IMAGE REGISTRATION FOR PERSPECTIVE DEFORMATION RECOVERY *

George Wolberg
Siavash Zokai

Department of Computer Science
City College of New York
New York, NY 10031
{wolberg|zokai}@cs-mail.engr.ccny.cuny.edu

Keywords: affine transformation, nonlinear least-squares, mosaics, perspective transformation, registration

ABSTRACT

This paper describes a hierarchical image registration algorithm to infer the perspective transformation that best matches a pair of images. This work estimates the perspective parameters by approximating the transformation to be piecewise affine. We demonstrate the process by subdividing a reference image into tiles and applying affine registration to match them in the target image. The affine parameters are computed iteratively in a coarse-to-fine hierarchical framework using a variation of the Levenberg-Marquadt nonlinear least squares optimization method. This approach yields a robust solution that precisely registers image tiles with subpixel accuracy. The corresponding image tiles are used to estimate a global perspective transformation. We demonstrate this approach on pairs of digital images subjected to large perspective deformation.

1. INTRODUCTION

Image registration refers to the geometric alignment of a set of images. The set may consist of two or more digital images taken of a single scene at different times, from different sensors, or from different viewpoints. The goal of registration is to establish geometric correspondence between the images so that they may be transformed, compared, and analyzed in a common reference frame. This is of practical importance in many fields, including remote sensing, medical imaging, and computer vision. Registration is often necessary for (1) integrating information taken from different sensors (i.e., multisensor data fusion), (2) finding changes in images taken at different times or under different conditions, (3) inferring three-dimensional information from images in which either the camera or the objects in the scene have moved, and (4) for model-based object recognition.[1] Common tasks associated with image registration include generating large panoramic images (image mosaics) from several overlapping images, producing super-resolution images from multiple images of the same scene, change detection, topographic mapping, and multisensor image fusion.

This paper describes a hierarchical image registration system based on parameter estimation techniques. We consider pairs of digital images that were acquired under a weak perspective camera model or a fixed center of projection. Both constraints yield images that are free of parallax. This is necessary to facilitate the registration of an image pair by a global perspective transformation. There are three classes of images that we can register with this method:

1. images of any 3-D scene taken at large distances, e.g., high altitude surveillance imagery.
2. images of any 3-D scene taken at any distance but from a fixed center of projection, e.g., any set of images acquired from a (swiveling) camera on a tripod.
3. images of flat objects taken at any distance.

Perspective transformations introduce space-variant scale changes to an image. This makes the solution to general perspective registration very difficult. We circumvent this problem by approximating perspective transformations to be piecewise affine. That is, a reference image is subdivided into tiles and affine registration is applied to find the best matching tiles in the target image. These tile pairs are used to estimate a global perspective transformation. This approach permits us to exploit recent results in affine registration to solve the more difficult perspective registration problem.

* Appears in *Proc. SPIE, Automatic Target Recognition X, Orlando, FL, April 2000.*
This work was supported by a NASA FAR Award (NAG-57129).

There is a vast literature of work in the related fields of image registration, motion estimation, image mosaics, and video indexing. Most algorithms exploit a hierarchical approach due to computational efficiency in handling large displacements. Hierarchical motion estimation[2–4] and image mosaic[5–12] algorithms usually assume small deformations among image pairs. For instance, a dense image sequence is required to stitch the frames together.[8,11] The problem of assembling a large set of images into a common reference frame is simplified when the inter-frame deformations are small. Our work addresses the problem of registering two input images in the presence of large deformations. We make use of hierarchical affine registration[13] to infer perspective. Although existing hierarchical affine registration work attempts to find local deformations,[14] we infer perspective transformations since they are a predominant deformation model found in a large class of images.

In Section 2, we review deformation models. Section 3 describes a nonlinear least-squares technique based on the Levenberg-Marquadt algorithm to solve the affine registration problem. Section 4 discusses how to assemble the results of the affine registration stage to infer the perspective parameters. Finally, Section 5 presents a summary and conclusions.

2. DEFORMATION MODELS

A deformation model is necessary to define the class of spatial transformations that may apply between any two images, I_1 and I_2, of the same scene. This defines a geometric relationship between each point in both images. The general form for the mapping function induced by the deformation is

$$[x, y] = [X(u, v), Y(u, v)] \tag{1}$$

where $[u, v]$ and $[x, y]$ denote corresponding pixels in I_1 and I_2, respectively, and X and Y, are arbitrary mapping functions that uniquely specify the spatial transformation. In registering I_1 and I_2, we shall be interested in recovering the inverse mapping functions U and V that transform I_2 back into I_1:

$$[u, v] = [U(x, y), V(x, y)] \tag{2}$$

In this sense, the registration process is akin to geometric correction, whereby we attempt invert the distortion.

The choice of deformation model must be closely tied to the imaging process by which the input data set was acquired. The complexity of the model dictates the choice of registration algorithm. Images which are known to differ by small translations, for example, require only a simple correlation technique to perform registration. Images which differ by elastic deformations, though, may require user-supplied correspondence points. These points, also known as control, fiducial, and tie points, are marked by the user at sparse and irregular landmark sites on the images. Since the mapping function is precisely known at only those points, a scattered data interpolation algorithm must be applied to smoothly propagate the mapping to all other points in the image. Note that in the former case the mapping function was derived directly from the image intensities, while in the latter case the mapping function was constructed from sparse correspondence points.

Geometric correspondence is achieved by determining the mapping function that governs the relationship of all points among a pair of images. There are several common mapping function models in image registration. They include (1) 3-parameter rigid transformations (translation, rotation), (2) 6-parameter affine transformations (translation, rotation, scale, shear), (3) 8-parameter perspective transformations, and (4) local nonrigid transformations.

In this paper, we shall address the problem of registering images misaligned due to a perspective transformation. The mapping function is given as

$$u = \frac{a_1 x + a_2 y + a_3}{a_7 x + a_8 y + 1} \tag{3a}$$

$$v = \frac{a_4 x + a_5 y + a_6}{a_7 x + a_8 y + 1} \tag{3b}$$

The eight parameters that govern the transformation of observed image I_2 into reference image I_1 will be estimated by approximating perspective to be piecewise affine. This will permit us to use a robust parameter estimation algorithm to recover the six unknown parameters relating each pair of image tiles in I_1 and I_2.

An affine transformation considers only the numerators in Eq. (3), i.e., $a_7 = a_8 = 0$. The affine parameters are estimated by minimizing the sum of squared differences between the registered tiles. A least squares method is applied to the collection of affine parameters recovered over all tile pairs in order to infer the eight parameters of a global perspective transformation.

3. AFFINE PARAMETER ESTIMATION

In this section, we demonstrate how to determine the affine parameters using nonlinear least-squares optimization. We use the sum of squared differences (SSD) as the objective criterion that establishes a similarity measure between two images (or regions):

$$\begin{aligned}\chi^2(\mathbf{a}) &= \iint_{\mathbf{u}\subset R^2} \left(I_1(\mathbf{u}) - I_2'(\mathbf{u})\right)^2 d\mathbf{u} \\ &= \iint_{\mathbf{u}\subset R^2} \left(I_1(\mathbf{u}) - T_{\mathbf{A}}\{I_2(\mathbf{x})\}\right)^2 d\mathbf{u} \\ &= \|I_1(\mathbf{u}) - T_{\mathbf{A}}\{I_2(\mathbf{x})\}\|^2 \end{aligned} \tag{4}$$

where T is a geometric transformation applied to observed image I_2 to map it from its $[x, y]$ coordinate system to the $[u, v]$ coordinate system of reference image I_2. The subscript $\mathbf{A}$ denotes that this transformation is affine. For discrete data, we have

$$\begin{aligned}\chi^2(\mathbf{a}) &= \sum_{i=1}^{N} \left[I_1(\mathbf{u}_i) - I_2'(\mathbf{u}_i)\right]^2 \\ &= \sum_{i=1}^{N} \left[I_1(\mathbf{u}_i) - T_{\mathbf{A}}\{I_2(\mathbf{x}_i)\}\right]^2 \\ &= \sum_{i=1}^{N} \left[I_1(u_i, v_i) - I_2(a_1 x_i + a_2 y_i + a_3, a_4 x_i + a_5 y_i + a_6)\right]^2 \end{aligned} \tag{5}$$

3.1. Newton-Raphson

The minimum of a function occurs at points where the first derivatives of the function vanishes to zero. We now solve for:

$$B_k(\mathbf{a}) = \frac{\partial \chi^2(\mathbf{a})}{\partial a_k} = -2\sum_{i=1}^{N} \left[I_1(\mathbf{u}_i) - I_2'(\mathbf{u}_i)\right] \frac{\partial I_2'(\mathbf{u}_i)}{\partial a_k} = 0 \tag{6}$$

where $k = 1, 2, ..., 6$. This gives us a system of six nonlinear equations:

$$B_k(\mathbf{a}) = 0 \quad k = 1, 2, ..., 6 \tag{7}$$

The linear part of the Taylor series gives us:

$$B_k(\mathbf{a} + \Delta\mathbf{a}) \approx B_k(\mathbf{a}) + \sum_{l=1}^{6} \frac{\partial B_k(\mathbf{a})}{\partial a_l} \Delta a_l \tag{8}$$

Six unknown variables and six linear equations gives us:

$$\mathbf{B}(\mathbf{a} + \Delta\mathbf{a}) = \mathbf{B}(\mathbf{a}) + \mathbf{H}(\mathbf{a})\Delta\mathbf{a} \tag{9}$$

If $(\mathbf{a} + \Delta\mathbf{a})$ is the solution, then $\mathbf{B}(\mathbf{a} + \Delta\mathbf{a}) = 0$ and we have:

$$\mathbf{H}(\mathbf{a})\Delta\mathbf{a} = -\mathbf{B}(\mathbf{a}) \tag{10}$$

$\mathbf{H}(\mathbf{a})$ is a 6×6 matrix and is known as the *Hessian* matrix:

$$h_{kl} = \frac{\partial B_k(\mathbf{a})}{\partial a_l} = \frac{\partial^2 \chi^2(\mathbf{a})}{\partial a_k \partial a_l} = 2\sum_{i=1}^{N} \left[\frac{\partial I_2'(\mathbf{u}_i)}{\partial a_k}\frac{\partial I_2'(\mathbf{u}_i)}{\partial a_l} - \left[I_1(\mathbf{u}_i) - I_2'(\mathbf{u}_i)\right] \frac{\partial^2 I_2'(\mathbf{u}_i)}{\partial a_k \partial a_l}\right] \tag{11}$$

Due to symmetry, $h_{kl} = h_{lk}$. We then solve for six values in $\Delta\mathbf{a}$ by Gauss elimination. The advantage of this method is fast convergence near the solution. Its disadvantages include difficulty in finding a good initial estimate, and possibly slow convergence or oscillations if the function is not well-behaved.

3.2. Steepest (Gradient) Descent

In any iterative minimization, the $(i+1)$st iterate is related to the ith iterate by the equation:

$$\mathbf{a}_{i+1} = \mathbf{a}_i + \Delta\mathbf{a}_i \tag{12}$$

We call $\{\mathbf{a}_{i+1}\}$ a *descending sequence* if $\chi^2(\mathbf{a}_{i+1}) < \chi^2(\mathbf{a}_i)$. One direction that surely produces a descent is the direction of the negative gradient. Therefore, $\mathbf{a}_{i+1} = \mathbf{a}_i - \mathbf{C}\nabla\chi^2(\mathbf{a})$, where $\mathbf{C}$ controls the step size along each parameter in $\mathbf{a}$.

$$\begin{aligned} \Delta\mathbf{a}_i &= \mathbf{a}_{i+1} - \mathbf{a}_i \\ &= -\mathbf{C}\nabla\chi^2(\mathbf{a}) \\ &= -\mathbf{C}\mathbf{B}(\mathbf{a}) \end{aligned} \tag{13}$$

An advantage of this method is that it always converges to a (local) minimum. A disadvantage, though, is its slow convergence.

3.3. Levenberg-Marquadt Algorithm

This method is based on the combination of the Newton-Raphson method and the steepest descent method. From the Newton-Raphson method, we have $\mathbf{H}(\mathbf{a})\Delta\mathbf{a} = -\mathbf{B}(\mathbf{a})$. From the steepest descent method, we have $\Delta\mathbf{a} = -\mathbf{C}\mathbf{B}(\mathbf{a})$. If we let $\mathbf{C}_k = \frac{1}{\lambda h_{kk}}$, then $\lambda h_{kk}\Delta\mathbf{a} = -\mathbf{B}(\mathbf{a})$. Marquadt proved that the following hybrid method minimizes $\chi^2(\mathbf{a})$ more robustly.[15]

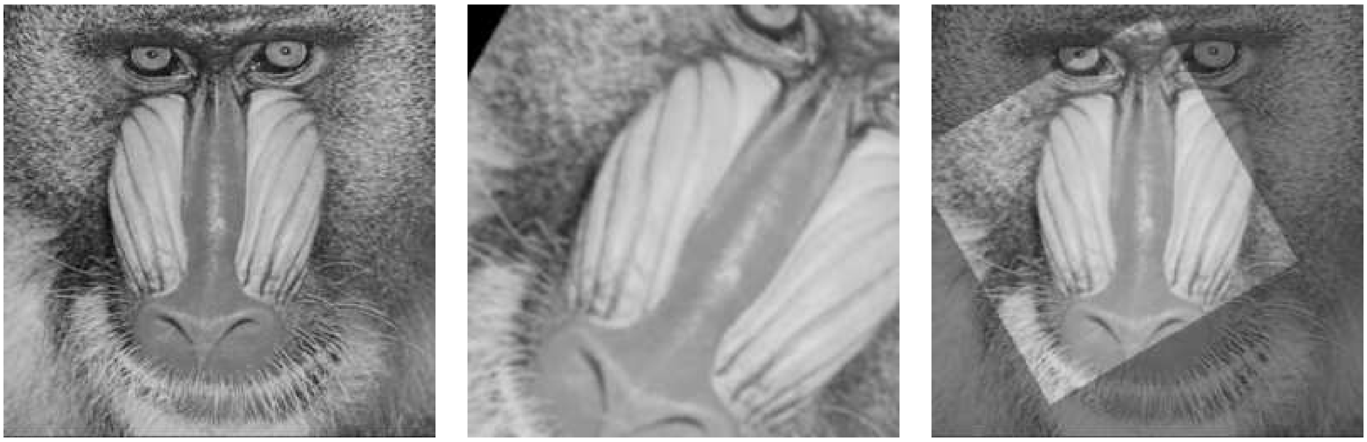
$$(\mathbf{H}(\mathbf{a}) + \lambda\mathbf{I})\Delta\mathbf{a} = -\mathbf{B}(\mathbf{a}) \tag{14}$$

Parameter $\lambda \geq 0$ controls the extent to which $\Delta\mathbf{a}$ conforms to either the Newton-Raphson or steepest descent methods. If the previous update succeeds in reducing $\chi^2(\mathbf{a})$, then λ is reduced to adopt the Newton-Raphson update. Otherwise, λ is increased to adopt a steepest descent update. The resulting method, known as the Levenberg-Marquardt algorithm (LMA), is given below. The algorithm requires threshold values T_1 and T_2 to specify stop conditions. The algorithm terminates when the change in $\Delta\chi^2$ falls below T_1 or λ rises above T_2.

Levenberg-Marquadt Algorithm
begin
  Initialize parameters to the identity matrix
  Initialize λ with a modest value, e.g., $\lambda = 0.1$
  while ($|\Delta\chi^2(\mathbf{a})| > T_1$ or $\lambda < T_2$) **do**
    Compute the 6×6 Hessian matrix $\mathbf{H}$
    Apply affine transformation on I_2
    Compute vector $\mathbf{B}$
    Solve linear equations $\mathbf{H}\Delta\mathbf{a} = -\mathbf{B}$ for $\Delta\mathbf{a}$
    Evaluate $\chi^2(\mathbf{a} + \Delta\mathbf{a})$
    if $\chi^2(\mathbf{a} + \Delta\mathbf{a}) < \chi^2(\mathbf{a})$ **then do**
      $\lambda \leftarrow \lambda/10$
      $\mathbf{a} \leftarrow \mathbf{a} + \Delta\mathbf{a}$
    else do
      $\lambda \leftarrow \lambda * 10$
    end if
  end while
end

3.4. Example

Fig. 1 demonstrates affine registration. Image pair I_1 and I_2 are shown in Figs. 1(a) and 1(b), respectively. Affine registration estimates the six affine parameters necessary to transform I_2 onto I_1. Fig. 1(c) shows the overlay of I_1 and the transformed I_2. Minimal double-exposure (or ghosting) effects illustrate that the image pair is registered with high fidelity.



(a) image I_1 (b) image I_2 (c) overlay of registered images

Figure 1. Affine registration.

4. PERSPECTIVE PARAMETER ESTIMATION

In this section, we extend the results of affine registration to handle perspective transformations. Fig. 2 demonstrates perspective registration. Image pair I_1 and I_2 are shown in Figs. 2(a) and 2(b), respectively. Perspective registration estimates the eight parameters necessary to transform I_2 onto I_1. Fig. 2(c) shows the overlay of I_1 and the transformed I_2. Minimal double-exposure (or ghosting) effects illustrate that the image pair is registered with high fidelity.

(a) image I_1 (b) image I_2 (c) overlay of registered images

Figure 2. Perspective registration.

This method cannot, in general, register images subjected to large perspective distortion. The primary difficulty is that the initial guess may not lie close to the solution and the minimization procedure may get stuck in a local minima. This problem is more severe for perspective than affine because perspective transformations are nonlinear.

We propose to circumvent this problem by approximating perspective to be piecewise affine. In this manner, we leverage the robust affine registration algorithm to handle the more difficult perspective registration problem. A local affine approximation is suggested by expanding the perspective transformation function (Eq. (3)) about a point using a first-order Taylor series. The approximation holds in a small neighborhood about point (x_0, y_0):

$$\begin{aligned} u &= U(x,y) \\ &= U(x_0,y_0) + \frac{\partial U(x_0,y_0)}{\partial x}(x-x_0) + \frac{\partial U(x_0,y_0)}{\partial y}(y-y_0) \\ &= A_1 x + A_2 y + A_3 \end{aligned} \tag{15}$$

where $A_1 = \frac{\partial U(x_0,y_0)}{\partial x}$, $A_2 = \frac{\partial U(x_0,y_0)}{\partial y}$, and $A_3 = U(x_0, y_0) - A_1 x_0 - A_2 y_0$. The expression for v is

$$\begin{aligned} v &= V(x,y) \\ &= V(x_0,y_0) + \frac{\partial V(x_0,y_0)}{\partial x}(x - x_0) + \frac{\partial V(x_0,y_0)}{\partial y}(y - y_0) \\ &= A_4 x + A_5 y + A_6 \end{aligned} \tag{16}$$

where $A_4 = \frac{\partial V(x_0,y_0)}{\partial x}$, $A_5 = \frac{\partial V(x_0,y_0)}{\partial y}$, and $A_6 = V(x_0, y_0) - A_4 x_0 - A_5 y_0$. Notice that the expressions for u and v match the linear form required of affine transformations.

In the next two sections, we review the process of determining the affine transformation for a four-corner (tile-to-tile) mapping and its use in inferring perspective parameters. We have already reviewed how to perform affine registration on a rectangular image domain, e.g., tile. For the purpose of demonstrating the feasibility of piecewise affine approximation, we will show how to infer the affine parameters given the correspondence of the tile corners. This permits us to analyze the proposed technique without concern for errors that may be due to the affine registration process using the LMA algorithm described in Section 3.

4.1. Inferring Affine Parameters

Consider a tile in I_1 and its corresponding quadrilateral in I_2. (Fig. 3).

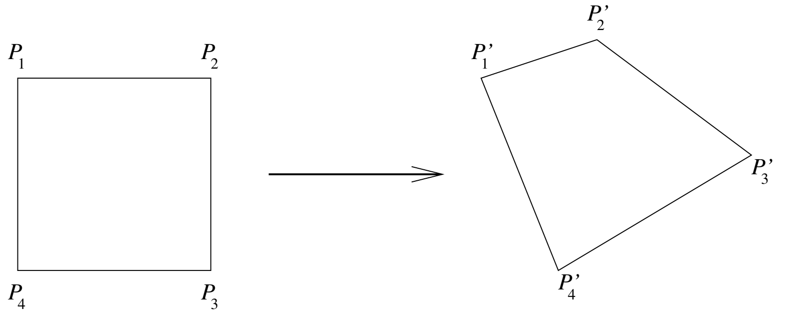


Figure 3. Four corner mapping.

Although a perspective transformation can fully account for this four corner mapping, we are interested in finding the best affine transformation that approximates this mapping. Given the four corners of a tile in observed image I_2 and their correspondences in reference image I_1, we may solve for the best affine fit by using the least squares approach. Let the affine mapping be given as:

$$u = a_1 x + a_2 y + a_3 \tag{17a}$$

$$v = a_4 x + a_5 y + a_6 \tag{17b}$$

We solve for the affine parameters by minimizing the expression for χ^2 below.

$$\chi^2(\mathbf{a}) = \sum_{i=1}^{4} \left[(u_i - a_1 x_i - a_2 y_i - a_3)^2 + (v_i - a_4 x_i - a_5 y_i - a_6)^2 \right] \tag{18}$$

We may relate these correspondences in the form $U = WA$:

$$\begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix} = \begin{bmatrix} x_1 & y_1 & 1 & 0 & 0 & 0 \\ x_2 & y_2 & 1 & 0 & 0 & 0 \\ x_3 & y_3 & 1 & 0 & 0 & 0 \\ x_4 & y_4 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & x_1 & y_1 & 1 \\ 0 & 0 & 0 & x_2 & y_2 & 1 \\ 0 & 0 & 0 & x_3 & y_3 & 1 \\ 0 & 0 & 0 & x_4 & y_4 & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \end{bmatrix} \tag{19}$$

The pseudoinverse solution $A = (W^T W)^{-1} W^T U$ is computed to solve for the six affine coefficients.

4.2. Inferring Perspective Parameters

The affine parameters computed in Eq. (19) apply to a single tile pair. We must infer the affine parameters for each of the N tile pairs in I_1 and I_2, and apply them to all tile centers. This establishes N point correspondences between I_1 and I_2, The eight unknown perspective parameters can be inferred by solving the resulting system of equations. Again, we may relate the correspondences in the form $U = WA$ by rewriting Eq. (3) in the form

$$u = a_1x + a_2y + a_3 - a_7ux - a_8uy \tag{20a}$$
$$v = a_4x + a_5y + a_6 - a_7vx - a_8vy \tag{20b}$$

This yields the following overdetermined system of equations.

$$\begin{bmatrix} u_0^i \\ \\ \vdots \\ \\ v_0^i \\ \\ \vdots \end{bmatrix}_{2N\times 1} = \begin{bmatrix} x_0^i & y_0^i & 1 & 0 & 0 & 0 & -u_0^i x_0^i & -u_0^i y_0^i \\ & & & & \vdots & & & \\ 0 & 0 & 0 & x_0^i & y_0^i & 1 & -v_0^i x_0^i & -v_0^i y_0^i \\ & & & & \vdots & & & \end{bmatrix}_{2N\times 8} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \\ a_7 \\ a_8 \end{bmatrix}_{8\times 1} \tag{21}$$

where (u_0^i, v_0^i) and (x_0^i, y_0^i) denote the centers of tile i in I_1 and I_2 for $1 \le i \le N$. The pseudoinverse solution $A = (W^TW)^{-1}W^TU$ is computed to solve for the eight perspective coefficients.

We may augment these results by including additional constraints in the least squares estimation. In addition to considering the correspondence of tile centers, we may also consider the local affine parameter estimates. From Eqs. (15) and (16), we know that the affine approximation at a tile is given as

$$u = A_1x + A_2y + A_3 \tag{22a}$$
$$v = A_4x + A_5y + A_6 \tag{22b}$$

where

$$\begin{aligned}
A_1 &= \frac{\partial U(x_0,y_0)}{\partial x} = \frac{a_1(a_7x_0 + a_8y_0 + 1) - a_7(a_1x_0 + a_2y_0 + a_3)}{(a_7x_0 + a_8y_0 + 1)^2} = \frac{a_1 - a_7u_0}{a_7x_0 + a_8y_0 + 1} \\
&= a_1 - a_7(A_1x_0 + u_0) - a_8A_1y_0 && \text{(23a)} \\
A_2 &= \frac{\partial U(x_0,y_0)}{\partial y} = \frac{a_2(a_7x_0 + a_8y_0 + 1) - a_8(a_1x_0 + a_2y_0 + a_3)}{(a_7x_0 + a_8y_0 + 1)^2} = \frac{a_2 - a_8u_0}{a_7x_0 + a_8y_0 + 1} \\
&= a_2 - a_7A_2x_0 - a_8(A_2y_0 + u_0) && \text{(23b)} \\
A_3 &= U(x_0,y_0) - A_1x_0 - A_2y_0 = \frac{a_1x_0 + a_2y_0 + a_3}{a_7x_0 + a_8y_0 + 1} - A_1x_0 - A_2y_0 \\
&= a_1x_0 + a_2y_0 + a_3 - a_7(A_1x_0^2 + A_2x_0y_0 + A_3x_0) - a_8(A_1x_0y_0 + A_2y_0^2 + A_3y_0) - (A_1x_0 + A_2y_0) \\
&= a_1x_0 + a_2y_0 + a_3 - a_7u_0x_0 - a_8u_0y_0 - (u_0 - A_3) && \text{(23c)} \\
A_4 &= \frac{\partial V(x_0,y_0)}{\partial x} = \frac{a_4(a_7x_0 + a_8y_0 + 1) - a_7(a_4x_0 + a_5y_0 + a_6)}{(a_7x_0 + a_8y_0 + 1)^2} = \frac{a_4 - a_7v_0}{a_7x_0 + a_8y_0 + 1} \\
&= a_4 - a_7(A_4x_0 + v_0) - a_8A_4y_0 && \text{(23d)} \\
A_5 &= \frac{\partial V(x_0,y_0)}{\partial y} = \frac{a_5(a_7x_0 + a_8y_0 + 1) - a_8(a_4x_0 + a_5y_0 + a_6)}{(a_7x_0 + a_8y_0 + 1)^2} = \frac{a_5 - a_8v_0}{a_7x_0 + a_8y_0 + 1} \\
&= a_5 - a_7A_5x_0 - a_8(A_5y_0 + v_0) && \text{(23e)} \\
A_6 &= V(x_0,y_0) - A_4x_0 - A_5y_0 = \frac{a_4x_0 + a_5y_0 + a_6}{a_7x_0 + a_8y_0 + 1} - A_4x_0 - A_5y_0 \\
&= a_4x_0 + a_5y_0 + a_6 - a_7(A_4x_0^2 + A_5x_0y_0 + A_6x_0) - a_8(A_4x_0y_0 + A_5y_0^2 + A_6y_0) - (A_4x_0 + A_5y_0) \\
&= a_4x_0 + a_5y_0 + a_6 - a_7v_0x_0 - a_8v_0y_0 - (v_0 - A_6) && \text{(23f)}
\end{aligned}$$

Notice that the terms for A_3 and A_6 actually produce equations of the form $u_0 = a_1x_0 + a_2y_0 + a_3 - a_7u_0x_0 - a_8u_0y_0$ and $v_0 = a_4x_0 + a_5y_0 + a_6 - a_7v_0x_0 - a_8v_0y_0$, respectively. These are the same point correspondence relations defined in Eq. (21). We may collect the expressions given above to yield the following system of equations that relates the affine parameters at all N tiles with the unknown parameters of the global perspective transformation.

$$\begin{bmatrix} A_1^i \\ \vdots \\ A_2^i \\ \vdots \\ u_0^i \\ \vdots \\ A_4^i \\ \vdots \\ A_5^i \\ \vdots \\ v_0^i \\ \vdots \end{bmatrix}_{6N\times 1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & -(A_1^i x_0^i + u_0^i) & -A_1^i y_0^i \\ & & & & & \vdots & & \\ 0 & 1 & 0 & 0 & 0 & 0 & -A_2^i x_0^i & -(A_2^i y_0^i + u_0^i) \\ & & & & & \vdots & & \\ x_0^i & y_0^i & 1 & 0 & 0 & 0 & -u_0^i x_0^i & -u_0^i y_0^i \\ & & & & & \vdots & & \\ 0 & 0 & 0 & 1 & 0 & 0 & -(A_4^i x_0^i + v_0^i) & -A_4^i y_0^i \\ & & & & & \vdots & & \\ 0 & 0 & 0 & 0 & 1 & 0 & -A_5^i x_0^i & -(A_5^i y_0^i + v_0^i) \\ & & & & & \vdots & & \\ 0 & 0 & 0 & x_0^i & y_0^i & 1 & -v_0^i x_0^i & -v_0^i y_0^i \\ & & & & & \vdots & & \end{bmatrix}_{6N\times 8} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \\ a_7 \\ a_8 \end{bmatrix}_{8\times 1} \tag{24}$$

where (u_0^i, v_0^i) and (x_0^i, y_0^i) denote the centers of tile i in I_1 and I_2 for $1 \leq i \leq N$. Note that Eq. (24) is a superset of Eq. (21). Again, the pseudoinverse solution is computed to solve for the eight perspective parameters. The solution to Eq. (24) is normally not stable if pixel coordinates are directly used. We therefore perform a simple normalization to bring the pixel coordinates into the [0,1] range.

4.3. Piecewise Affine Approximation

Consider an observed image I_2 subdivided into a regular grid of tiles. For each tile T_2^i, we perform affine registration using the Levenberg-Marquadt algorithm to search for the most similar tile T_1^i in reference image I_1. This yields a collection of affine parameters and tile centers. Note that all tiles must share the same dimensions in order to compute χ^2. The algorithm is outlined below.

Piecewise Affine Approximation Algorithm
begin
    Partition I_2 into a collection of N tiles: T_2^i, for $1 \leq i \leq N$
    $i \leftarrow 1$
    while $(i < N)$ **do**
        Select tile T_2^i
        for all positions (x, y) in I_1 **do**
            Crop tile T_1^i in I_1
            Register T_2^i with T_1^i using LMA
            if $\chi^2(\mathbf{a})$ is minimum **then do**
                $x^i \leftarrow x$
                $y^i \leftarrow y$
                $\mathbf{a}^i \leftarrow \mathbf{a}$
            end if
        end for
        $i \leftarrow i + 1$
    end while
    Solve Eq. (21) or Eq. (24) for perspective parameters using pseudoinverse solution (linear least squares)
end

4.4. Examples

Fig. 4 demonstrates the proposed algorithm. The top row of the figure depicts an image subdivided into a uniform grid of 2×2, 4×4, and 8×8 tiles, respectively. We consider these three resolutions to demonstrate the relationship between tile size and approximation error.

The second row shows the effects of a perspective warp applied to the input images. The overlaid grids and the marked tiles centers clearly illustrate the foreshortening effects of perspective.

The piecewise affine approximation of a perspective transformation is shown in the third row. Notice that each input tile has undergone an affine warp that best approximates the four corner mapping to its corresponding tile in the warped image. The true perspective grids (second row) are superimposed to illustrate the error with respect to the ground truth.

The positions of the tile centers from the input (first row) and piecewise affine images (third row) are supplied to Eq. (21) to infer a global perspective transformation. The results are shown in the fourth row, with the true perspective grids (and tile centers) overlaid on the figure. The mapping is due to the perspective transformation inferred directly from the correspondences of the tile centers. The last row of the figure shows the improved approximation due to Eq. (24), when the local affine parameter estimates are used alongside the point correspondences.

The Taylor series approximation is adequate in only a small neighborhood around a point, i.e., tile center. Clearly, a 2×2 tile subdivision is too coarse, as depicted in the first column of Fig. 4. Finer subdivision into 16 and 64 tiles, as shown in the second and third columns of the figure, produce greatly improved results. The use of small tiles thus enforces the assumption implicit in the Taylor series, and thereby helps reduce errors.

Fig. 5 demonstrates the algorithm on a pair of images subjected to a large perspective deformation. Fig. 5(c) illustrates a set of manually chosen tiles drawn from Fig. 5(a). These tiles then undergo affine registration to establish correspondence with Fig. 5(b). The point correspondences of the tile centers were sufficient to achieve the perspective registration shown in Fig. 5(d). Notice that the minimal double-exposure effect in the figure is evidence of a good match.

The choice of tiles is important to the process. The perspective transformation computed above was applied to a regular grid. A piecewise affine approximation was applied to the warped grid to yield Fig. 6. Notice that many of the image tiles are lacking proper visual structure that would have made accurate matches possible. Therefore, only tiles replete with statistically significant data must be registered. The tiles shown in Fig. 5 were selected for their rich visual content.

5. SUMMARY AND CONCLUSIONS

A method to register image pairs subjected to large perspective deformations has been presented. The images are assumed to be acquired under a weak perspective camera model or a fixed center of projection. These are necessary conditions to avoid parallax and facilitate registration in the presence of a global perspective transformation.

We have demonstrated the feasibility of approximating a perspective transformation as piecewise affine. A reference image is subdivided into tiles and affine registration is applied to find the best matching tiles in the target image. These tile pairs are used to estimate a global perspective transformation. In this manner, we exploit recent results in affine registration to solve the more difficult perspective registration problem.

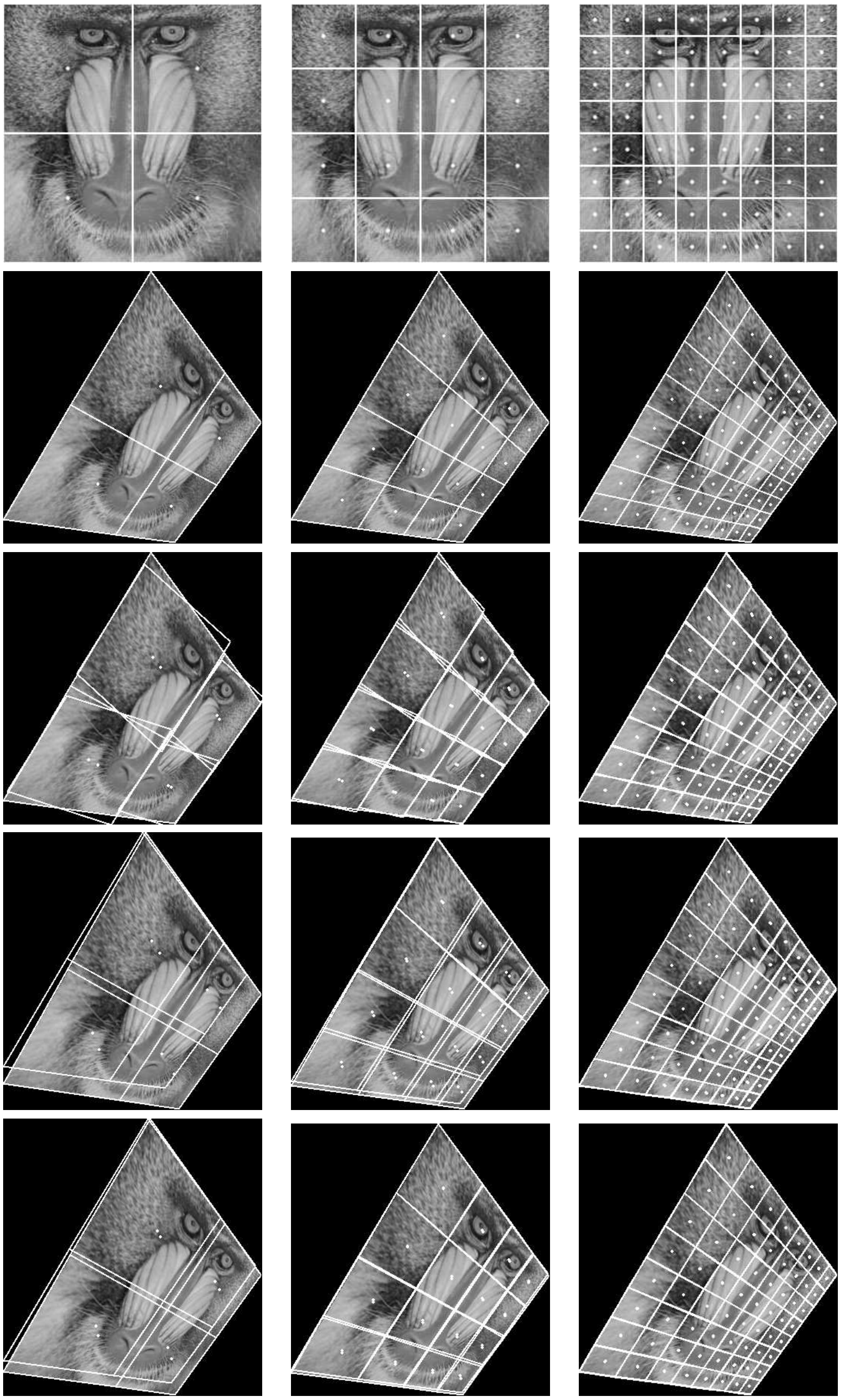
In affine registration, we estimate the affine parameters necessary to register any two digital images misaligned due to rotation, scale, shear, and translation. The parameters are selected to minimize the sum of squared differences between the two images. They are computed iteratively in a coarse-to-fine hierarchical framework using a variation of the Levenberg-Marquadt nonlinear least squares optimization method. This approach yields a robust solution that precisely registers images with subpixel accuracy.

After applying affine registration to all tiles, the correspondence of tile centers are used to solve for the eight perspective parameters. The accuracy of the technique is further refined by considering the local affine parameter estimates.

A critical component to this work is the affine registration among tiles. As we pointed out in Section 4, the choice of tiles is important to the process. Future work will address the tile selection process. In particular, we will investigate the minimal set of tiles necessary to achieve accurate registration. The tile size is also critical to the accuracy and computational performance of the algorithm. There is a tradeoff between registration accuracy and the Taylor series approximation error. Although the use of small tiles conforms more closely to the Taylor series approximation, it lowers the registration accuracy because it offers less structure and visual detail. An automated method of determining this tradeoff will be investigated. We will also investigate the use of more robust optimization techniques for eliminating outliers. In particular, we will compare the least median squares with our existing linear least squares (pseudoinverse) approach.

REFERENCES

1. L. G. Brown, "A survey of image registration techniques," *ACM Computing Surveys* **24**, pp. 325–376, December 1992.
2. J. R. Bergen and E. Adelson, "Hierarchical, computationally efficient motion estimation algorithm," *J. Opt. Soc. Am.* **4**(35), 1987.
3. P. Anandan, "A computational framework and an algorithm for the measurement of visual motion," *Intl. J. Computer Vision* **2**, pp. 283–310, 1989.
4. J. R. Bergen, P. Anandan, K. J. Hanna, and R. Hingorani, "Hierarchical model-based motion estimation," *Proc. Euro. Conf. Computer Vision (ECCV)* , pp. 237–252, 1992.
5. S. Chen, "Quicktime VR: An image-based approach to virtual environment navigation," *Proc. Siggraph '95* , pp. 29–38, 1995.
6. M. Irani and P. Anandan, "Video indexing based on mosaic representations," *Proc. IEEE* **86**, pp. 237–252, May 1998.
7. R. Szeliski, "Image mosaicing for tele-reality applications," *Proc. IEEE Workshop on Applications of Computer Vision* , pp. 230–236, 1994.
8. R. Szeliski and H.-Y. Shum, "Video mosaics for virtual environments," *IEEE Computer Graphics and Applications* **16**, pp. 22–30, 1996.
9. R. Szeliski and H.-Y. Shum, "Creating full view panoramic image mosaics and environment maps," *Proc. Siggraph '97* , pp. 251–258, 1997.
10. H.-Y. Shum and R. Szeliski, "Construction and refinement of panoramic mosaics with global and local alignment," *Proc. Intl. Conf. Computer Vision* , pp. 953–958, 1998.
11. S. Mann and R. W. Picard, "Video orbits of the projective group: A simple approach to featureless estimation of parameters," *IEEE Trans. Image Processing* **6**, pp. 1281–1295, September 1997.
12. J. Davis, "Mosaics of scenes with moving objects," *IEEE Conf. Computer Vision and Pattern Recognition (CVPR)* , 1998.
13. P. Thévenaz, U. E. Ruttimann, and M. Unser, "A pyramid approach to subpixel registration based on intensity," *IEEE Trans. Image Processing* **7**(1), pp. 27–41, 1998.
14. D. Aiger and D. Cohen-Or, "Mosaicing ultrasonic volumes for visual simulation," *IEEE Computer Graphics and Applications* **20**, pp. 53–61, March/April 2000.
15. D. W. Marquadt, "An algorithm for least-squares estimation of nonlinear parameters," *J. Society for Industrial and Applied Mathematics* **11**(2), pp. 431–441, 1963.



(a) (b) (c)

Figure 4. Piecewise affine approximation of perspective with (a) 2 × 2 tiles, (b) 4 × 4 tiles, and (c) 8 × 8 tiles.

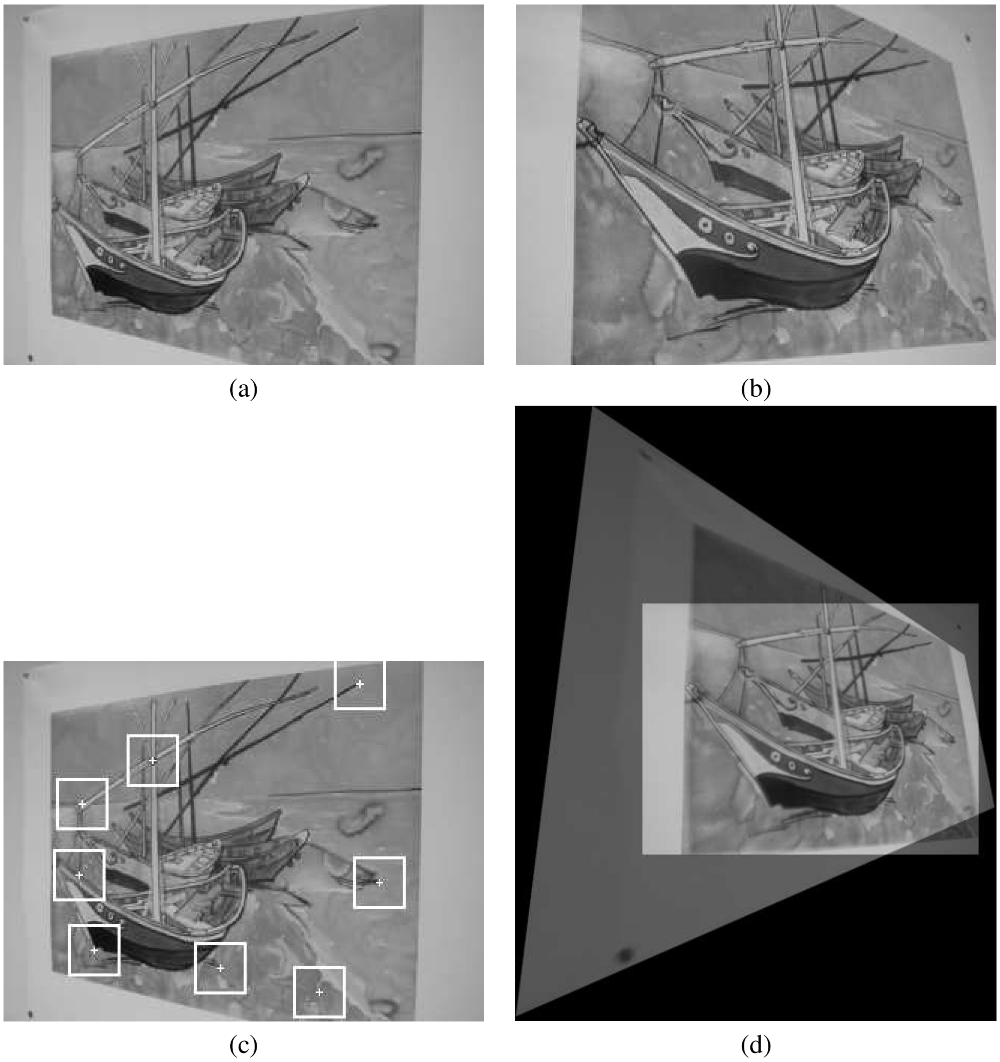


Figure 5. Perspective registration. (a) Image I_1; (b) Image I_2; (c) Selected tiles; (d) Overlay of registered I_1 on I_2.

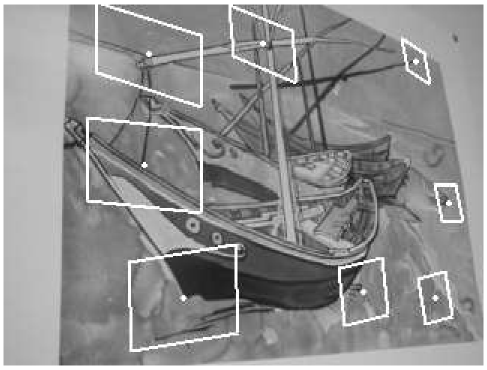


Figure 6. Piecewise affine approximation.