_2+u_1)\left(u_2^2+u_1^2+u_2-u_1\right) \\
&= \frac{1}{2}\frac{\pi f_{true}}{N}\left[\left(\frac{2\pi f_{true}}{N}-\frac{l}{2}\right)^2+\frac{l^2}{4}+\frac{2\pi f_{true}}{N}-l\right] \qquad (9)
\end{aligned}
$$

Hence

$$
\begin{aligned}
f_{true}-f' &\approx \left\{\frac{\overline{u^2}}{f_{true}}+\left[\overline{u}-\frac{1}{3}\left(5\frac{f_{true}^2}{f^2}-3\right)\frac{\overline{u^3}}{f_{true}^2}\right]\alpha\right. \\
&+ \left.\left[\frac{f_{true}}{3}-2\left(\frac{f_{true}^2}{f^2}-\frac{2}{3}\right)\frac{\overline{u^2}}{f_{true}}\right]\alpha^2\right\}\left(\frac{f_{true}^2}{f^2}-1\right) \\
&\approx \frac{f_{true}}{f}\frac{f_{true}+f}{2f}2\left\{\frac{\overline{u^2}}{f_{true}^2}+\left(\frac{\overline{u}}{f_{true}}-\frac{2}{3}\frac{\overline{u^3}}{f_{true}^3}\right)\alpha+\left(\frac{1}{3}-\frac{2}{3}\frac{\overline{u^2}}{f_{true}^2}\right)\alpha^2\right\}(f_{true}-f) \\
&\approx 2\left\{\frac{\overline{u^2}}{f_{true}^2}+\left(\frac{\overline{u}}{f_{true}}-\frac{2}{3}\frac{\overline{u^3}}{f_{true}^3}\right)\alpha+\frac{1}{3}\left(1-2\frac{\overline{u^2}}{f_{true}^2}\right)\alpha^2\right\}(f_{true}-f) \qquad (10)
\end{aligned}
$$

Let

$$\beta_1 = \frac{\overline{u}}{f_{true}} = \frac{\pi}{N}, \qquad (11)$$

$$\beta_2 = \frac{\overline{u^2}}{f_{true}^2} = \frac{1}{3}\left[\frac{4\pi^2}{N^2}-\left(\frac{2\pi}{N}-\frac{l}{2f_{true}}\right)\frac{l}{2f_{true}}+\frac{1}{f_{true}}\left(\frac{\pi}{N}-\frac{l}{2f_{true}}\right)\right] \qquad (12)$$

and

$$\beta_3 = \frac{\overline{u^3}}{f_{true}^3} = \frac{1}{2}\frac{\pi}{N}\left[\left(\frac{2\pi}{N}-\frac{l}{2f_{true}}\right)^2+\left(\frac{l}{2f_{true}}\right)^2+\frac{2}{f_{true}}\left(\frac{\pi}{N}-\frac{l}{2f_{true}}\right)\right] \qquad (13)$$

Thus (10) becomes

$$f_{true}-f' \approx 2\left[\beta_2+\left(\beta_1-\frac{2}{3}\beta_3\right)\alpha+\frac{1}{3}(1-2\beta_2)\alpha^2\right](f_{true}-f) = \eta(f_{true}-f) \qquad (14)$$

If N is large, which is typical (in practice, N is about 50), then $|f_{true}-f'| \ll |f_{true}-f|$. This implies that the estimated focal length based on the composited length is a significantly better estimate.

3.2 Image compositing approach to camera calibration

The previous result suggests a direct, iterative method of simultaneously determining the camera focal length and constructing a panoramic image. This *iterative image compositing approach to camera calibration* has the advantages of not requiring both feature detection and separate prior calibration. The pseudocode associated with this method is as follows:

```
Let the initial estimate of focal length be f_0.
Determine compositing length L_0 from f_0.
Set k = 1.
    1.  Calculate f_k = L_{k-1}/(2π).
    2.  Determine compositing length L_k from f_k.
    3.  If (|L_k − L_{k-1}| ≥ ε) {
            k ← k + 1
            Go to Step 1.
        }
        else
            f_k is the final estimated focal length.
```

Since we know that the iterated value of f_k converges toward f_{true}, it would be interesting to determine its rate of convergence. (14) can be rewritten as a recurrence equation (assuming equality rather than approximation)

$$f_{true} - f_k = \eta \left(f_{true} - f_{k-1}\right) \tag{15}$$

Rearranging, we have

$$f_k - \eta f_{k-1} = (1-\eta) f_{true}, \tag{16}$$

from which the solution can be found to be

$$f_k = f_{true} + \left(f_0 - f_{true}\right) \eta^k \tag{17}$$

Hence, the convergence of f_k towards f_{true} is exponential in the vicinity of the true focal length, as shown by (17). This also indicates that the convergence is faster if the number of frames N increases, the image length l decreases, or the true focal length f_{true} increases. As an example, for $N = 50$, $l = 232$, and $f_{true} = 274.5$, $\eta = 0.117$.

The graph in Figure 6 shows the convergence of estimated focal length from different initial estimates. (A sequence of the synthetic room is shown in Figure 7 and the corresponding composited image is shown in Figure 8.) It is interesting to note that the actual estimated focal lengths are *smaller* than theoretically predicted ones. One of the reasons could be due to effects of resampling using bilinear interpolation. In addition, we also make the assumption that each point is "fully

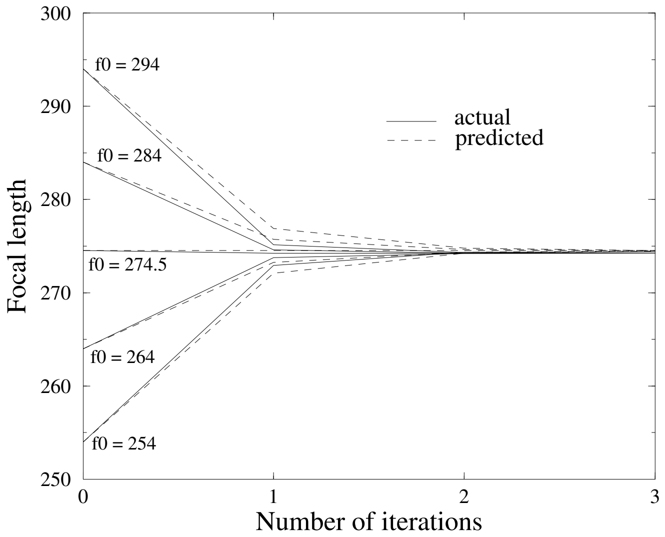


Figure 6: Graph of estimated focal length vs. number of iteration. f0 is the initial estimated focal length; the actual focal length is 274.5. The solid lines represent actual values whereas dashed lines represent predicted values.

textured," which is difficult to realize in practice and even in simulations. Finally, shifts greater than 1 pixel are not likely to influence the net shift correctly.

A panorama of the synthetic room is shown in Figure 9. As can be seen, the effect of misestimating the focal length in compositing is a blurring effect, presumably about the correct locations. When a rectilinear image is projected onto a cylindrical surface of the wrong cross-sectional radius, which is also the estimated focal length, the error in pixel placement increases away from the central image column. Having many cylindrical-converted images superimposed would thus have the effect of locally smearing the correct locations. This suggests that a good scheme of compositing many images to form a panorama is to down-weight the pixels away from the central image column during compositing. Indeed, Figure 10 shows the effect of using such a simple scheme. Here each pixel is weighted by a factor proportional to $|c - c_{center}|^{\xi}$, where c is the current pixel column, c_{center} the central pixel column, and $\xi = -5$. This yields a panorama that visually appears almost as good as that shown in Figure 8. Note, however, that the panorama in Figure 10 is still not quite physically correct; the *aspect ratio* is still not exact.

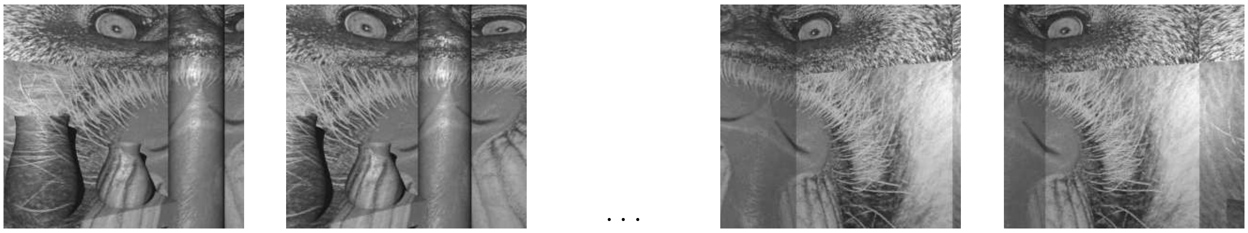


Image 1 Image 2 Image (N-1) Image N

Figure 7: Example undistorted image sequence of synthetic room.

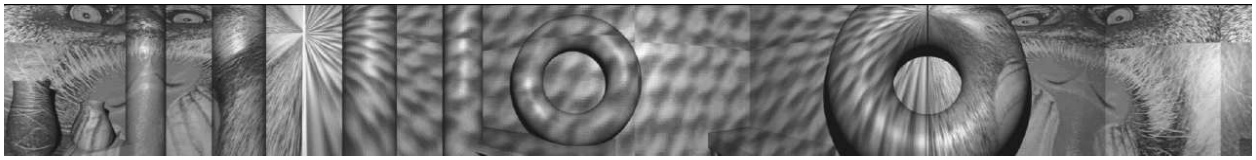


Figure 8: Panorama of synthetic room after compositing the sequence in Figure 7.

To further illustrate the robustness of this approach, we have also started the iterative process with the *original rectilinear images*, (i.e., $f_0 = \infty$), which would be the worst case focal length initialization. The convergence of the focal length value is: $\infty \to 281.18 \to 274.40 \to 274.24$. As before, the actual focal length is 274.5. The process arrives at virtually the correct focal length in just two iterations. This result is very significant; it illustrates that in principal, we can start without a focal length estimate.

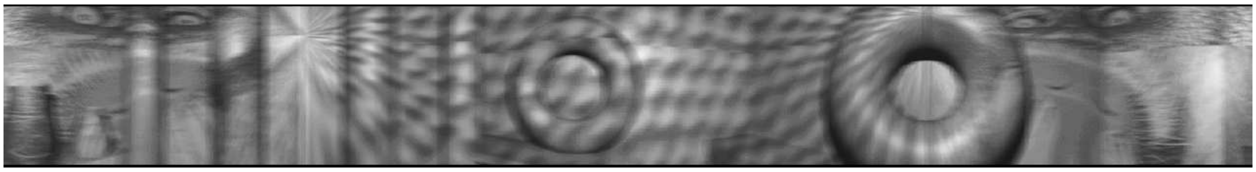


Figure 9: Panorama of synthetic room corresponding to an erroneous focal length.

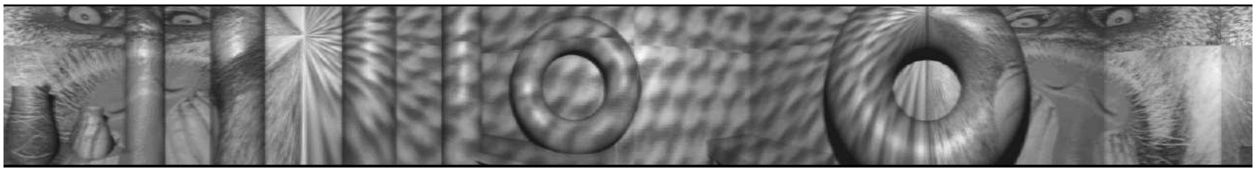


Figure 10: Panorama of synthetic room corresponding to an erroneous focal length, but using a simple weighted compositing technique.

4 Compositing errors due to misestimation of radial distortion coefficient

Another important camera intrinsic parameter that could cause errors in the compositing length is the radial lens distortion. If (x_u, y_u) is the undistorted image location and (x_d, y_d) is its radially distorted counterpart, then

$$\begin{aligned} x_u &= x_d(1 + \kappa_1 r_d^2 + \kappa_2 r_d^4 + \ldots) \\ y_u &= y_d(1 + \kappa_1 r_d^2 + \kappa_2 r_d^4 + \ldots) \end{aligned} \tag{18}$$

where

$$r_d = \sqrt{x_d^2 + y_d^2}$$

For our work, we use only the first radial coefficient term $\kappa = \kappa_1$:

$$\begin{aligned} x_u &= x_d(1 + \kappa r_d^2) \\ y_u &= y_d(1 + \kappa r_d^2) \end{aligned} \tag{19}$$

with the inverse

$$\begin{aligned} x_d &= \frac{x_u}{1 + \kappa r_d^2} \\ y_d &= \frac{y_u}{1 + \kappa r_d^2} \end{aligned} \tag{20}$$

where

$$r_d^2 = \left(\sqrt{\left(\frac{1}{27\kappa^3} + \frac{r_u^2}{2\kappa^2} \right)^2 - \frac{1}{729\kappa^6}} + \frac{1}{27\kappa^3} + \frac{r_u^2}{2\kappa^2} \right)^{\frac{1}{3}}$$

$$\begin{tikzcd}
\mathrm{Cyl}_2 \arrow[r, "f^{-1}"] & \mathrm{Rect}_2 \arrow[d, "\kappa^{-1}"'] & \mathrm{Cyl}_{2,\mathrm{true}} \arrow[r, "\text{rotate by } \alpha"] & \mathrm{Cyl}_{1,\mathrm{true}} \arrow[d, "f^{-1}_{\mathrm{true}}"] & \mathrm{Rect}_1 \arrow[r, "f"] & \mathrm{Cyl}_1 \\
& \mathrm{Rect0}_2 \arrow[r, "\kappa_{\mathrm{true}}"'] & \mathrm{Rect'}_2 \arrow[u, "f_{\mathrm{true}}"'] & \mathrm{Rect'}_1 \arrow[r, "\kappa^{-1}_{\mathrm{true}}"'] & \mathrm{Rect0}_1 \arrow[u, "\kappa"'] &
\end{tikzcd}$$

Figure 11: Mapping of pixels from the second cylindrical image to the first. The transformation f indicates mapping from cylindrical to rectilinear coordinates with focal length f while transformation κ indicates the radial distortion mapping with radial distortion factor κ. Terms with the subscript "true" represent the correct entities while those without this subscript represent estimated ones. Rect0 is the undistorted rectilinear image. See text.

$$+\frac{1}{9\left(\sqrt{\left(\frac{1}{27\kappa^3}+\frac{r_u^2}{2\kappa^2}\right)^2-\frac{1}{729\kappa^6}}+\frac{1}{27\kappa^3}+\frac{r_u^2}{2\kappa^2}\right)^{\frac{1}{3}}\kappa^2}-\frac{2}{3\kappa} \qquad (21)$$

(21) is found using Mathematica$^{\mathrm{TM}}$ [Wolfram, 1991]. Details of lens distortion modeling can be found in [Slama, 1980].

The transformations required to show the effect of incorrect focal length and radial distortion coefficient are depicted in Figure 11. We assume that the cylindrical images are displaced by an angular amount α. To see how these transformations come about, consider the right half of the series of transformations beyond "rotate by α." We require the mapping from the correct cylindrical image point to the actual cylindrical image point, given estimates of f and κ. To generate the correct undistorted rectilinear image, we have to unproject from the cylindrical surface to the flat rectilinear surface (f_{true}^{-1}) and then radially undistort (κ_{true}^{-1}). Subsequently we perform radial distortion (κ) and cylindrical projection (f) to arrive at the estimated cylindrical image. This is similarly done for the second image. Equations (3), (19), (20), and (21) are used in series to determine the theoretical displacement, as is similarly done in section 3. The difference is that in calculating the mean displacement, the displacements are averaged over *all* the pixels in the image. This is because radial distortion changes both x and y coordinates, while the cylindrical projection changes the x component independently of y. In addition, if the camera axis passes through the

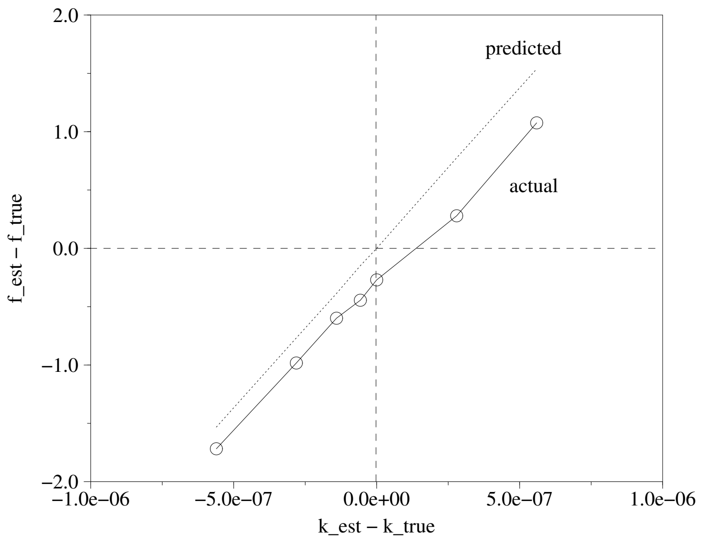


Figure 12: Graph of equivalent focal length error vs. error in κ, the radial distortion factor. The true focal length (f_{true}) is 274.5 and the true radial distortion factor (κ_{true}) is 2.8×10^{-7}.

image center row, the average displacement in y is zero.

The effect of misestimating the radial distortion coefficient κ for a typical value of $f = 274.5$ and $\kappa = 2.8 \times 10^{-7}$ is shown in Figure 12. As can be seen, the effect is almost linear, and despite significant errors in κ, the resulting error in the effective focal length is small ($< 1\%$). This illustrates that for typical real focal lengths and radial distortion coefficients, the dominant factor in the compositing length error is the accuracy of the focal length.

The appearance of the panorama due to error in radial distortion coefficient κ is not very perceptible if the radial distortion is typically small (of the order of 10^{-7}). An extreme case that corresponds to a large error in radial distortion coefficient (by 10^{-5}) can be seen in Figure 13. Here, a simple scheme of compositing by direct averaging is performed, and there is a perceptible ghosting effect. However, using the weighted compositing scheme results in a much sharper image, as shown in Figure 14. There is still some blurring effects, which is more pronounced away from the central horizontal row of the panorama, but this is to be expected with errors in κ.

There are two ways of measuring compositing length error: mismatch between observed compositing length and expected compositing length based on estimated focal length, and mismatch

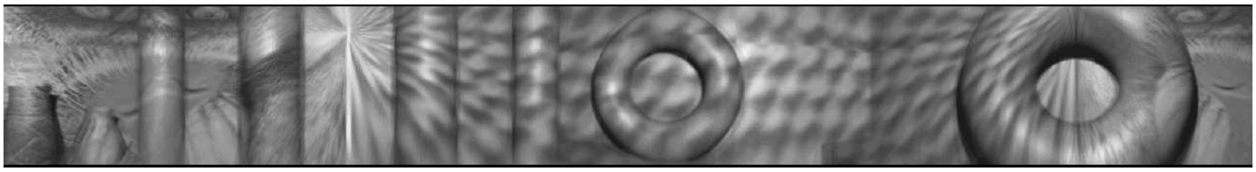


Figure 13: Another panorama of synthetic room corresponding to a large erroneous radial distortion coefficient (by 1.0×10^{-5}).

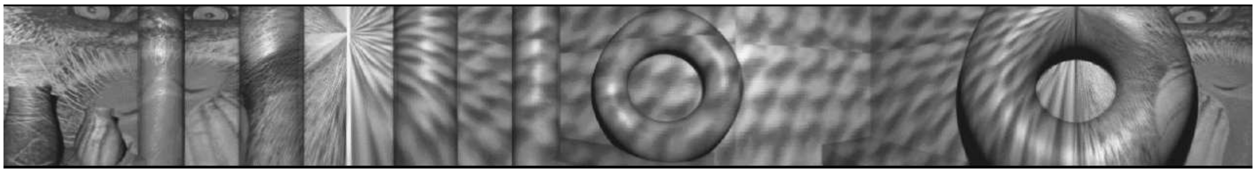


Figure 14: Panorama of synthetic room (same camera location as in Figure 13) corresponding to an erroneous focal length.

between the correct compositing length and expected compositing length. The first error ($dL0$) measures the consistency between the estimated focal length and the observed composite length. The second error ($dL1$) metric measures the error due to the current estimate of the focal length, and cannot be calculated unless the true focal length is known. Figure 15 shows the variation of both types of compositing length error as a function of errors in estimated focal length and radial distortion coefficient. (The nominal focal length and radial distortion coefficient are 274.5 and 2.8×10^{-7} respectively.) The error in focal length is expressed as $df = f - f_{true}$. The error $dL0$ is $2\pi f - L$, where L is the compositing length and f is the estimated focal length. This is relevant if the estimated focal length is assumed to be correct and the composited length is adjusted to be compatible with the estimated focal length. In this case, the image displacement errors are distributed over all the frames (the simpliest method being uniform distribution). This procedure involves the least amount of computation as the images do not require reprojection onto a cylindrical surface of a difference cross-sectional radius (i.e., focal length). It may be used in the case of accurately estimated focal lengths. Meanwhile, the error $dL1$ is $2\pi(f - f_{true})$, f_{true} being the correct focal length. This is relevant in the case of using the newly estimated focal length based on the composited length. As can be observed, both types of compositing length errors are more sensitive to the error in the estimated focal length, with $dL0$ much more so.

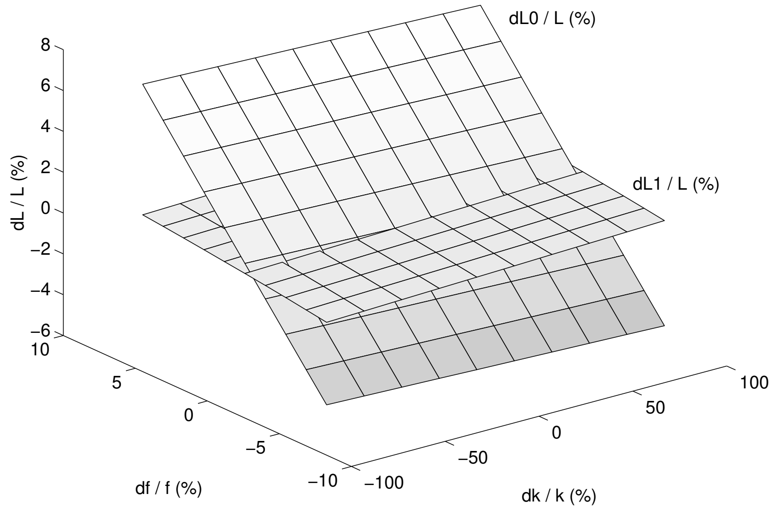


Figure 15: Variation of compositing length error vs. errors in both focal length and radial distortion coefficient. The deviations are all in terms of percentages. The nominal focal length and radial distortion coefficient are 274.5 and 2.8×10^{-7} respectively. See text for descriptions of $dL0$ and $dL1$.

5 Effect of error in focal length and radial distortion coefficient on 3-D data

The recovered 3-D data do depend on the accuracy of the estimated focal length and radial distortion coefficient. This can be seen from Figures 17 and 18. The length and breadth of the synthetic room are 10 and 8 units respectively. Stereo data was recovered from 3 camera locations; two camera locations are $0.3\sqrt{2}$ units away from the first or reference camera location. An example of a distribution of recovered stereo data corresponding to the correct focal length of 274.5 and no radial distortion is shown in Figure 16. Surprisingly, despite the increased numerical errors, the recovered 3-D data corresponding to the other (erroneous) focal lengths and radial distortion coefficient do not appear significantly different from that shown in Figure 16. This suggests that if exact reconstruction is not required and that the panorama does not have to be of high quality, then just using the estimated focal length directly would suffice.

From Figure 17, it can be observed that the effect of underestimating the focal length is greater

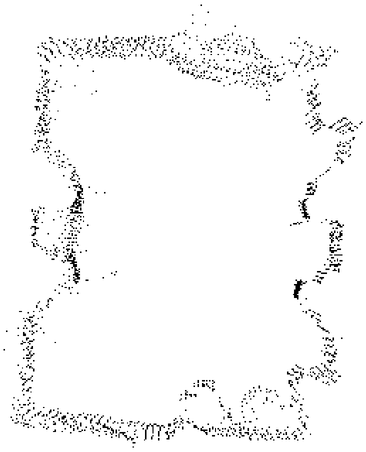


Figure 16: Example recovered 3-D data (corresponding to the correct focal length of 274.5).

than overestimating it. This is most likely due to the greater relative change in the curvature error (the curvature of the cylindrical surface being inversely proportional to the cross-sectional radius, which is the focal length) in underestimating the focal length. In addition, the effect of misestimating the focal length appears to be more significant on the accuracy of the reconstructed 3-D points than does misestimating the radial distortion coefficient (assuming typical values of κ of the order of 10^{-7}). This suggests that as long as the field of view is not too large as to result in significant radial distortion, we can get by with a simple estimation of κ, or by assuming no radial distortion.

6 Summary

We have analyzed the compositing error in terms of two intrinsic camera parameters, namely the focal length and the radial distortion coefficient. Given typical values of the radial distortion coefficient, the effect of the focal length on the compositing error is more significant than that of the radial distortion coefficient. An important discovery from this analysis is that the relative compositing length error due to camera focal length error is disproportionately much less (i.e., in terms of percentages) than the relative focal length error. This enables the use of the resulting compositing length to recover a better estimate of the camera focal length, and forms the basis of the iterative compositing approach to camera calibration. This method has the advantage of not having

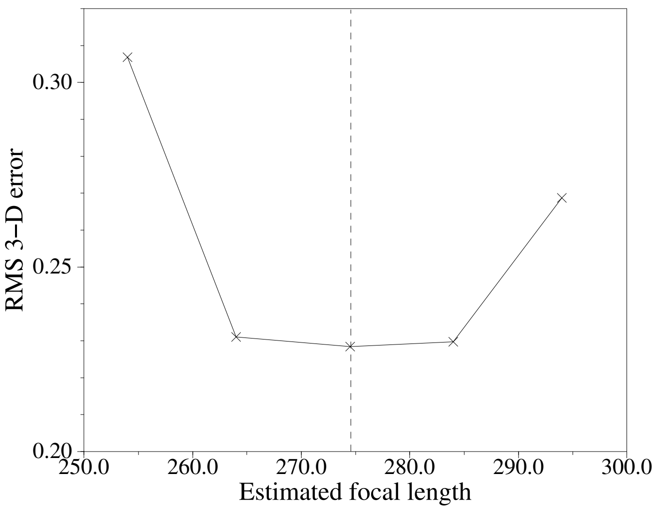


Figure 17: Graph of RMS 3D error of recovered stereo data vs. estimated focal length. The true focal length is 274.5 (indicated by the vertical dashed line).

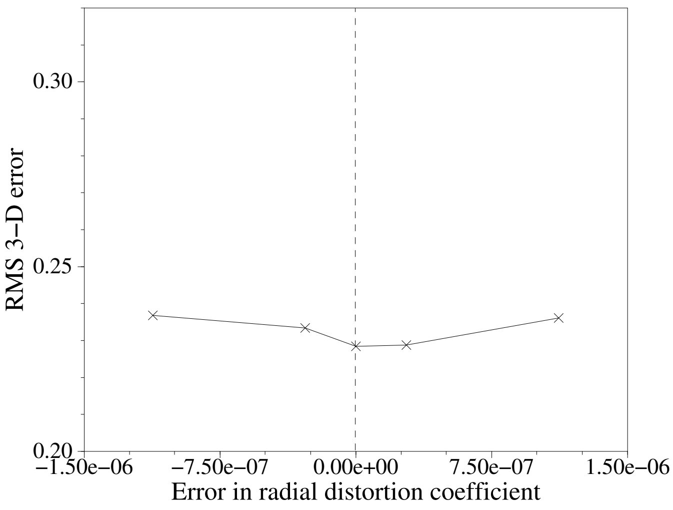


Figure 18: Graph of RMS 3D error of recovered stereo data vs. error in radial distortion coefficient κ. Typical real values of κ is of the order of 10^{-7}. The zero error in κ is indicated by the vertical dashed line.

to know the camera focal length when a panorama is to be generated from a sequence of images. In addition, it does not rely on feature detection and tracking and on a separate prior calibration process.

It has also been found that the resulting composite panorama is of a much higher visual quality if a weighted scheme in combining overlapping regions is used. Specifically, in blending images, we employ a weighting distribution of an exponential form that favors pixels closer to the central column of the image to which they belong.

References

[Beardsley and Murray, 1992] P. Beardsley and D. Murray. Camera calibration using vanishing points. In D. Hogg and R. Boyle, editors, *British Machine Vision Conference*, Springer-Verlag, Sept. 1992.

[Brown, 1971] D.C. Brown. Close-range camera calibration. *Photogrammetric Engineering*, 37:855–866, 1971.

[Caprile and Torre, 1990] B. Caprile and V. Torre. Using vanishing points for camera calibration. *International Journal of Computer Vision*, 4(2):127–139, March 1990.

[Chen, 1995] S.E. Chen. QuickTime VR – An image-based approach to virtual environment navigation. *Computer Graphics (SIGGRAPH'95)*, :29–38, Aug. 1995.

[Kang and Szeliski, 1995] S. B. Kang and R. Szeliski. *3-D Scene Data Recovery using Omnidirectional Multibaseline Stereo*. Technical Report 95/6, Digital Equipment Corporation, Cambridge Research Lab, October 1995.

[Kang *et al.*, 1995] S. B. Kang, A. Johnson, and R. Szeliski. *Extraction of Concise and Realistic 3-D Models from Real Data*. Technical Report 95/7, Digital Equipment Corporation, Cambridge Research Lab, October 1995.

[McMillan and Bishop, 1995] L. McMillan and G. Bishop. Plenoptic modeling: An image-based rendering system. *Computer Graphics (SIGGRAPH'95)*, :39–46, August 1995.

[Penna, 1991] M.A. Penna. Camera calibration: A quick and easy way to determine the scale factor. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(12):1240–1245, 1991.

[Slama, 1980] Chester C. Slama, editor. *Manual of Photogrammetry*. American Society of Photogrammetry, Falls Church, Virginia, fourth edition, 1980.

[Stein, 1995] G. Stein. Accurate internal camera calibration using rotation, with analysis of sources of error. In *Fifth International Conference on Computer Vision (ICCV'95)*, pages 230–236, Cambridge, Massachusetts, June 1995.

[Szeliski, 1994] R. Szeliski. Image mosaicing for tele-reality applications. In *IEEE Workshop on Applications of Computer Vision (WACV'94)*, pages 44–53, IEEE Computer Society, Sarasota, Florida, December 1994.

[Szeliski and Kang, 1994] R. Szeliski and S. B. Kang. Recovering 3D shape and motion from image streams using nonlinear least squares. *Journal of Visual Communication and Image Representation*, 5(1):10–28, March 1994.

[Tsai, 1987] R.Y. Tsai. A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf TV cameras and lenses. *IEEE Journal of Robotics and Automation*, RA-3(4):323–344, 1987.

[Wang and Tsai, 1991] L. L. Wang and W. H. Tsai. Camera calibration by vanishing lines for 3-D computer vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-13(4):370–376, April 1991.

[Weng *et al.*, 1992] J. Weng, P. Cohen, and M. Herniou. Camera calibration with distortion models and accuracy evaluation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(10):965–980, Oct. 1992.

[Wolfram, 1991] Stephen Wolfram. *Mathematica*$^{\mathrm{TM}}$*: A System for Doing Mathematics by Computer*. Addison-Wesley, Redwood City, California, second edition, 1991.