Fast Focal Length Solution in Partial Panoramic Image Stitching

Kirk L. Duffin
Northern Illinois University
duffin@cs.niu.edu

William A. Barrett
Brigham Young University
barrett@cs.byu.edu

Abstract

Accurate estimation of effective camera focal length is crucial to the success of panoramic image stitching. Fast techniques for estimating the focal length exist, but are dependent upon a close initial approximation or the existence of a full circle panoramic image sequence. Numerical solutions of the focal length demonstrate strong coupling between the focal length and the angles used to position each component image about the common spherical center.

This paper demonstrates that parameterizing image position using distance over the sphere surface instead of angles effectively decouples the focal length from the image position. This new parameterization does not require an initial focal length estimate for quick convergence, nor does it require a full circle panorama in order to refine the focal length. Experiments with synthetic and real image sets demonstrate the robustness of the method and a speedup of 5 to 20 times over angle based positioning.

Keywords: *Focal length estimation, image stitching, partial panoramas, zoom lenses*

1 Introduction

Image stitching or *image mosaicing* is the process of transforming and compositing a set of images, each a subset of a scene, into a single larger image. The transformation for each image maps the local coordinate system present in each image onto the global coordinate system in the final composite.

There are several image transformation types reported in the literature. *Panoramic transformations*, where the images are acquired from a single view point, are most common. Panoramic mosaics can be made on cylinders, as found in QuickTime VR[3, 2] and plenoptic modeling [11]. Full panoramas can be placed on piecewise planar surfaces[7, 19]. Composition of image strips onto planar surfaces under affine transformations has also been investigated[14, 8]. Arbitrary images of planar surfaces can also be composited[10].

In the field of aerial photogrammetry, solution techniques for finding projective transformations are well developed[1]. However, correspondence with global points of known coordinates is used to give accuracy to the final composition.

Image stitching can be *incremental* or *global*. Incremental stitching works with an image pair. One of the images is considered to be fixed. The fixed image coordinate system is equivalent to the coordinate system of the final composite image. The transformation of one other image is calculated with respect to the fixed image based on their common area of overlap, typically at least 50% in the previous work mentioned. The two images are composited together and the result is used as the fixed image for stitching with the next image in the sequence. A drawback of incremental stitching is the accumulation of error in the image transformation parameters. This is often seen as ghosting of image features in the final composite.

Global stitching attempts to find the simultaneous solution of transformations for all images in the image set[16, 4]. Globally optimized stitching greatly reduces the ghosting errors in the final composite image.

A necessary step in creating panoramic composites is estimating the focal length of the camera. This can be done as an *a priori* camera calibration step or as an error correction after creating a transformation solution. Both [19] and [9] demonstrate ways of correcting the focal length estimate based on the error of matched features on opposite ends of the panorama. Of necessity, a *full* 360° panorama must be acquired and stitched in order to determine the error and the focal length correction.

1.1 High Resolution Partial Panoramas

Most of the stitching work mentioned above is used to create hemispherical panoramas using a relatively large camera field of view and small (≈ 50) number of images. This paper examines the more restrictive problem of creating high resolution partial panoramas with zoom lenses. In this problem, the camera field of view is very narrow ($< 10°$), there are a large number of images (often 100 or more) and the resulting composite fills only a small part of the hemispherical field of view.

Focal length estimates in these situations are often nonexistent. An appropriate zoom lens setting is chosen as a compromise between speed in the image acquisition and the amount of image detail desired. Because a full circle image sequence does not exist, focal length estimates cannot be directly calculated. In addition, the narrow field of view makes an estimate from overlapping image pairs very inaccurate.

The rest of this paper describes a reparameterization of the standard panoramic stitching formulas that allows for a quick solution with no initial focal length estimate.

The work is illustrated with three image sets, one of a mountain scene in the visible spectrum, an infrared filtered set from a Mayan archaeological site in Bonampak, Mexico[12][1], and a synthetic image set for the purpose of comparing calculated results with known values. For each image set, we examine the speed of convergence to a solution using both new and previously techniques.

2 Image Transformation and Solution

Creating a panoramic image from an image set is the same as finding a position on the surface of a sphere for every image in the set such that when the images are reprojected onto the sphere, the original view from the center of the sphere is recreated.

Projective matrix transformations[6] are used to transform points in the coordinate system of each image into points surrounding the sphere. Mann and Picard[10] and others have shown how arbitrary views of planar surfaces and panoramic views of a 3D scene can be described as 2D projective transformations.

Projective transforms offer a flexible set of transformations in the 2D plane. Szeliski [18] gives a good overview of the different classes of transformations that can be achieved through the use of 2D homogeneous matrices. A full projective transform offers eight degrees of freedom per image. Panoramic image transforms, as developed in Section 2.1, require only four degrees of freedom per image: three for rotation and one for focal length. It is reasonable to assume however, that the focal length is common for all images in a panoramic set.

The global solution of the parameters describing the matrix transformations is known as *bundle adjustment*[16] and is arrived at in an iterative fashion. In bundle adjustment, a set of point pairs $(\mathbf{p}_{i_k}, \mathbf{p}_{j_k})$ is identified in overlapping images i and j such that when the points are transformed to

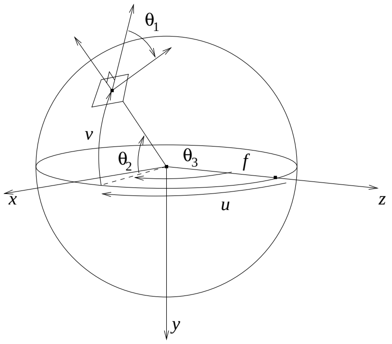

Figure 1. Panoramic image transformation. Both position angle and arc distance parameterizations shown.

their final positions, $\mathbf{p}'_{i_k}$ and $\mathbf{p}'_{j_k}$ and normalized, the distance between the points in each pair is minimized. An overall metric of the value of the solution is given by sum of squares of the point pair distances after transformation:

$$\varepsilon(\cdot) = \sum_{i,j,k} \left\| \mathrm{norm}(\mathbf{p}'_{i_k}) - \mathrm{norm}(\mathbf{p}'_{j_k}) \right\|^2 \qquad (1)$$

where i and j range over pairs of overlapping images and k ranges over a set of matched point pairs for each image pair (i, j). In this metric, the transformations are from individual image coordinate systems to the composite coordinate system.

Levenberg-Marquardt minimization [15, 13], a generalization of gradient descent and the Newton-Raphson solution of a quadratic form, is used to find the solution.

2.1 Panoramic Image Transformation

In this section we present a detailed description of the transformation from 2D image coordinates to the 3D coordinate system of the panoramic image. This description is similar to others elsewhere. The main purpose of this exposition is to provide a point of reference when we describe our reformulation of the transformation in section 3.1.

An illustration of the transformation is shown in Figure 1. The composition coordinate system is 3D, Cartesian, and right handed, with x positive to the right; y positive down, coincident with standard image pixel ordering schemes; z positive into the scene. The optic center of the image to be transformed is placed at the origin with x and y image axes parallel to those of the scene. For convenience we take the convention that the image pixel coordinates are renumbered so that the image origin is at the optic center.

The image is translated in z by the focal length f in pixels and then rotated about the origin. The rotation is almost universally parameterized as a set of three angles. A notable

[1] All infrared video images in this paper are courtesy of Mary Miller, Yale University; Stephan Houston, Brigham Young University Anthropology Department; and the Bonampak Documentation Project.

exception to this practice is [4] who use quaternions to avoid the singularities that occur when using position angles. The rotation decomposition that we use here is first a rotation θ_1 about the optic axis in the *xy* plane, followed by a rotation in the *yz* plane and lastly by a rotation in the *xz* plane.

The transformation of an image point $\mathbf{p}$ to a 3D composite coordinate system point $\mathbf{p}'$ is

$$\mathbf{p}' = \mathbf{M}\mathbf{p} = \mathbf{R}\mathbf{T}\mathbf{p} \tag{2}$$

with

$$\mathbf{R} =, \begin{pmatrix} r_{11} & r_{12} & r_{13} & 0 \\ r_{21} & r_{22} & r_{23} & 0 \\ r_{31} & r_{32} & r_{33} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \mathbf{T} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & f \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

resulting in a final transformation matrix $\mathbf{M}$ of

$$\mathbf{M} = \begin{pmatrix} r_{11} & r_{12} & r_{13} & fr_{13} \\ r_{21} & r_{22} & r_{23} & fr_{23} \\ r_{31} & r_{32} & r_{33} & fr_{33} \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{3}$$

A homogeneous initial image point $\mathbf{p}$ is always of the form $(x,y,0,1)^T$ and the transformed point $\mathbf{p}'$ of the form $(x',y',z',1)^T$. Consequently, the third column and fourth row of $\mathbf{M}$ can be eliminated, creating a 2D homogeneous transformation from $(x,y,1)^T$ to $(x',y',z')^T$.

After transformation, the points on the image plane have been transformed to points on a plane tangent to a sphere of radius f as shown in Figure 1. Matched points in different images could easily have different distances along the same ray from the center of the sphere. Consequently, the transformed points must be normalized before they can be compared.

The points can not be normalized to the surface of the sphere. because the radius of the sphere, f, is changing as part of the solution process. A drawback of matched point distance error metrics is that any transformation parameter such as f that globally reduces the magnitude of all transformed points tends to reduce their distance as well, providing a false solution. These problems can be ameliorated with modified distance error metrics. [5] presents such a metric that prevents individual image scaling parameters from converging to zero.

However, a much better solution is to normalize the transformed point pairs to lie on the unit sphere before comparison. This again is the approach taken in bundle adjustment. Because the transformation is a rigid body transformation, the magnitude of the point $(x',y',z')^T$ is the same as that of the point $(x,y,f)^T$. So the normalization can be done using untransformed points instead of transformed points which greatly simplifies the derivative calculations needed in each non-linear solution step.

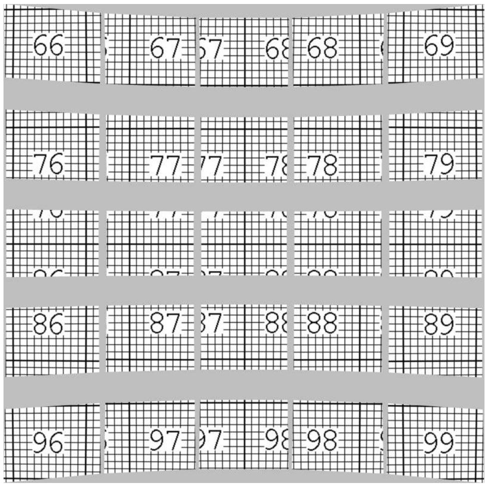

Figure 2. An illustration of the error induced by a change of focal length and constant angle positions.

The final error metric used is thus

$$\varepsilon(\theta_{\mathbf{1}},u,v,f) = \sum_{i,j,k} \left\| \frac{\mathbf{p'}_{i_k}}{\sqrt{x_{i_k}^2 + y_{i_k}^2 + f^2}} - \frac{\mathbf{p'}_{j_k}}{\sqrt{x_{j_k}^2 + y_{j_k}^2 + f^2}} \right\|^2 \tag{4}$$

where k ranges over the matched points for image pair (i,j) and the $\mathbf{p}'$ are transformed as in Equation 2.

3 A New Parameterization

One severe problem with the bundle adjustment as presented is that it converges very slowly. This is due to the strong coupling between the focal length and the position angles. This coupling is demonstrated in Figure 2 where the angle parameters for a correct stitching solution of a synthetic grid are left intact and only the focal length is changed from its correct value. This strong coupling constrains changes in focal length to be small because in focal length drastically increase the final error measurements.

3.1 Arc Distance Parameterization

Our solution, and the key point to this paper is to decouple the position parameters from the focal length by using arc distance along the sphere surface instead of angles. These distances, labeled as u and v and measured in pixels, are used as parameters for image position on the sphere. The parameter v is equivalent to distance along a longitude line from the equator while u is the distance from the longitude line, along a parallel. The new transformation parameters are also illustrated in Figure 1.

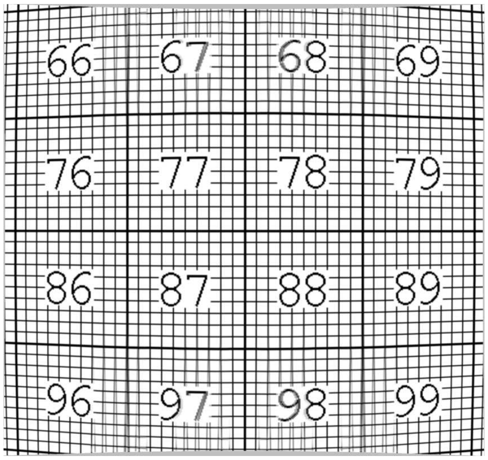

Figure 3. An illustration of the error induced by a change of focal length and constant arc distance positions.

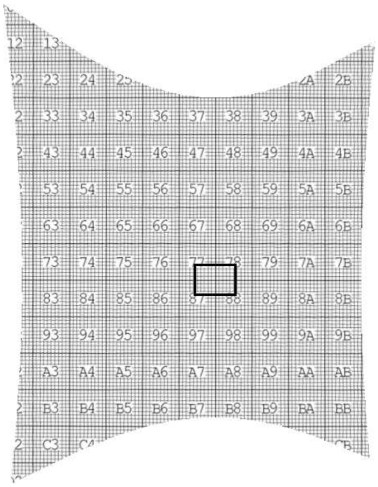

Figure 4. A partial panoramic stitch of a synthetic image set. The component image with a fixed identity transformation is outlined in black.

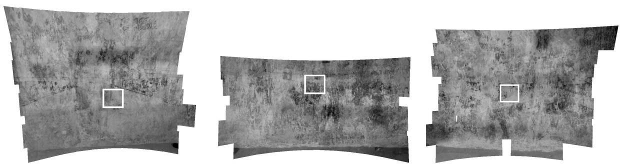

Figure 5. The Bonampak partial panoramic stitches. From left to right, Bonampak 1, 2, and 3. In each composite, the image with a fixed identity transformation is outlined in white.

Only the rotation matrix **R** in Equation 2 is changed by the u and v parameters. Angle θ_2 is replaced by v/f while θ_3 is replaced by u/f.

Using an arc distance parameterization, the relative distances between images remain comparatively unaffected by changes in focal length. A helpful analogy is to envision a flexible sheet of images wrapped around the sphere that readjusts as the sphere changes radius. Figure 3 demonstrates that the new parameterization is uncoupled by altering the focal length of a correct stitching solution of the same synthetic grid in Figure 2. The arc distances are left constant. The change made to the focal length is the same in both parameterizations.

4 Application and Comparison

In this section we compare the arc distance parameterization with the standard angle-based bundle adjustment method. We compute panoramic transformations for several image sets using both parameterizations and examine the convergence of the focal length parameter. All panoramic transformations in this section were computed by Levenberg-Marquardt minimization with an extremely conservative stopping criterion — no change in the parameter vector to within 10^{-9}.

In each image set, point pairs are chosen from overlapping image pairs. In the synthetic image set to be shown, salient feature point pairs are chosen automatically. In the real world image sets, matched point pairs are chosen by hand. In all cases, point coordinates are refined to subpixel precision using intensity based matching in a small region about each pair point. The region average is subtracted out during the matching to help compensate for large scale, spatially varying bias in the sensor.

Figure 4 shows a panoramic composite of the synthetic grid image created specifically to test focal length accuracy. This image set has a 10° field of view with a stepping angle of 8° between images. With images of size 640 by 480 pixels, the true focal length is 2743.213 pixels.

Figure 5 shows three infrared composites of contiguous sections of a Mayan mural in Bonampak, Mexico. The images contain complex, low contrast, background texture. The images were captured with a video camera with a zoom lens and an IR filter. The heavy filter pushed the image sensor close to its threshold of operation, resulting in noisy images with accentuated spatially dependent bias. Our approach of hand picking matched point pairs was designed in direct response to these image sets. During image acquisition, at each imaging position, the zoom was maximized to focus on the wall and then reduced slightly to fit more content into each frame. Consequently, the true focal length is unknown and varies with each set; within each set, the f is assumed to remain constant.

Features in these composites are difficult to appreciate at the scale reproduced here. However, the intersection of the mural wall with the floor can be seen as a straight line at the bottom of each image composite. These straight lines are correctly reconstructed artifacts of the image data. The composition does not use this image edge in any way.

Figure 6 shows a video composite of a mountain peak. High zoom magnification was used to acquire these images, resulting in a very narrow field of view of $\approx 5°$. The true focal length is again unknown. The full resolution size of this image is 16126 by 3210 pixels.

For each image set, an initial solution is computed, allowing only translation. The same initial solution is used for both the spherical and projective methods. Both methods start out with an initial focal length estimate of 100,000 pixels in all cases. Table 1 summarizes the results of the experiments.

Figure 7 shows the convergence of the focal length estimation in the synthetic Grid image. Both angle and arc distance methods arrive at the same focal length estimate, but the arc distance method converges over 7.5 times faster. The spherical method allows the focal length estimation more freedom to change, leading to oscillations in the estimate. But the same freedom lets the estimate settle down to within .1 pixel of the final value after only 30 iterations. Residual oscillations dampen out until no change occurs within 10^{-9}.

The final focal length estimate in this image set is 2747.548 pixels. The actual focal length is 2743.213 pixels. Two points need to be stated regarding this relative error of 0.16% in the focal length, which is considerably less than errors achieved with real image sets. First, the eyepoint and center of rotation are coincident. Stein [17] has shown the estimation error that results when the two points are not coincident. Secondly, the estimation error in this case can be traced to inaccuracies in the refinement of point pair coordinates using the best match of small image regions about the points. When exact *a priori* priori coordinates are used, the focal length estimate matches the exact value to within 3.1×10^{-4}. And sum squared error drops to within 0.002448.

Figure 10 shows the sum squared error for the solution of the Grid image set.

Figures 8 and 11 show the progression of focal length estimates and total SSQ error for the three Bonampak image sets of Figure 5. The focal length estimate for the arc distance parameterization converges 12 to 16 times faster to its final value than the angular parameterization.

Figures 9 and 12 show the focal length estimates and total SSQ error for the Mountain image set of Figure 6. In this example, The arc distance based estimate converges over 20 times faster than the solution based on angle parameterization.

5 Conclusion

In this paper, we have presented a reparameterization of the partial panoramic stitching problem based on arc distance. We have shown how the new formulation results in robust estimates of system focal length without the need for approximate initial estimates. We have also demonstrated a significant increase (roughly an order of magnitude) in the rate of convergence of focal length estimates over standard angle based parameterizations.

Quick, robust convergence of focal length estimates extends image stitching techniques to the use of zoom lenses, where focal lengths are unknown.

Initial work implementing the ideas in this paper showed that arc distance parameterization alone is responsible for the freedom of movement exhibited by the focal length parameter.

Future work will include applying the spherical distance parameterization to intensity based error metrics, determining whether or not such a change will reduce the need for *a priori* focal length estimates for this important class of metrics.

6 Acknowledgments

This work was funded by a grant from the Utah State Centers of Excellence, the Computer Science Department at Brigham Young University and the Center for Research In Vision and Imaging Technologies (RIVIT) at BYU. Infrared video images of Bonampak were provided courtesy Stephan Houston, BYU Anthropology Department; Mary Miller, Yale University; and the Bonampak Documentation Project.

References

[1] C. Burnside. *Mapping from Aerial Photographs*. Collins, 2nd edition, 1985.

[2] S. E. Chen. QuickTime VR — An Image-Based Approach to Virtual Environment Navigation. In *Computer Graphics Proceedings, Annual Conference Series*, pages 29–38. ACM SIGGRAPH, ACM Press, August 1995.

[3] S. E. Chen and L. Williams. View Interpolation for Image Synthesis. In *Computer Graphics Proceedings, Annual Conference Series*, pages 279–288. ACM SIGGRAPH, ACM Press, August 1993.

[4] S. Coorg and S. Teller. Spherical Mosaics with Quaternions and Dense Correlation. *International Journal of Computer Vision*, 37(3):259–273, June 2000.

Image Set	Image	Point	Trans.		Final	Final	Pan.
Images	Pairs	Pairs	Steps		f (Pixels)	SSQ Error	Steps
Grid				angle	2747.548	9913.592	384
100	180	10235	22	arc	2747.548	9913.592	51
Bonampak 1				angle	3378.902	22657.818	598
91	163	1507	16	arc	3378.902	22657.818	50
Bonampak 2				angle	3427.450	2265.846	828
65	114	759	17	arc	3427.450	2265.846	53
Bonampak 3				angle	3866.855	18138.216	658
89	171	1040	16	arc	3866.855	18138.216	41
Mountain				angle	4993.679	42452.350	1112
177	364	4401	16	arc	4933.679	42452.350	53

Table 1. A comparison of panoramic stitching over several image sets. The number of iterative steps to obtain an initial translation-only solution are given. The number of additional steps to obtain a panoramic solution is also given for both the planar projective error metric and the spherical surface 3D error metric.

[5] K. L. Duffin and W. A. Barrett. Globally Optimal Image Mosaics. In *Proceedings, Graphic Interface '98*, pages 217–222. Canadian Human-Computer Communications Society, June 1998.

[6] L. E. Garner. *An Outline of Projective Geometry*. North Holland, 1981.

[7] M. Irani, P. Anandan, and S. Hsu. Mosaic Based Representations of Video Sequences and Their Applications. In *International Conferance on Computer Vision*, pages 605–611, 1995.

[8] B. Jones. Texture Maps from Orthographic Video. In *Visual Proceedings, Annual Conference Series*, page 161. ACM SIGGRAPH, ACM Press, August 1997.

[9] S. B. Kang and R. Weiss. Characterization of Errors in Compositing Panoramic Images. Technical Report 96/2, Digital Equipment Corporation, Cambridge Research Lab, June 1996.

[10] S. Mann and R. Picard. Virtual Bellows: Constructing High Quality Stills from Video. In *International Conference on Image Processing*, pages 363–367, 1994.

[11] L. McMillan and G. Bishop. Plenoptic Modeling: An Image-Based Rendering System. In *Computer Graphics Proceedings, Annual Conference Series*, pages 39–46. ACM SIGGRAPH, ACM Press, August 1995.

[12] M. Miller. Maya Masterpiece Revealed at Bonampak. *National Geographic*, 187(2):50–69, February 1995.

[13] J. C. Nash. *Compact Numerical Methods for Computers*. Adam Hilger, 1990.

[14] S. Peleg and J. Herman. Panoramic Mosaics by Manifold Projection. In *IEEE Computer Vision and Pattern Recognition*, pages 338–343, 1997.

[15] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C*. Cambridge University Press, 2nd edition, 1992.

[16] H. Shum and R. Szeliski. Construction and Refinement of Panoramic Mosaics with Global and Local Alignment. In *International Conference on Computer Vision*, pages 953–958, 1998.

[17] G. P. Stein. Accurate Internal Camera Calibration using Rotation with Analysis of Sources of Error. In *International Conference on Computer Vision*, pages 230–236, 1995.

[18] R. Szeliski. Video Mosaics for Virtual Environments. *IEEE Computer Graphics and Applications*, pages 22–30, March 1996.

[19] R. Szeliski and H.-Y. Shum. Creating Full View Panoramic Image Mosaics and Environment Maps. In *Computer Graphics Proceedings, Annual Conference Series*, pages 251–258. ACM SIGGRAPH, ACM Press, August 1997.

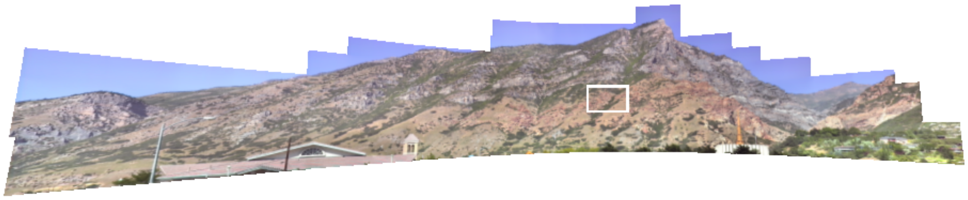

Figure 6. The Mountain partial panoramic stitch. The image with a fixed identity transformation is outlined in white. The slant of vertical features in the image is due to the non-zero angle between the world up vector and the fixed image plane.

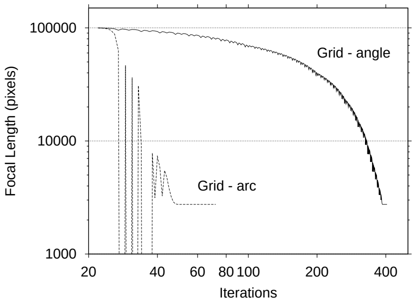

Figure 7. Focal length estimation in the Grid image set.

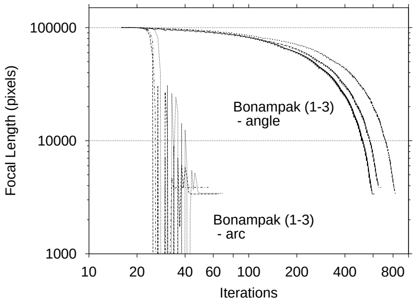

Figure 8. Focal length estimation in the Bonampak image sets.

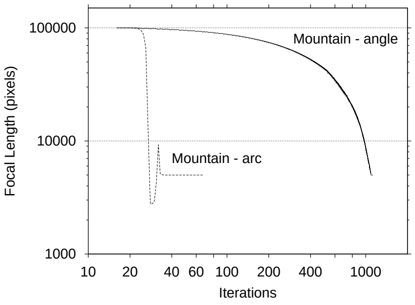

Figure 9. Focal length estimation in the Mountain image set.

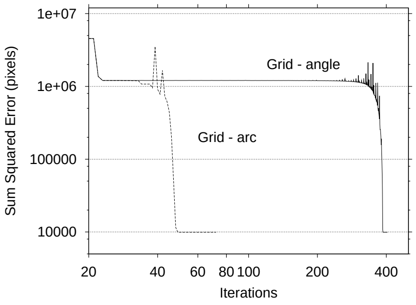

Figure 10. Sum squared error in the Grid image set.

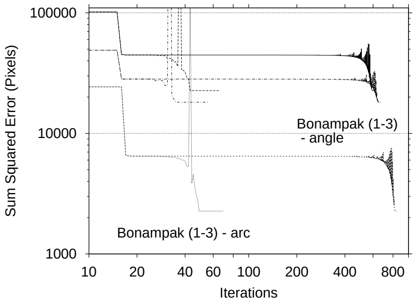

Figure 11. Sum squared error in the Bonampak image sets.

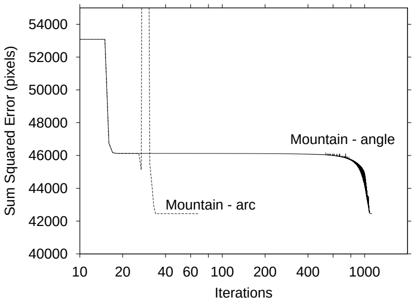

Figure 12. Sum squared error in the Mountain image set.