AN OVERVIEW OF THE STEREO CORRELATION AND TRIANGULATION FORMULATIONS USED IN DICE

D. Z. TURNER

ABSTRACT. This document provides a detailed overview of the stereo correlation algorithm and triangulation formulation used in the Digital Image Correlation Engine (DICe) to triangulate three dimensional motion in space given the image coordinates and camera calibration parameters. [Report No. SAND2017-1875 R]

1. Introduction

The objective of this document is to outline, in detail, the process used in DICe to perform stereo correlation, including the triangulation step. From an abstract perspective, this process involves the following steps:

- (1) A projective mapping is optimized that relates image points in the left camera image to the right camera image assuming a perfect projection.
- (2) A cross correlation is performed between the left camera image and the right camera image for the first frame. In cases where the lens distortion is low and the object of interest is a flat plane, this step can be skipped and the projective mapping is used to determine the q and r fields which represent the displacement of a subset centroid from its position in left camera sensor coordinates to right sensor coordinates.
- (3) The left camera images are correlated from one frame to the next and the right camera images are correlated from one frame to the next.
- (4) Given the sensor positions of the subsets in the left and right images, the three dimensional position of each subset centroid is triangulated.
- (5) (optional) The three dimensional model coordinates are projected to a best fit plane among all the data points.
- (6) The strains are computed using the final model coordinate system, which may be a best fit plane.

Items 3, 4, (5) and 6 are repeated for each frame in the sequence.

2. Triangulation

This section describes how three dimensional points in model space are triangulated in DICe, given the projection of that point into two images and the intrinsic and extrinsic camera calibration parameters. We also discuss lens distortion correction. We assume that the calibration parameters are known and do not discuss estimating the calibration parameters or bundle adjustment. Obtaining the calibration parameters can be done in DICe using the OpenCV stereo calibration routines that have been incorporated.

3. Preliminaries

In this section, we describe some fundamental concepts that will be helpful to understand the workflow of the triangulation process.

Date: February 21, 2017.

1

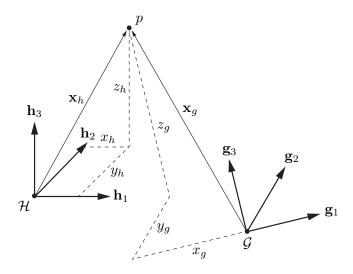


FIGURE 1. Two coordinate systems with origins \mathcal{H} and \mathcal{G} , and a point, p, defined in terms of a position vector in each coordinate system.

- 3.1. Coordinate systems. A three dimensional coordinate system is defined in terms of an origin and three basis vectors or cardinal directions. For simplicity, we assume that the basis vectors are orthogonal and that they follow the right hand rule. Consider the coordinate system shown in Figure 1 with origin, \mathcal{H} , and basis vectors, \mathbf{h}_i . A point, p, in space, is defined in terms of a position vector, \mathbf{x}_h . The components of \mathbf{x}_h are relative to a particular coordinate system, in this case $\mathbf{x}_h = x_h \mathbf{h}_1 + y_h \mathbf{h}_2 + z_h \mathbf{h}_3$. A second coordinate system is also shown in Figure 1 with origin \mathcal{G} and basis vectors \mathbf{g}_i . The same point, p, is defined in this coordinate system with position vector $\mathbf{x}_g = x_g \mathbf{g}_1 + y_g \mathbf{g}_2 + z_g \mathbf{g}_3$. In the formulation that follows, we will refer to the coordinate system with origin \mathcal{H} as coordinate system \mathcal{H} .
- 3.2. Coordinate transformation. We define a coordinate transformation matrix, \mathbf{T} , that converts the components of a position vector in one coordinate system to another. The matrix, \mathbf{T} , has the following form

$$\mathbf{T} = \begin{bmatrix} R_{11} & R_{12} & R_{13} & t_x \\ R_{21} & R_{22} & R_{23} & t_y \\ R_{31} & R_{32} & R_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} , \tag{1}$$

where the components R_{ij} represent rotations that, when applied to the source coordinate axes, align the source and target coordinate axes. The components t_i represent translations along each of the source basis vectors from the source origin to the origin of the target coordinate system (see Figure 2.

The components of **T** are determined using the Cardan-Bryant angles as described in Section 3.3 and the translations t_x, t_y , and t_z are given in the calibration parameters. In general, for a coordinate transformation from \mathcal{H} to \mathcal{G} the translations are needed in reference to coordinate system \mathcal{G} and the rotations are

$$R_{i,j} = \cos\theta(\mathbf{g}_i, \mathbf{h}_j). \tag{2}$$

In words, the rotation matrix is defined in terms of the angles formed between the basis axes in \mathcal{G} to \mathcal{H} . Since for a the transformation from \mathcal{H} to \mathcal{G} we know the translations in terms of \mathcal{H} , it is easier to compute the transformation from \mathcal{G} to \mathcal{H} and take the inverse.

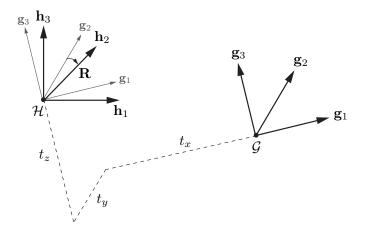


FIGURE 2. Definition of terms in the transformation matrix $\mathbf{T}_{\mathcal{H}\to\mathcal{G}}$ that transforms position vectors in the \mathcal{H} coordinate system to \mathcal{G} .

To make the transformation notation more descriptive in terms of the direction of the transform, we denote the transform from coordinate system \mathcal{H} to \mathcal{G} as $\mathbf{T}_{\mathcal{H}\to\mathcal{G}}$. The transformation of position vector components proceeds as follows,

$$\begin{bmatrix} x_g \\ y_g \\ z_g \\ 1 \end{bmatrix} = \mathbf{T}_{\mathcal{H} \to \mathcal{G}} \begin{bmatrix} x_h \\ y_h \\ z_h \\ 1 \end{bmatrix} . \tag{3}$$

The inverse transform, from \mathcal{G} to \mathcal{H} , is given by taking the inverse of the transform matrix as follows,

$$\begin{bmatrix} x_h \\ y_h \\ z_h \\ 1 \end{bmatrix} = \mathbf{T}_{\mathcal{H} \to \mathcal{G}}^{-1} \begin{bmatrix} x_g \\ y_g \\ z_g \\ 1 \end{bmatrix} . \tag{4}$$

In other words, the following relationship holds: $\mathbf{T}_{\mathcal{H}\to\mathcal{G}}^{-1} = \mathbf{T}_{\mathcal{G}\to\mathcal{H}}$.

3.3. Cardan-Bryant angles. The nine-component rotation matrix, \mathbf{R} , can be uniquely defined in terms of three angles known as the Cardan-Bryant angles, α , β , and γ . The following relationships can be used to determine the rotation matrix components from the Cardan-Bryant angles

$$R_{11} = \cos(\alpha)\cos(\beta) \tag{5a}$$

$$R_{12} = \cos(\beta)\sin(\alpha) \tag{5b}$$

$$R_{13} = -\sin(\beta) \tag{5c}$$

$$R_{21} = \cos(\alpha)\sin(\beta)\sin(\gamma) - \cos(\gamma)\sin(\alpha) \tag{5d}$$

$$R_{22} = \sin(\alpha)\sin(\beta)\sin(\gamma) + \cos(\alpha)\cos(\gamma) \tag{5e}$$

$$R_{23} = \cos(\beta)\sin(\gamma) \tag{5f}$$

$$R_{31} = \cos(\alpha)\cos(\gamma)\sin(\beta) + \sin(\alpha)\sin(\gamma) \tag{5g}$$

$$R_{32} = \cos(\gamma)\sin(\alpha)\sin(\beta) - \cos(\alpha)\sin(\gamma) \tag{5h}$$

$$R_{33} = \cos(\beta)\cos(\gamma) \tag{5i}$$

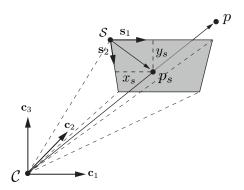


FIGURE 3. Definition of terms in the pinhole camera model and the projective transform to the image plane (the gray shape represents the image).

3.4. Projective transformation to the image plane. Consider a point, p, in space with a position vector, \mathbf{x}_c , relative to the coordinate system \mathcal{C} , where \mathcal{C} represents the origin of a pinhole camera model. The projective transform takes the point, p, given in coordinates relative to \mathcal{C} , and transforms it to the sensor coordinates of point p_s in the image plane. The point p_s is given in pixel coordinates (x_s, y_s) . This is shown in Figure 3, where \mathcal{S} represents the sensor or image coordinate system origin. The projective transformation from \mathcal{C} to \mathcal{S} is defined in terms of a matrix, \mathbf{F} , with the following components

$$\mathbf{F} = \begin{bmatrix} f_x & f_s & c_x & 0 \\ 0 & f_y & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} , \tag{6}$$

where f_x and f_y are the focal length parameters, f_s is the skew parameter, c_x is x coordinate, along \mathbf{s}_1 , of the center of the image plane (relative to \mathcal{S}) or the principal point, c_y is the y coordinate, along \mathbf{s}_2 , of the center of the image plane (relative to \mathcal{S}), and ψ is a scale factor. The components of \mathbf{F} are the *intrinsic* parameters associated with the camera that are determined in the camera calibration. The transformation is then given as

$$\psi \begin{bmatrix} x_s \\ y_s \\ 1 \end{bmatrix} = \mathbf{F} \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} . \tag{7}$$

4. Triangulation

In the section above we outlined two basic tools needed to triangulate a three dimensional point in space, coordinate transformation and projective transformation. Coordinate transforms enable converting coordinates from one system to another. Projective transforms connect the sensor or image coordinates to coordinates in physical or model space relative to the pinhole camera coordinate system. Our goal, in this section, is to determine the physical coordinates, $\mathbf{x}_c = (x_c, y_c, z_c)$, of a three dimensional point, p, in space, relative to the coordinate system \mathcal{C}_1 of camera 1, using as input the sensor coordinates of the projection of that point in the image plane associated with camera 1, $p_s^1 = (x_s^1, y_s^1)$, and the sensor coordinates of the projection of that point in the image plane associated with camera 2, $p_s^2 = (x_s^2, y_s^2)$. To do so, we require three pieces of information: the sensor coordinates in both images, the intrinsic calibration parameters from camera 1 and 2 $(c_x^1, c_y^1, f_x^1, f_y^1, f_s^1, c_x^2, c_y^2, f_x^2, f_y^2, f_s^2)$, and the transformation matrix, $\mathbf{T}_{\mathcal{C}_1 \to \mathcal{C}_2}$ (with elements \mathbf{R}, t_x, t_y ,

¹Note that the superscript notation is in reference to the camera id not an exponent.

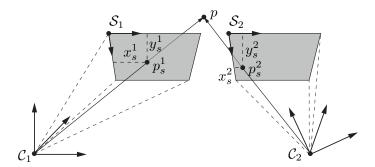


FIGURE 4. Projection of physical or model point p into the sensor planes of camera 1 and 2.

and t_z). The transformation matrix is also referred to as being composed of the extrinsic calibration parameters. A diagram of the triangulation geometry is shown in Figure 4

4.1. Minimization procedure. Were there no errors in the estimation of the sensor coordinates (x_s^2, y_s^2) , one could directly solve for \mathbf{x}_c using the two projective transforms for sensor planes 1 and 2, but because of small errors in the estimation of (x_s^2, y_s^2) and the calibration parameters, a minimization approach must be used. There are many possibilities in regards to which degrees of freedom to optimize. In this formulation, we assume that the calibration parameters are known precisely and aim to determine the triangulated point that minimizes the euclidean distance between the projection of that point from camera 1 and camera 2. In other words, we find the optimal point that is closest to where the projective transforms from each camera would place the model coordinates (which may not coincide precisely).

The projective transform of point p onto sensor plane 1 is given as

$$\psi^1 \begin{bmatrix} x_s^1 \\ y_s^1 \\ 1 \end{bmatrix} = \mathbf{F}^1 \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} . \tag{8}$$

The projective transform of point p onto sensor plane 2 is given as

$$\psi^2 \begin{bmatrix} x_s^2 \\ y_s^2 \\ 1 \end{bmatrix} = \mathbf{F}^2 \mathbf{T}_{\mathcal{C}_1 \to \mathcal{C}_2} \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} , \qquad (9)$$

where the transformation matrix, $\mathbf{T}_{\mathcal{C}_1 \to \mathcal{C}_2}$, is necessary because we wish to express all physical coordinates in reference to \mathcal{C}_1 . Note that all the coordinates on the right hand side of both (8) and (9) are relative to the coordinate system for camera 1.

The matrix form in (8) represents three equations that, after making the substitution $\psi^1 = z_{c1}$, can be written as two residuals

$$r_1(\mathbf{x}_c) \equiv f_x^1 x_c + f_s^1 y_c + (c_x^1 - x_s^1) z_c \tag{10}$$

$$r_2(\mathbf{x}_c) \equiv f_y^1 y_c + (c_y^1 - y_s^1) z_c.$$
(11)

A similar set of residuals can be derived from (9) by making the substitution, $\psi^2 = R_{31}x_c + R_{32}y_c + R_{33}z_c + t_z$,

$$r_3(\mathbf{x}_c) \equiv (f_x^2 R_{11} + f_s^2 R_{21} + (c_x^2 - x_s^2) R_{31}) x_c + (f_x^2 R_{12} + f_s^2 R_{22} + (c_x^2 - x_s^2) R_{32}) y_c$$

$$+ (f_x^2 R_{13} + f_s^2 R_{23} + (c_x^2 - x_s^2) R_{33}) z_c + f_x^2 t_x + f_s^2 t_y + (c_x^2 - x_s^2) t_z$$

$$(12)$$

$$r_4(\mathbf{x}_c) \equiv (f_y^2 R_{21} + (c_y^2 - y_s^2) R_{31}) x_c + (f_y^2 R_{22} + (c_y^2 - y_s^2) R_{32}) y_c$$

$$+ (f_y^2 R_{23} + (c_y^2 - y_s^2) R_{33}) z_c + f_y^2 t_y + (c_y^2 - y_s^2) t_z$$

$$(13)$$

Given the residuals above, the minimization problem can be stated, find the components of the position vector, $\mathbf{x}_c^* = (x_c^*, y_c^*, z_c^*)$, such that

$$\mathbf{x}_c^* = \arg\min_{\mathbf{x}_c} \|\mathbf{r}(\mathbf{x}_c)\|^2, \tag{14}$$

where $\mathbf{r}(\mathbf{x}_c) = (r_1, r_2, r_3, r_4)$. Since (14) represents a linear least squares problem, we can re-write the minimization problem in terms of a generalized inverse: Find \mathbf{x}_c^* , such that

$$\mathbf{x}_c^* = \arg\min_{\mathbf{x}_c} \|\mathbf{M}\mathbf{x}_c - \mathbf{b}\|^2, \tag{15}$$

with normal equations

$$\mathbf{M}^T \mathbf{M} \mathbf{x}_c^* = \mathbf{M}^T \mathbf{b} \,, \tag{16}$$

and solution

$$\mathbf{x}_c^* = (\mathbf{M}^T \mathbf{M})^{-1} \mathbf{M}^T \mathbf{b}. \tag{17}$$

The components of \mathbf{M} are as follows

$$M_{11} = f_x^1 (18a)$$

$$M_{12} = f_s^1 (18b)$$

$$M_{13} = (c_x^1 - x_s^1) (18c)$$

$$M_{21} = 0$$
 (18d)

$$M_{22} = f_y^1 (18e)$$

$$M_{23} = (c_y^1 - y_s^1) (18f)$$

$$M_{31} = f_x^2 R_{11} + f_s^2 R_{21} + (c_x^2 - x_s^2) R_{31}$$
(18g)

$$M_{32} = f_x^2 R_{12} + f_s^2 R_{22} + (c_x^2 - x_s^2) R_{32}$$
(18h)

$$M_{33} = f_x^2 R_{13} + f_s^2 R_{23} + (c_x^2 - x_s^2) R_{33}$$
 (18i)

$$M_{41} = f_y^2 R_{21} + (c_y^2 - y_s^2) R_{31} (18j)$$

$$M_{42} = f_y^2 R_{22} + (c_y^2 - y_s^2) R_{32}$$
(18k)

$$M_{43} = f_y^2 R_{23} + (c_y^2 - y_s^2) R_{33} (181)$$

and **b** is given as

$$b_1 = 0 \tag{19a}$$

$$b_2 = 0 \tag{19b}$$

$$b_3 = -f_x^2 t_x - f_s^2 t_y - (c_x^2 - x_s^2) t_z (19c)$$

$$b_4 = -f_y^2 t_y - (c_y^2 - y_s^2) t_z (19d)$$

5. Lens distortion correction

In the triangulation formulation above, the sensor locations (x_s, y_s) may be inaccurate due to lens distortions. Prior to triangulating, the sensor positions are corrected using a lens distortion model. The lens distortion correction implemented in DICe is based on a radial distortion model, \mathbf{d} , of the following form

$$\mathbf{d} = \begin{bmatrix} c_x \tilde{x} d \\ c_y \tilde{y} d \end{bmatrix} , \tag{20}$$

where

$$d = \kappa_1(\tilde{x}^2 + \tilde{y}^2) + \kappa_2(\tilde{x}^2 + \tilde{y}^2)^2 + \kappa_3(\tilde{x}^2 + \tilde{y}^2)^3.$$
 (21)

 κ_1, κ_2 and κ_3 are the distortion parameters, and $\tilde{x} = (x_s - c_x)/c_x$ and $\tilde{y} = (y_s - c_y)/c_y$ are the normalized sensor coordinates. Note that in the distortion model above, all of the superscripts are exponents, not camera ids. Given the radial distortion model, the ideal (undistorted) sensor position, $\mathbf{x}_{si} = (x_{si}, y_{si})$ is determined as

$$\mathbf{x}_{si} = \mathbf{x}_s - \mathbf{d} \,, \tag{22}$$

where \mathbf{x}_s is the distorted sensor location in the image plane.

6. Example triangulation

Consider a stereo calibration with the parameters shown in Table 1. The sensor coordinates of a point in space in the left camera are (190,187). The sensor coordinates in the right camera are (193.877,186.094). The results for triangulating the example point are given in Tables 2 and 3. The error is less than 0.02% when compared with the results from Vic3d (a commercial DIC code).

7. Definition of coordinate systems and transforms between them

The primary coordinate systems used for stereo correlation in DICe are shown in Figure 5. If the user does not specify a transformation to a particular coordinate system, the default is to use the left camera coordinate system (with the origin at the left camera sensor) as the model or physical coordinates. Alternatively, the user may define any arbitrary model coordinate system or use a best fit plane of the data points (described below).

8. Projective mapping

This section describes the projective mapping that is used to initialize the cross-correlation between the left and the right cameras. In most cases, an eight parameter model is used to characterize the projection of a sensor position in the left image, (xs^1, y_s^1) , to the corresponding sensor position in the right image, (x_s^2, y_s^2) . The projection is defined as follows

$$x_s^2 = \frac{ax_s^1 + by_s^1 + c}{gx_s^1 + hy_s^1 + 1} \quad ; \quad y_s^2 = \frac{dx_s^1 + ey_s^1 + f}{gx_s^1 + hy_s^1 + 1}$$
 (23)

The initialize the process, the AKAZE feature matching algorithm from OpenCV is used to match features from the left to the right image. These correspondences are then analyzed using an SVD process to obtain initial guesses for a, b, c, d, e, f, g, and h. These eight parameters are then optimized using a multidimensional simplex optimization where the objective function compares intensity values over the entire image from the left and right images in the first frame. In the simplex optimization objective evaluation, the entire image is treated pixel by pixel, not using subsets. The optimal solution for the parameters is the one that produces the least difference, in terms of intensity values, between the left image and the corresponding pixel positions in the right image given by the projective transform above. Whichever interpolant has been selected by the user in the correlation options is used to interpolate the right image in the optimization process.

Intrinsic param camera 0	value
c_x	182.694
c_y	184.935
f_x	2428.49
f_y	2431.72
f_s	3.39496
κ_1	0.0881161
κ_2	0.0
κ_3	0.0
Intrinsic param camera 1	value
c_x	211.986
c_y	191.741
f_x	2479.9
f_y	2484.28
f_s	2.99616
κ_1	0.112661
κ_2	0.0
κ_3	0.0
Extrinsic param	value
α	-0.66612
β	29.5027
γ	-1.37165
t_x	-3674.08
t_y	-20.9658
t_z	1040.45

Table 1. Triangulation example calibration parameter set

Code	X	Y	Z
Vic3d	46.1199	-25.5283	-6543.5
DICe	46.12	-25.52	-6543.5

Table 2. Triangulation example results given in model or world coordinates

Code	X	Y	Z
DICe	21.9899	6.01185	7310.42

TABLE 3. Triangulation example results given in terms of the left camera's coordinate system

If the images are severely distorted (for example for extremely large stereo angles), an additional twelve parameter warp model is used. This additional warp model is defined as follows:

$$x'_{s} = a_{0} + a_{1}x + a_{2}y + a_{3}x^{2} + a_{4}xy + a_{5}y^{2}$$

$$y'_{s} = a_{6} + a_{7}x + a_{8}y + a_{9}x^{2} + a_{10}xy + a_{11}y^{2}.$$
(24)

This additional warp is superimposed on the projective transform in 23 in and again uses feature matching to initialize the parameter estimation as well as simplex optimization.

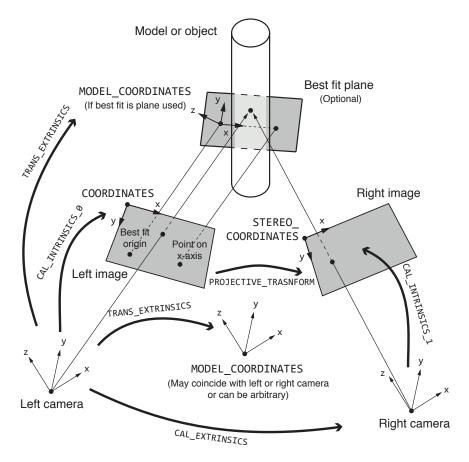


FIGURE 5. Definition of the coordinate systems and transforms used in stereo correlation in DICe.

9. Using a best fit plane for the model coordinates

If requested by the user, a best fit plane of the data points is used as the model coordinate system. The user specifies the origin of this system and a point along the x-axis in the left image. This information is used to build a coordinate system that represents the best fit plane to the three dimensional data points in space. The equation of this plane is given as

$$Z = -1(C_1X + C_2Y + C_3) (25)$$

The three coefficients, C_1 , C_2 , and C_3 are determined using a least-squares fit of the triangulated data points after an initial triangulation is performed. This sub-process proceeds as follows

- (1) The cross-correlation between the left and right images is performed.
- (2) An initial triangulation is performed for the first frame to get corresponding three dimensional points in physical coordinates. (It is assumed that the first frame represents the reference state of the object of interest.)
- (3) These points are used to determine the best fit plane.
- (4) This information is used with the origin and x-axis specified above to determine a transformation from the current physical coordinate system to the best fit plane system.
- (5) The triangulation is repeated using the new transformation to the best fit plane.
- (6) Strain calculations are completed.

To determine the plane coefficients, the following least squares problem is solved:

$$\sum \begin{bmatrix} X_i X_i & X_i Y_i & X_i \\ Y_i X_i & Y_i Y_i & Y_i \\ X_i & Y_i & 1 \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \\ C_3 \end{bmatrix} = \sum \begin{bmatrix} -X_i Z_i \\ -Y_i Z_i \\ -Z_i \end{bmatrix}$$
 (26)

where the summation is over every three dimensional point, i, from the initial triangulation. The normal vector to this plane, \mathbf{n} , is then $(C_1, C_2, 1)$. This is also the basis vector for the best fit plane, \mathbf{g}_3 . The basis vector representing the x-direction, \mathbf{g}_1 , in the best fit plane is determined as follows: The selected origin and point on the x-axis in left image sensor coordinates are projected to the right image using the projective transform above. These points are then triangulated in the original physical or model coordinate system. The X and Y component of these points are used to determine the Z positions in the best fit plane. Once these two positions are known in the best fit plane, the x-basis is simply the vector pointing between these two points. The third vector, to complete the best fit plane basis, is determined by taking the cross product of \mathbf{g}_3 with \mathbf{g}_1 .

To define the transformation from the original model coordinate system to the best fit plane system, the translations and rotations between the two systems must be determined. The translations are simply the coordinates of the origin as determined above. The rotations are given by the following relationship:

$$R_{i,j} = \cos\theta(\mathbf{e}_i, \mathbf{g}_j) \tag{27}$$

The rotation components are the cosines of the angles between the basis vectors from both coordinate systems. The basis vectors in the original model coordinate system are $\mathbf{e}_1 = (1, 0, 0), \mathbf{e}_2 = (0, 1, 0), \mathbf{e}_3 = (0, 0, 1).$

This rotations and translations are used to define a new transformation matrix in the triangulation process that maps the triangulated points onto the best fit plane. As mentioned previously, this transformation must be inverted before it is applied.

Acknowledgment

Sandia National Laboratories is a multi-mission laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

CORRESPONDENCE TO: DANIEL Z. TURNER, CENTER FOR COMPUTING RESEARCH, SANDIA NATIONAL LABORATORIES, P.O. BOX 5800; ALBUQUERQUE, NEW MEXICO 87185. *E-mail address:* dzturne@sandia.gov, *Phone:* (505) 845-7446