Globally Optimal Image Mosaics*

Kirk L. Duffin William A. Barrett

Computer Science Department
Brigham Young University
Provo, UT 84602
e-mail: {kirkℓ,barrett}@cs.byu.edu

Abstract

In this paper we examine the simultaneous solution of a set of image transformations with the goal of creating a globally optimal image mosaic. We examine an alternative parameterization of the full projective matrix transformation that leads to elimination of independent skew and aspect ratio parameters for each image. We also create a scale-free distance error metric which prevents the tendency of simultaneously solved systems to tend toward the zero solution.

Keywords: Image stitching, image mosaics, projective transformations

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 used. Panoramic transformations onto cylinders are most common, as found in QuickTime VR[3, 2] and plenoptic modeling[8]. Mosaics on planar surfaces are also popular. Panoramas with a stationary eyepoint can be placed on piecewise planar surfaces[5, 13]. Arbitrary images of planar surfaces can also be composited[7]. Composition of image strips onto planar surfaces under affine transformations has also been investigated[11, 6].

The previous work mentioned has focused on composition of images with large image overlap, typically at least 50%. The distance, in terms of number of images, between any subimage and a subimage whose transformation is fixed, is often very small. Often this distance is one, i.e., each new image is combined with the previous composite image which is held fixed.

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

The focus of this paper is to examine the simultaneous solution of the full set of component image transformations needed for a composite image. Many component images are distant from a fixed image. Unlike cylindrical panoramas, we will examine mosaics extending both horizontally and vertically from the fixed image. We will use photogrammetric techniques for simultaneous solution of the image transformations. However, like the previous stitching work mentioned, we will assume no known global control points. We will also demonstrate how a new error measurement function and a reparameterization of the standard projective transformation produce composites with less warping than traditional formulas.

We will illustrate our work with two image sets, one of a mountain scene in the visible spectrum, and an infrared[1] filtered set from a Mayan archaeological site in Bonampak, Mexico[9]. The first set is a typical video mosaicing set with overall good contrast, clean signal, and generous overlap. It contains 79 images. The Bonampak set contains 91 frames, many with low contrast, heavy filtering, uneven sensor response, and abundant noisy background texture. This set has proved to be an excellent test set for our methods. Examples of both image sets can be seen in Figure 1. The relative positions and overlaps of the individual images can be seen in Figures 4d and 5d.

Standard Projective Transform Solution

Solution of an image transformation involves two components:

- An error function which measures the difference between features in overlapping image pairs. This error metric is usually the sum squared error for individual image pair features.

*This paper presented at 1998 Graphics Interface.

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

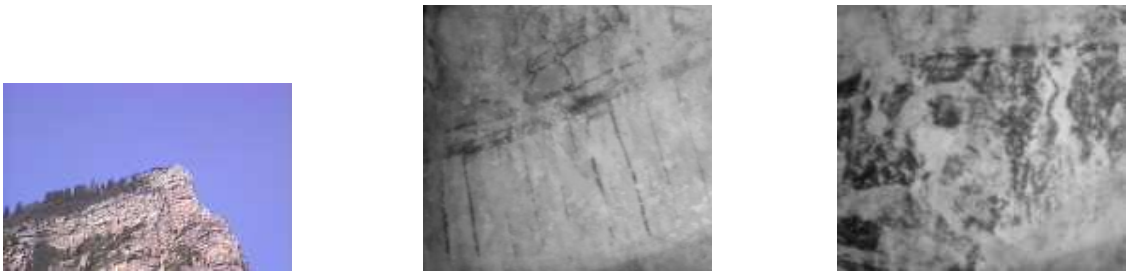


Figure 1: Three video frames from two image stitching sets. The left image is the summit of Squaw Peak in Provo, Utah. The center and right images are two infrared video frames of a Mayan archaeological mural in Bonampak, Mexico. This center image is typical and a good example of a composition set of low contrast, uneven sensor response, and ambiguity from irregular background texture. The right image shows a more interesting, but more infrequent frame from the same data set. Note the light-colored profile of a Mayan figure in the lower left area of the frame.

- A set transformation functions between the coordinate systems of the component subimages and the coordinate system of the composite image.

A non-linear least squares minimization technique such as Levenberg-Marquardt[12, 10] is used to compute the transformation function parameters for each image.

In order to compute a unique solution, it is necessary to fix the transformation of at least one image with respect to the global coordinate system. This is usually done by giving one image the identity transformation.

Error Functions

A commonly used error function minimizes the difference in image intensity between transformed image pairs:

$$\epsilon() = \sum_{i,j,\mathbf{x}} [I_i(f_i(\mathbf{x})) - I_j(f_j(\mathbf{x}))]^2 \qquad (1)$$

where I_i and I_j are a pair of overlapping images and f_i and f_j the corresponding transformations from global coordinates $\mathbf{x}$ to local image coordinates.

Error functions such as equation 1 are the function of choice for interactive video mosaics and wherever no explicit point matching is performed between image frames. Such functions work well when contrast is high, sensor response is even, noise is low, and ambiguity is minimal. Image sequences that respond poorly to automatic matching techniques require explicit point matching and a different error function.

An error function that minimizes the global distance between matched points in image pairs is

$$\epsilon() = \sum_{\substack{i,j,n \\ (\mathbf{x}_{i_n},\mathbf{x}_{j_n})\in P_{ij}}} [g_i(\mathbf{x}_{i_n}) - g_j(\mathbf{x}_{j_n})]^2 \qquad (2)$$

where g_i and g_j are the image transformations from local to global image coordinates and P_{ij} is the set of matched point pairs for images i and j.

Matrix Transformations

In the plane, all projective transformations can be expressed as 3 by 3 homogeneous matrix transforms[4]. These transforms are equivalent up to a non-zero scale factor, leaving eight independent variables in the transformation:

$$\begin{pmatrix} m_1 & m_2 & m_3 \\ m_4 & m_5 & m_6 \\ m_7 & m_8 & 1 \end{pmatrix} \qquad (3)$$

The transformed coordinates of a point (x, y) using this transformation are

$$x' = \frac{m_1 x + m_2 y + m_3}{m_7 x + m_8 y + 1}, \qquad y' = \frac{m_4 x + m_5 y + m_6}{m_7 x + m_8 y + 1} \qquad (4)$$

Examples of global mosaics computed for the two test data sets using the error metric of equation 2 and the transformation parameterization of equation 4 can be found in Figures 4a and 5a.

Modified Projective Transform Solution

Using the direct matrix entry parameterization of equation 3 reveals two problems. First, there are sufficient degrees of freedom to allow an independent image skew angle and independent x and y scale factors for each image. Cameras used in image stitching application can be considered to have a constant, and often negligible, skew angle for each image. A zoom lens may provide an isotropic scale factor, but the aspect ratio can be assumed to be fixed.

Secondly, a valid globally optimal solution is the zero solution, i.e when all free matrix entries go to 0. In practice, the existence of the zero solution is seen as a tendency for images to shrink with distance from the fixed

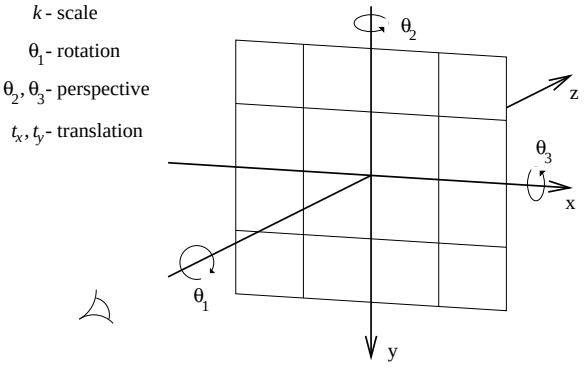


Figure 2: A reduced set of projective transformation parameters. Individual skew and aspect ratio are ignored for each image. The perspective parameters are reformulated as rotations in the z direction.

image. This phenomenon can be seen in Figures 4a and 5a.

Reparameterized Projective Transforms

An equivalent representation of the full projective transform with a different parameterization uses the composition of the following matrices with their accompanying parameters: image rotation $\mathbf{R}$,θ; isotropic scale factor $\mathbf{S}$,k; skew angle $\mathbf{L}$,θ_s; anisotropic scale factor or aspect ratio $\mathbf{A}$,a; translation $\mathbf{T}$, t_x and t_y; and perspective or keystone factors $\mathbf{P}$, p_x and p_y. A common transformation composition would be

$$\mathbf{x}' = \mathbf{TPRALSx} \tag{5}$$

Note that this parameterization has the same number of free parameters, 8, as the direct parameterization.

An alternative view to the perspective factors treats them as rotations in the xz and yz planes of a camera centered coordinate system as shown if Figure 2. These rotations, θ_2 and θ_3, can be combined with the rotation in the xy plane, θ_1, into a single 3 by 3 rotation matrix. Casting out the skew and aspect ratio parameters the new parameterization for the projective transformation matrix becomes the matrix shown in Figure 3. The new parameterization has 6 parameters instead of 8, which reduces the size of the Jacobian matrices used in the Levenberg-Marquardt routine and helps compensate for the complexity of the new formulas. In addition, our experiments have shown that without the extra degrees of freedom, the transformation solver converges to its solution in fewer steps.

Using this reduced parameterization for projective transforms results in the composite images shown in Figures 4b and 5b.

Scale-free Error Function

Even with the new parameterization, the undesirable zero solution still exists. As the scale of any image goes to zero, so does the overall error. A new inverse scaled distance function ameliorates this problem. The new error function for each point pair in an image pair is calculated as the difference in transformed point coordinates divided by the geometric average of the scale factors for each image transformation, i.e.,

$$\epsilon() = \sum_{\substack{i,j,n \\ (\mathbf{x}_{i_n},\mathbf{x}_{j_n})\in P_{ij}}} \left[\frac{1}{\sqrt{k_i k_j}}\left(g_i(\mathbf{x}_{i_n}) - g_j(\mathbf{x}_{j_n})\right)\right]^2 \tag{6}$$

This formula measures the relative distance between points in an image pair. The overall error is unaffected by scaling both images equally. As the scale factor approaches 0, the relative distance error increases. In an iterative solver, changes to scale will be made to reduce the overall distance error for a set of points rather than the distance between any given point pair.

Composite images using transformations calculated using the inverse scaled error function are shown in Figures 4c and 5c. Note the improvement in the wall-floor intersection line in the Bonampak image. Also note the reduced keystone effects in the Squaw Peak image and the flattening of the bottom edge of the composite. This is proper behavior for this image set, as the fixed image with identity transformation for this set is located on the bottom row below the summit.

Conclusion

In this paper we have examined the simultaneous solution of sets of image transformations under perspective projection for the purpose of image composition. We have reformulated the standard projective matrix transformation in order to reduce degrees of freedom that do not exist in the initial image set. Using the new transformation parameters, we have developed a new error metric that keeps the global solution from sliding towards the global zero solution. The combination of new error metric and projective parameterization produces projective transformation composites without as much warping as the full projective matrix solution. In addition, because the solutions are globally optimal, transformation error is distributed throughout the composite image and is thus less noticeable.

Acknowledgments

This work was funded by the Computer Science Department at Brigham Young University and the Center for Research In Vision and Imaging Technologies (RIVIT) at

$$\mathbf{x}' = \mathbf{TPRSx} = \begin{pmatrix} k\cos\theta_1\cos\theta_2 & -k\sin\theta_1\cos\theta_2 & t_x \\ \begin{pmatrix} -k\cos\theta_1\sin\theta_2\sin\theta_3 \\ +k\sin\theta_1\cos\theta_3 \end{pmatrix} & \begin{pmatrix} k\sin\theta_1\sin\theta_2\sin\theta_3 \\ +k\cos\theta_1\cos\theta_3 \end{pmatrix} & t_y \\ \begin{pmatrix} k\cos\theta_1\sin\theta_2\cos\theta_3 \\ +k\sin\theta_1\sin\theta_3 \end{pmatrix} & \begin{pmatrix} -k\sin\theta_1\sin\theta_2\cos\theta_3 \\ +k\cos\theta_1\sin\theta_3 \end{pmatrix} & 1 \end{pmatrix} \mathbf{x}$$

Figure 3: The matrix formulation of a reduced set of projective transformation parameters. See the text for a description of the parameters.

BYU. Infrared video images of Bonampak were provided courtesy of Stephan Houston, BYU Anthropology Department; Mary Miller, Yale University; and the Bonampak Documentation Project.

References

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

[2] Shenchang Eric 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] Shenchang Eric Chen and Lance Williams. View interpolation for image synthesis. In *Computer Graphics Proceedings, Annual Conference Series*, pages 279–288. ACM SIGGRAPH, ACM Press, August 1993.

[4] Lynn E. Garner. *An Outline of Projective Geometry*. North Holland, 1981.

[5] 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.

[6] Brian Jones. Texture maps from orthographic video. In *Visual Proceedings, Annual Conference Series*, page 161. ACM SIGGRAPH, ACM Press, August 1997.

[7] S. Mann and R.W. Picard. Virtual bellows: Constructing high quality stills from video. In *International Conference on Image Processing*, pages 363–367, 1994.

[8] Leonard McMillan and Gary Bishop. Plenoptic modeling: An image-based rendering system. In *Computer Graphics Proceedings, Annual Conference Series*, pages 39–46. ACM SIGGRAPH, ACM Press, August 1995.

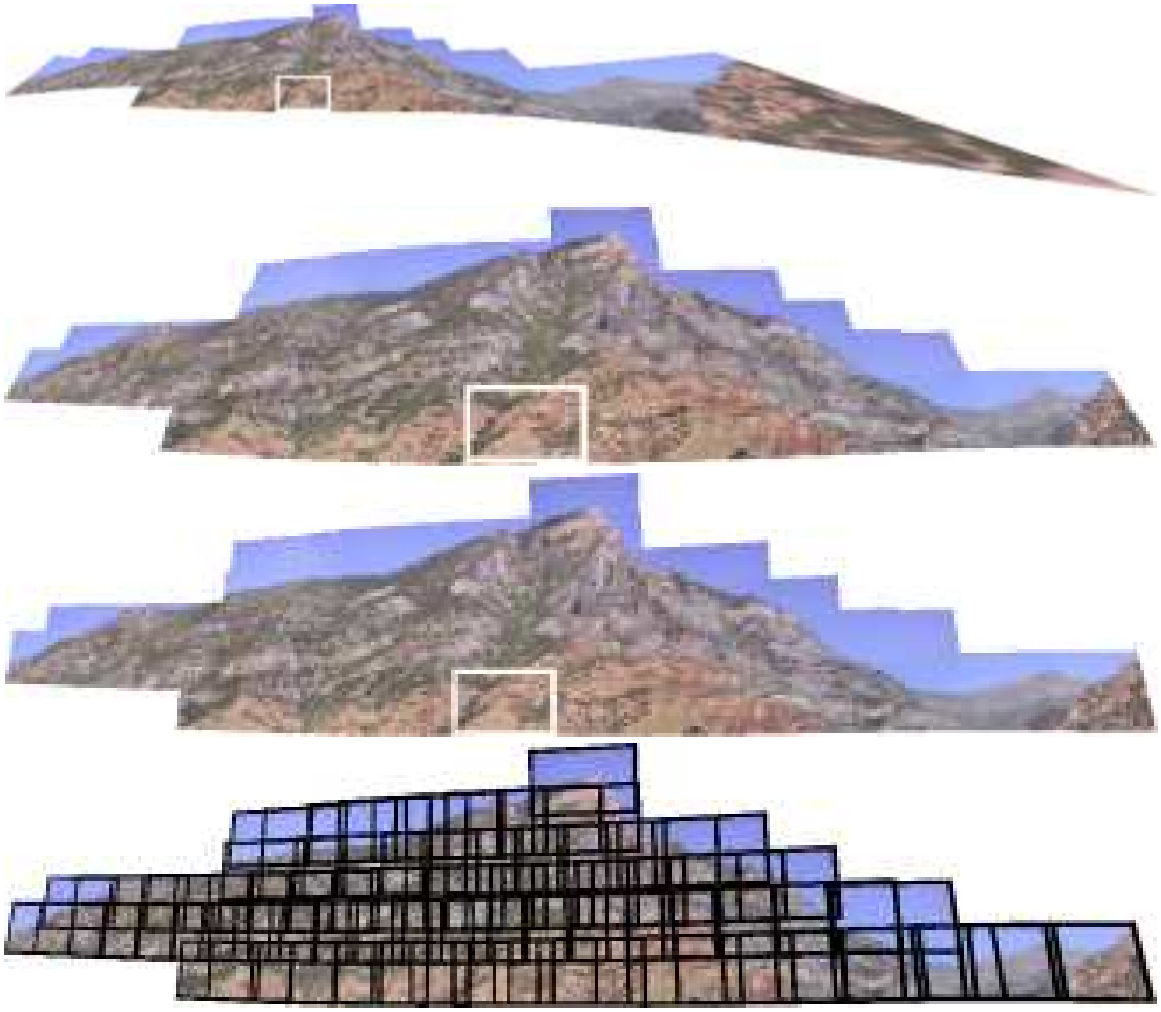
[9] Mary Miller. Maya masterpiece revealed at bonampak. *National Geographic*, 187(2):50–69, February 1995.

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

[11] S. Peleg and J. Herman. Panoramic mosaics by manifold projection. In *IEEE Computer Vision and Pattern Recognition*, pages 338–343, 1997.

[12] William H. Press, Saul A. Teukolsky, Vetterling William T., and Flannery Brian P. *Numerical Recipes in C*. Cambridge University Press, 2nd edition, 1992.

[13] Richard Szeliski and Heung-Yeung 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.



(a

(b

(c

(d

Figure 4: Globally optimal Squaw Peak image composites. Image a was computed using a simple distance error metric and matrix entry transformation parameterization. Image b was computed with a simple distance error metric and constrained projective transformation parameterization. Image c was computed with an inverse scale error metric and constrained projective transformation parameterization. The fixed image in each composite is outlined in white. Each composite contains 79 images, which are outlined in image d.

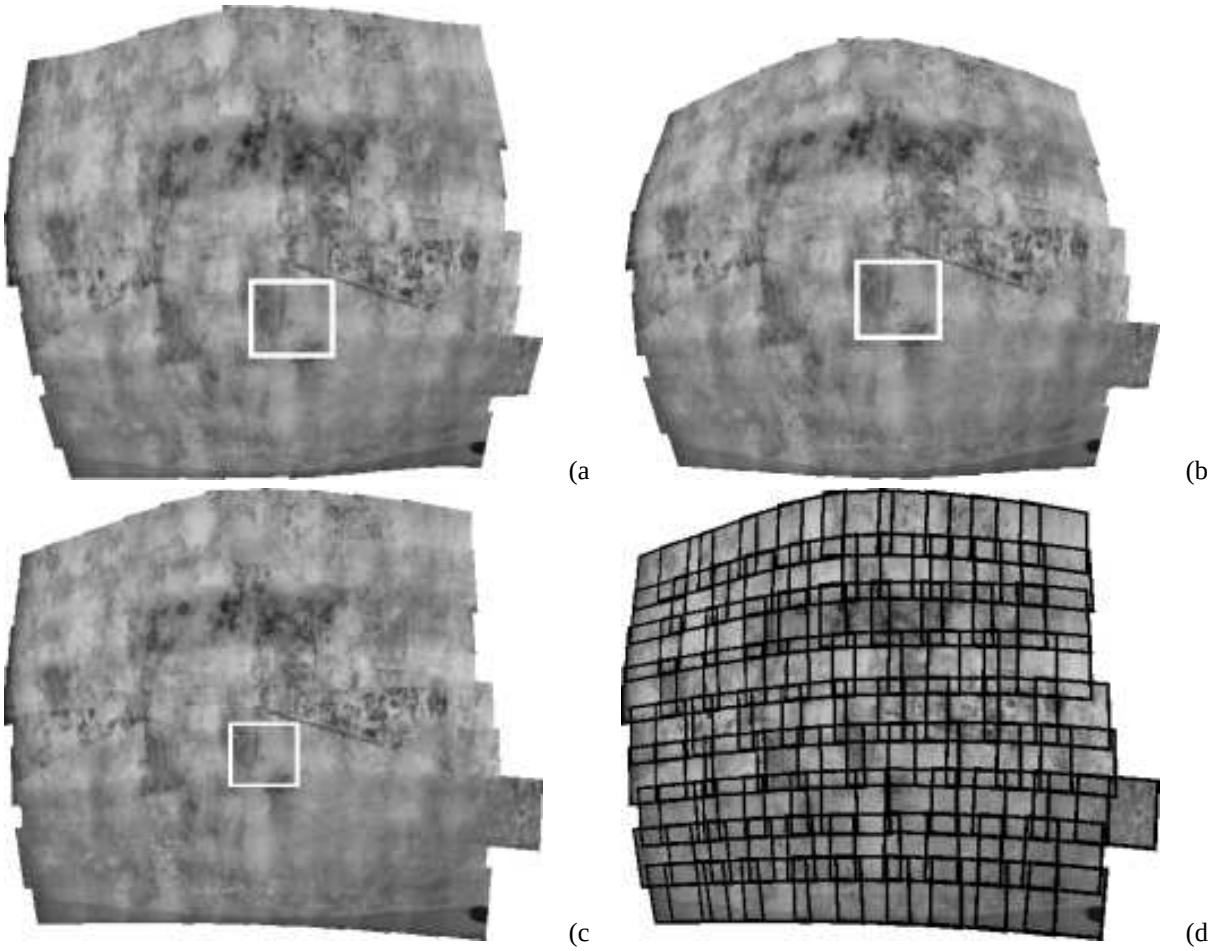


Figure 5: Globally optimal Bonampak image composites. Image a was computed using a simple distance error metric and matrix entry transformation parameterization. Image b was computed with a simple distance error metric and constrained projective transformation parameterization. Image c was computed with an inverse scale error metric and constrained projective transformation parameterization. The fixed image in each composite is outlined in white. Each composite contains 91 images, which are outlined in image d. Note particularly the line in the images where the wall meets the floor. This line should be straight. (The dark spot in the lower right hand corner is an ancient Mayan camera lens cap.)