Fitting Spheres to Range Data From 3-D Imaging Systems

Marek Franaszek, Geraldine S. Cheok, Kamel S. Saidi, and Christoph Witzgall

Abstract—Two error functions used for nonlinear least squares (LS) fitting of spheres to range data from 3-D imaging systems are discussed: the orthogonal error function and the directional error function. Both functions allow unrestricted gradient-based minimization and were tested on more than 40 data sets[1] collected under different experimental conditions (e.g., different sphere diameters, instruments, data density, and data noise). It was found that the orthogonal error function results in two local minima and that the outcome of the optimization depends on the choice of starting point. The centroid of the data points is commonly used as the starting point for the nonlinear LS solution, but the choice of starting point is sensitive to data segmentation and, for some sparse and noisy data sets, can lead to a spurious minimum that does not correspond to the center of a real sphere. The directional error function has only one minimum; therefore, it is not sensitive to the starting point and is more suitable for applications that require fully automated sphere fitting.

Index Terms—Directional error function, orthogonal error function, sphere fitting, target-based registration, 3-D imaging systems.

I. INTRODUCTION

RANGE data from 3-D imaging systems (e.g., laser scanners) may be used to obtain detailed information about the shape and location of objects within a region of interest. Data sets acquired by these instruments are essentially 2-D range images $I(\varphi, \theta)$, where I denotes the distance from an instrument to a point on a surface of the object, while φ and θ are the azimuth and the elevation angles to that point [1]. Typically, a point cloud in the instrument coordinate frame is derived from the range image. Within such a point cloud, a match to a given 3-D model of an object is frequently sought. The model may be constructed as a collection of either geometric primitives (e.g., planes, spheres, and cylinders) or other analytically defined surfaces. The goal is to determine parameters that define a given model. Fitting methods based on the minimization of a specified error function are widely used for this purpose.

In this paper, we are interested in fitting spheres of known radii to 3-D imaging data where such data may be both sparse

[1]All data sets investigated in this paper are available upon request by contacting the first author.
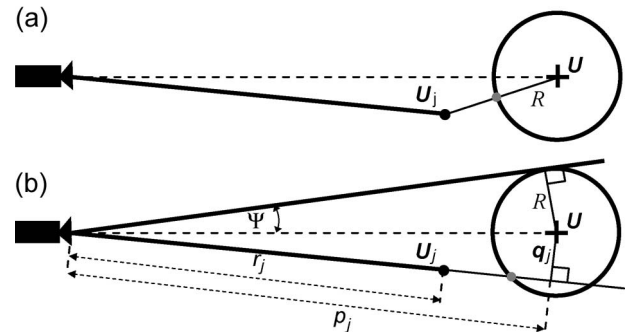

Fig. 1. Two types of error functions used in sphere fitting. Both are based on a distance between an experimental point (black dot) and its projection on a sphere surface (gray dot). (a) Orthogonal fitting—projection along normal to a surface. (b) Directional fitting—projection along scanning direction.

and noisy. Three-dimensional imaging systems are line-of-sight instruments and require multiple scans from different locations to obtain a comprehensive representation of a scene. As a result, registration of two or more data sets is often required. Spheres are traditionally used as high-precision fiducial objects for such tasks, as they can be viewed from all angles. They can be quickly identified in different scans and can roughly be segmented because they appear the same from all directions. When at least three fitted sphere centers (common to two or more data sets) are determined, the registration of these data sets is readily accomplished [2]. Sphere targets are also used when calibration is required for applications where the coordinate frame of the instrument has to be related to a given external coordinate frame representative of the ground truth. The current trend is to automate such tasks, in which case the robustness of the fitting procedure, as well as the speed of computation, is critical.

We compare the performance of two error functions used for sphere fitting: the directional and the orthogonal error function. The directional error function, which was originally introduced in [3], is based on the distance between the measured point and its projection onto a theoretical modeled sphere surface along the direction of the instrument's line-of-sight to the scanned sphere (see Fig. 1). For some sphere locations, it may not be possible to define projections of some measured points, and this limits the search space to a region in which projections exist for all measured points. To overcome this limitation, the directional error function has to be extended for measured points that cannot be projected on a sphere surface. The extension introduced in [3] resulted in a discontinuous error function. The new extension introduced in this paper ensures a continuous directional error function that is almost everywhere differentiable.

This enables the use of efficient gradient-based minimizations such as quasi-Newton. We investigate the performance of the directional error function for 47 data sets, and we compare it to the commonly used and widely commercialized orthogonal error function. We observe that the directional error function has only one minimum, while the orthogonal error function has two minima.

There are methods that purport to determine the best or "global" minimum for a specified resolution and search space [4]–[6]. Such a global minimizer, however, requires computational effort well beyond that of, e.g., a gradient-based iteration method, which yields one minimum depending on the starting point. Such global minimization methods are thus not suitable for real-time applications.

The existence of an incorrect minimum for the orthogonal sphere fitting is a consequence of the definition of the orthogonal error function and the way 3-D imaging systems collect data. These systems are line-of-sight instruments, and a single scan contains points measured only on one half of a sphere. Other classes of instruments (e.g., coordinate measurement machines) are able to provide a point cloud (measured in one coordinate frame) that may cover the whole surface of a sphere. For data sets acquired with these types of instruments, in practically all relevant experimental conditions, the orthogonal error function has only one minimum unless the measured points are collected from only one side of the sphere surface, as is the case for a single scan acquired with a 3-D imaging system. A sphere may also be fitted to range data containing two or more data sets acquired from different instrument locations and registered to one coordinate frame. In this case, the resulting data may cover the whole sphere surface, and the orthogonal error function has one minimum. Two minima are observed only when a sphere is fitted with the orthogonal function to data covering one hemisphere.

This paper is organized as follows. In Section II, the explicit expression for the sphere center derived from the algebraic fitting is briefly discussed, followed by the introduction of the orthogonal and directional error functions. Section III describes the experimental conditions under which the data were collected. Section IV contains details of the numerical calculations, and the results are presented in Section V, followed by a discussion and conclusions in Sections VI and VII, respectively. Appendix A contains the proposition and proof that all stationary points of the orthogonal error function lie within a distance less than the sphere radius from the centroid of data points. Finally, Appendix B shows that all stationary points of the directional error function lie in a confined region centered at the centroid. In addition, Appendix B shows that the location of a stationary point of the directional error function is uniquely determined only by the azimuth and the elevation angle of that point.

II. Sphere Fitting Algorithms

Fitting a sphere to a set of points has extensively been studied [7]–[16]. A sphere of radius R and center $U = [X, Y, Z]$ can be fitted to measured points $U_j = [X_j, Y_j, Z_j]$, where $j = 1, \ldots, N$, by minimizing a properly defined error function.

Two main strategies have been developed: algebraic fitting and geometric fitting [10], [17]. The former is based on the algebraic equation of a sphere

$$x^2 + y^2 + z^2 + ax + by + cz + d = 0 \qquad (1)$$

where the coefficients of the equation a, b, c, and d are related to the sphere center and radius by

$$U = -0.5[a, b, c] \qquad (2a)$$
$$R^2 = 0.25(a^2 + b^2 + c^2) - d. \qquad (2b)$$

In [3], it was shown that the right side of (2b) is always positive, and therefore, the existence of a positive radius R is guaranteed.

The determination of the unknown radius R and sphere center U can be accomplished by minimizing the following algebraic error function:

$$ErA(a, b, c, d) = (1/N) \sum_{j=1}^{N} \Delta_j^2 \qquad (3a)$$

where

$$\Delta_j(a, b, c, d) = X_j^2 + Y_j^2 + Z_j^2 + aX_j + bY_j + cZ_j + d. \qquad (3b)$$

At a minimum, the gradient of the error function has to be zero, and because the jth deviation Δ_j is linear in its parameters, i.e., a, b, c, and d, the location of the minimum can analytically be calculated. This is why this method falls under the category of linear least squares (LS). This linear LS approach can be used only in a 4-D search space where the sphere radius R is unknown and needs to be treated in the same way as the unknown sphere center U. When the radius R is known, the gradient of the algebraic error function (3a) depends on parameters a, b, and c in a nonlinear relationship, so nonlinear LS optimization is needed [10].

In the following sections, we discuss two different error functions that can be used in nonlinear LS fitting. Both of them are differentiable, so the standard quasi-Newton optimization can be applied.

A. Orthogonal Fitting

For orthogonal fitting, the error function is based on the distance between the measured point U_j and its orthogonal projection onto a sphere surface, as shown in Fig. 1(a), and can be written as

$$
\begin{aligned}
&ErO(X, Y, Z) \\
&= (1/N) \sum_{j=1}^{N} \left(\sqrt{(X - X_j)^2 + (Y - Y_j)^2 + (Z - Z_j)^2} - R \right)^2
\end{aligned}
\qquad (4)
$$

and the gradient of ErO is

$$
\begin{aligned}
&\nabla ErO(X, Y, Z) \\
&= (1/N) \sum_{j=1}^{N} f_j(X, Y, Z)[X - X_j, Y - Y_j, Z - Z_j]
\end{aligned}
\qquad (5)
$$

where

$$f_j(X, Y, Z)$$
$$= 2\left(1 - \frac{R}{\sqrt{(X-X_j)^2 + (Y-Y_j)^2 + (Z-Z_j)^2}}\right). \quad (6)$$

When $U = U_j$, the jth term in (5) does not have a value (the limit of the jth term depends on how U approaches U_j). In any arbitrarily small neighborhood of U, the jth term remains bounded. If a data set does not contain two identical points $U_k = U_j$ for any $k \neq j$, there can only be one term with a singular behavior at a time, and the contribution of all other terms in (5) remains continuous. In the numerical implementation, we set the jth term in (5) to zero when $U = U_j$.

B. Directional Fitting

For directional fitting, the error function is based on the distance between the measured point U_j and its projection onto the surface of the sphere closer to the instrument along the direction of U_j [see Fig. 1(b)]. The problem with this definition of the error function is that the projected point may not exist, but the function must still be defined and has to smoothly vary. To design such a function, we introduce the following notation: $u_j = [x_j, y_j, z_j]$, $u_j = U_j/r_j$, where r_j is the length of the vector U_j. We then define the dot and the cross products $p_j = u_j \circ U$ and $q_j = u_j \times U$, respectively. The length of the vector q_j, $q_j = \|q_j\|$, is the distance from the sphere center to the line defined by the unit vector u_j. In Cartesian coordinates, the expression for p_j and q_j (see Fig. 1) can be written as

$$p_j(X, Y, Z) = Xx_j + Yy_j + Zz_j \quad (7)$$

and

$$q_j(X, Y, Z)$$
$$= \sqrt{(y_j Z - z_j Y)^2 + (z_j X - x_j Z)^2 + (x_j Y - y_j X)^2}. \quad (8)$$

The directional error function ErD may now be expressed as

$$ErD(X, Y, Z) = (1/N) \sum_{j=1}^{N} E_j(X, Y, Z) \quad (9)$$

where

$$E_j(X, Y, Z) = \begin{cases} \left(p_j - \sqrt{R^2 - q_j^2} - r_j\right)^2, & \text{if } q_j < R \\ (p_j - r_j)^2 + (q_j - R)^2, & \text{if } q_j \geq R. \end{cases} \quad (10)$$

Let us note that every E_j is a continuous and increasing function as the value of q_j varies from less than R to greater than R. This corresponds to the transition between the two configurations where U_j can and cannot be projected on a sphere surface along the U_j direction. Thus, the whole directional error function ErD is also continuous, and it allows for

the unconstrained search for the minimum. The corresponding gradient may be evaluated as

$$\nabla ErD(X, Y, Z) = (1/N) \sum_{j=1}^{N} \left[\frac{\partial E_j}{\partial X}, \frac{\partial E_j}{\partial Y}, \frac{\partial E_j}{\partial Z}\right] \quad (11)$$

where the individual derivatives can be calculated using the chain rule as follows:

$$\frac{\partial E_j}{\partial X} = \frac{\partial E_j}{\partial p_j}\frac{\partial p_j}{\partial X} + \frac{\partial E_j}{\partial q_j}\frac{\partial q_j}{\partial X}$$
$$\frac{\partial E_j}{\partial Y} = \frac{\partial E_j}{\partial p_j}\frac{\partial p_j}{\partial Y} + \frac{\partial E_j}{\partial q_j}\frac{\partial q_j}{\partial Y}$$
$$\frac{\partial E_j}{\partial Z} = \frac{\partial E_j}{\partial p_j}\frac{\partial p_j}{\partial Z} + \frac{\partial E_j}{\partial q_j}\frac{\partial q_j}{\partial Z}. \quad (12)$$

Using the definitions of p_j and q_j and the Lagrange formula for the triple cross product, the explicit equation for derivatives can be written as

$$\frac{\partial E_j}{\partial X} = \begin{cases} 2\left(p_j - \sqrt{R^2 - q_j^2} - r_j\right) \\ \quad \times \left(x_j + \frac{X - x_j p_j}{\sqrt{R^2 - q_j^2}}\right), & \text{if } q_j < R \quad (13a) \\ 2(p_j - r_j)x_j + 2(q_j - R)\frac{X - x_j p_j}{q_j}, & \text{if } q_j \geq R. \end{cases}$$

In the same way, the derivatives of E_j with respect to Y and Z can be calculated, yielding similar equations

$$\frac{\partial E_j}{\partial Y} = \begin{cases} 2(p_j - \sqrt{R^2 - q_j^2} - r_j) \\ \quad \times \left(y_j + \frac{Y - y_j p_j}{\sqrt{R^2 - q_j^2}}\right), & \text{if } q_j < R \\ 2(p_j - r_j)y_j + 2(q_j - R)\frac{Y - y_j p_j}{q_j}, & \text{if } q_j \geq R \end{cases}$$
$$\quad (13b)$$

$$\frac{\partial E_j}{\partial Z} = \begin{cases} 2\left(p_j - \sqrt{R^2 - q_j^2} - r_j\right) \\ \quad \times \left(z_j + \frac{Z - z_j p_j}{\sqrt{R^2 - q_j^2}}\right), & \text{if } q_j < R \\ 2(p_j - r_j)z_j + 2(q_j - R)\frac{Z - z_j p_j}{q_j}, & \text{if } q_j \geq R. \end{cases}$$
$$\quad (13c)$$

Contrary to the orthogonal fitting method, the jth contribution to a gradient in (11) has a well-defined value for any location of the sphere center U. Even when $q_j = R$, the corresponding jth term in (11) has a finite value $2(p_j - r_j)[x_j, y_j, z_j]$. When, however, $q_j \to R_{(-)}$, the jth term in (11) can have an arbitrarily large value, which may happen not just for one point U_j but simultaneously for many U_k points for which $q_k \to R_{(-)}$. This, in turn, may result in a discontinuity in the gradient for the directional error function for some isolated locations of the sphere center U. In the numerical implementation, this property does not adversely affect the convergence of the minimization. This statement is based on two observations. First, as q_j approaches R from below, the corresponding jth term in (11) shows a rather weak divergence of type $\varepsilon^{-1/2}$, with $\varepsilon \to 0$. Second, for most data sets, due to the symmetry of a sphere, many terms with large absolute values in the sum (11) have opposite signs, and therefore, their net contribution to the gradient will be moderate. This subject is further discussed in Section VI.
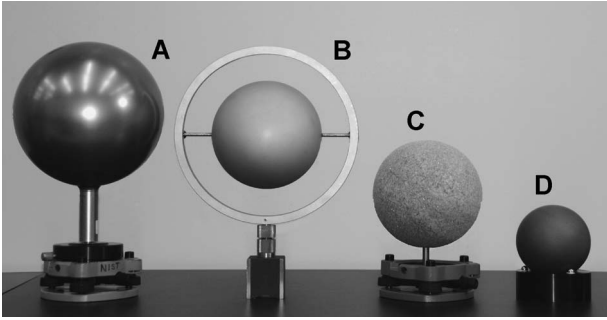
Fig. 2. Four spheres used in the experiments.

III. EXPERIMENT

Four different spheres were scanned with three different instruments at various distances (between 5 and 100 m). The three instruments used in the experiments fall into two categories. Instrument $In1$ is the most accurate of the three instruments and has a maximum range of 24 m and a manufacturer-specified measurement error of about 100 μm. This instrument is typically used in indoor applications where highly accurate measurements are required (e.g., in manufacturing facilities). The other two instruments, i.e., $In2$ and $In3$, have longer maximum ranges and larger manufacturer-specified range errors. The maximum ranges of these instruments are greater than 100 m, and the range errors are between 7 and 10 mm. They are used in both indoor and outdoor applications (e.g., at construction sites).

The four spheres used in the experiments are made of different materials, and they have different sizes and surface finishes (see Fig. 2). The radii for spheres A, B, C, and D were $R_A = 101.6$ mm, $R_B = 76.2$ mm, $R_C = 76.2$ mm, and $R_D = 50.8$ mm, respectively. Spheres A and B are made of anodized aluminum, sphere C is made of styrofoam, and sphere D is made of titanium. The surface of sphere C is rough compared to that of the other spheres.

The combination of the three instruments, four spheres, and many scanning distances resulted in 47 data sets, varying from very noisy to very clean. All scans were collected indoors under controlled conditions. The sphere in each acquired data set was manually segmented, with the number of points in each segmented point cloud varying from a few hundreds to a few thousands.

IV. NUMERICAL CALCULATIONS

Each segmented data set contained the Cartesian coordinates of N points. The process began with fitting a sphere of known radius R to the data set using the orthogonal error function. The centroid of the point cloud $\boldsymbol{U}_0$ was selected as the starting point for the minimization. Once the location of a minimum $\boldsymbol{U}_c = [X_c, Y_c, Z_c]$ was found, its spherical coordinates $[D_c, \Phi_c, \Theta_c]$ were determined, where distance $D_c = \|\boldsymbol{U}_c\|$, and Φ_c and Θ_c are the azimuth and the elevation angle of the fitted sphere center. Next, the angle Ψ, which is depicted in Fig. 1(b), was calculated using $\sin \Psi = R/D_c$ and the error function $ErO(\boldsymbol{U})$ for the grid of points $\boldsymbol{U} = [D, \Phi, \Theta]$, centered on $[D_c, \Phi_c, \Theta_c]$. The magnitudes of the gradients were also calculated on the grid.
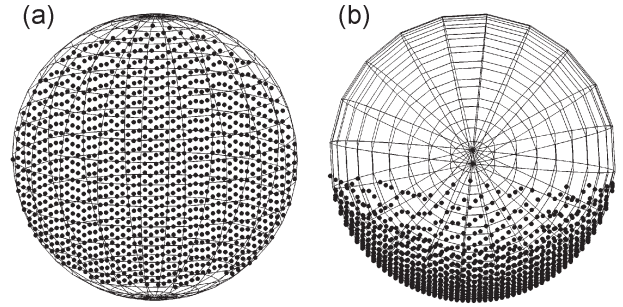

Fig. 3. Point cloud $P1$ (number of points $N = 1176$). (a) View from the instrument toward the fitted sphere center. (b) View from the top.
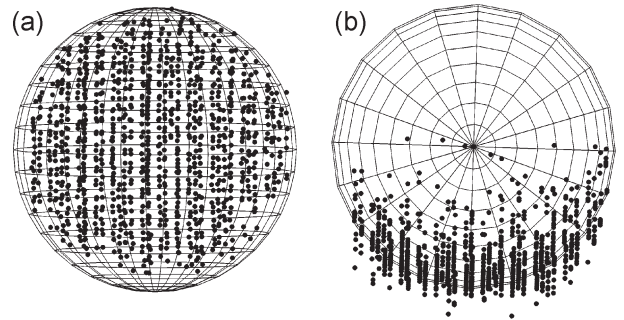

Fig. 4. Point cloud $P2$ (number of points $N = 1096$). (a) View from the instrument toward the fitted sphere center. (b) View from the top.
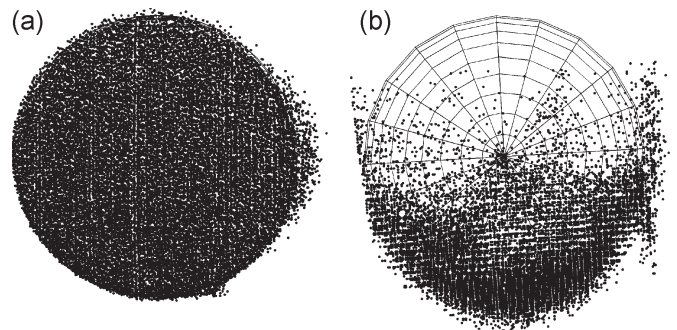

Fig. 5. Point cloud $P3$ (number of points $N = 20\,112$). (a) View from the instrument toward the fitted sphere center. (b) View from the top.

The whole procedure was then repeated with the directional error function. For both error functions, the same quasi-Newton minimization procedure was used [18]. Exit conditions for the optimization process were defined by two parameters, i.e., the relative length of the step and the relative gradient magnitude, both of which were set to 10^{-7}. All calculations were performed to double precision on a 32-bit computer.

V. RESULTS

Figs. 3–5 show examples of typical data sets used for testing the behavior of the two error functions examined in this paper and also show the fitted spheres. Point cloud $P1$ in Fig. 3 was acquired by scanning sphere D from a distance of about 11 m with instrument $In1$, point cloud $P2$ in Fig. 4 was obtained by scanning sphere B from a distance of 100 m with instrument $In2$, and data set $P3$ in Fig. 5 was obtained by scanning sphere A from a distance of about 6 m with instrument $In3$.
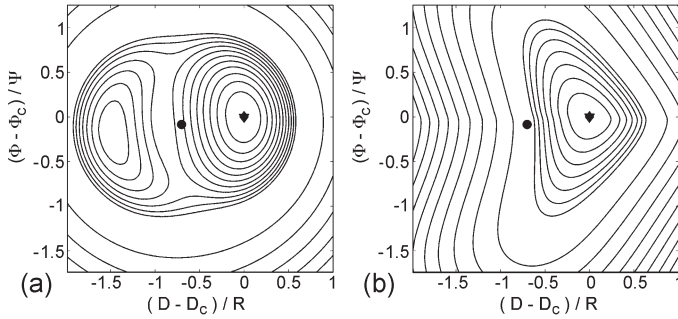
Fig. 6. Contour plots of two error functions for data set $P1$. (a) Orthogonal error. (b) Directional error. For this data set, the linear LS solution coincides with the nonlinear LS for both error functions.
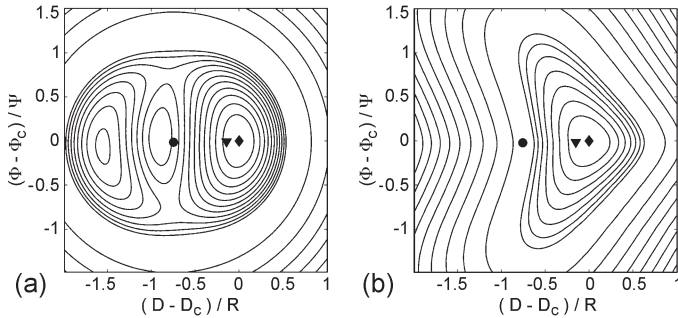

Fig. 7. Contour plots of two error functions for data set $P2$. (a) Orthogonal. (b) Directional. For this case, the linear LS solution (triangle) is close to the nonlinear LS (diamond).
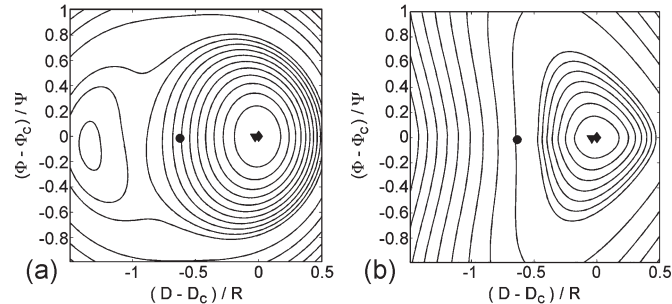

Fig. 8. Contour plots of two error functions for data set $P3$. (a) Orthogonal. (b) Directional. For this case, the linear LS solution almost coincides with the nonlinear LS.

Figs. 6–8 show contour plots of both error functions $ErO(D, \Phi, \Theta = \Theta_c)$ and $ErD(D, \Phi, \Theta = \Theta_c)$ for the data sets shown in Figs. 3–5. In these figures, the dot represents the starting point (centroid of the experimental points $\boldsymbol{U}_0$), the diamond represents the end point of optimization, and the triangle represents the algebraic solution (linear LS). Both axes of each error plot are normalized so that $(0, 0)$ corresponds to the fitted sphere center $\boldsymbol{U}_c$ indicated by a diamond. For visualization purposes, we keep the elevation angle fixed ($\Theta = \Theta_c$) and vary D and Φ such that $|D_c - D| < \alpha R$ and $|\Phi_c - \Phi| < \alpha \Psi$, where α is a dimensionless constant that defines the span of the grid points around the fitted sphere center $\boldsymbol{U}_c$. To enhance the visualization and to show the region close to the fitted center, the contour lines are not equally spaced. In Appendix A, we prove that all stationary points (including all minima) of ErO must lie within a sphere of radius R centered at $\boldsymbol{U}_0$. In Appendix B, we show that, for most practical applications
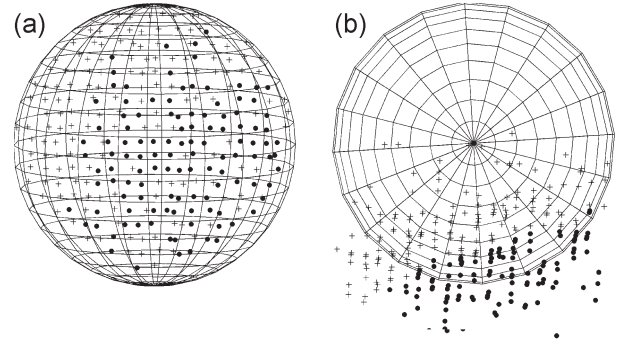

Fig. 9. Point cloud $P4$ (marked by both dots and crosses, total number of points $N = 276$). (a) View from the scanner toward the fitted sphere center. (b) View from the top. Dots only correspond to subset $P4s$ ($N = 117$).
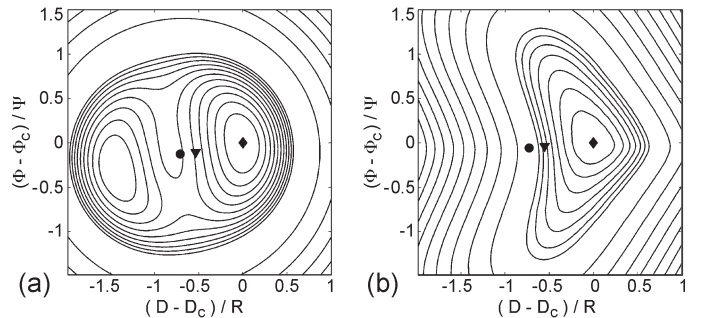

Fig. 10. Contour plots of two error functions for data set $P4$. (a) Orthogonal. (b) Directional. For the included data set, the linear LS solution (triangle) is much closer to the centroid (circle) than to the nonlinear LS solution (diamond).

(where the sphere distance to the instrument is much larger than R), all stationary points of ErD must be within a bounded region centered at $\boldsymbol{U}_0$. Therefore, it is sufficient to plot contour plots of both error functions in the bounded region only.

Contour plots for the orthogonal error function in Figs. 6(a), 7(a), and 8(a) reveal the existence of two local minima: one marked with a diamond at $(0, 0)$ and the second (not shown) located at approximately $(-1.5, -0.1)$. The first minimum (referred hereafter as $\boldsymbol{U}_G$) is the correct one because it corresponds to the location of a real sphere. The second one (referred hereafter as $\boldsymbol{U}_W$) is the wrong one because $\|\boldsymbol{U}_W\|$ is systematically less than a distance from the instrument to the real sphere center. This spurious minimum $\boldsymbol{U}_W$ is shallower than the correct minimum at $\boldsymbol{U}_G$, and the optimization may converge to it when a point that lies in the basin of attraction of $\boldsymbol{U}_W$ is chosen as the starting point. The LS solution (marked by a triangle) is close to $\boldsymbol{U}_G$ if it is a good approximation of a real sphere center.

Fig. 9 shows a point cloud, i.e., $P4$, acquired by scanning sphere C from a distance of 6 m with instrument $In3$. Two different symbols are used in the figure to differentiate between the full data set $P4$ (both dots and crosses) and the subset $P4s$ (dots only), which was created by deleting points $\boldsymbol{U}_j$ that have a large angular distance to $\boldsymbol{U}_c$. Figs. 10 and 11 show the error plots for both error functions for the full data set $P4$ and the subset $P4s$.

Fig. 12 shows another full point cloud, i.e., $P5$, which was acquired by scanning the sphere D from a distance of 73 m using instrument $In2$, and its subset, i.e., $P5s$, was created in a
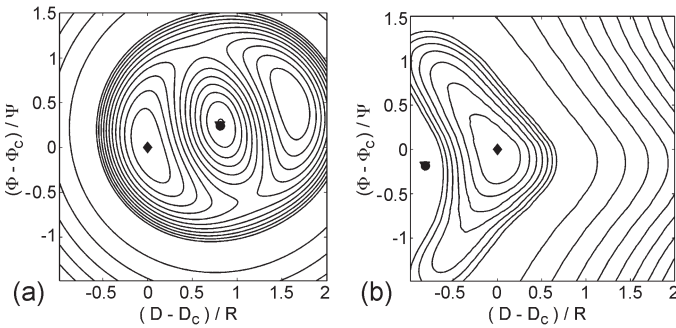
Fig. 11. Contour plots of two error functions for data set $P4s$. (a) Orthogonal. (b) Directional. For the included data set, the linear LS solution (triangle) coincides with the centroid U_0 (circle). Selecting either of them as a starting point for nonlinear LS minimization yields a wrong solution for the orthogonal error function (diamond) in (a).
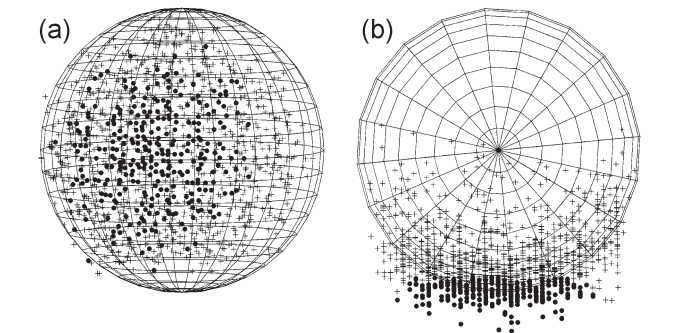

Fig. 12. Point cloud $P5$ (marked by both dots and crosses, total number of points $N = 1268$). (a) View from the scanner toward the fitted sphere center. (b) View from the top. Dots only correspond to subset $P5s$ ($N = 321$).
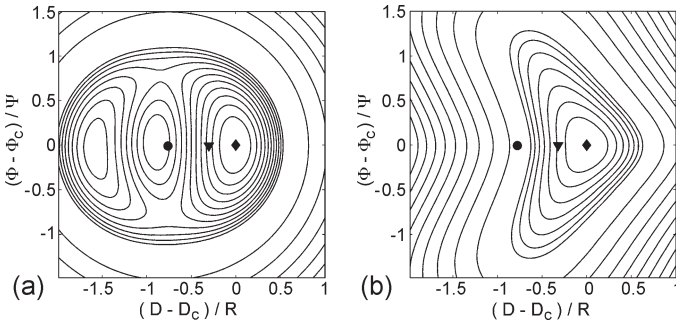

Fig. 13. Contour plots of two error functions for data set $P5$. (a) Orthogonal. (b) Directional. For this data set, the linear LS solution (triangle) is located approximately halfway between the centroid (circle) and the nonlinear LS solution (diamond).

similar way as $P4s$. Figs. 13 and 14 show error function plots corresponding to the data set $P5$ and the subset $P5s$.

In Figs. 10(a) and 13(a), a diamond indicating the fitted sphere center U_c is located at the right minimum U_G, while in Figs. 11(a) and 14(a), a diamond is located at the wrong minimum U_W. For the data shown in Figs. 11(a) and 14(a), the LS solution marked by a triangle is a bad approximation of the real sphere position.

The actual location of the fitted sphere center U_c depends on the error function used in the optimization, the starting point, and the data set (e.g., full or subset). The last column in Table I contains the normalized distance between two sphere centers $\Delta = \|U_c(1) - U_c(2)\|/R$, where $U_c(1)$ is the fitted sphere center for a given error function, starting point, and
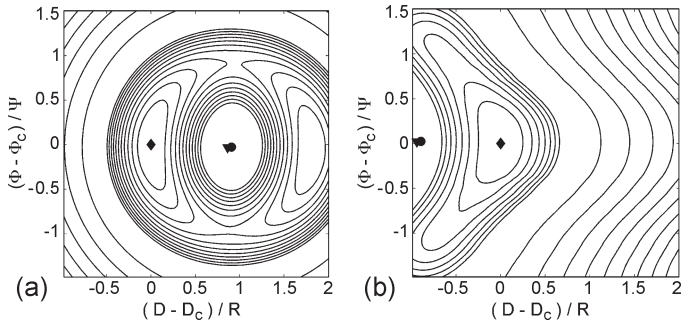

Fig. 14. Contour plots of two error functions for data set $P5s$. (a) Orthogonal. (b) Directional. For this data set, the linear LS solution (triangle) coincides with the centroid (circle). Selecting either of them as a starting point for nonlinear LS minimization yields a wrong solution for the orthogonal error function.

TABLE I
EFFECT OF THE ERROR FUNCTION, STARTING POINT,
AND DATA SET ON THE FITTED SPHERE CENTERS

| Case | Run 1 | | Run 2 | | |
	Dataset	Minimum 1	Dataset	Minimum 2	Δ [%]
Constant: Dataset, error function Variable: starting point					
1	P1	O+	P1	O−	149.0
2	P2	O+	P2	O−	156.4
3	P3	O+	P3	O−	135.2
4	P4	O+	P4	O−	148.1
5	P4s	O+	P4s	O−	169.8
6	P5	O+	P5	O−	157.9
7	P5s	O+	P5s	O−	181.4
Constant: Dataset, starting point Variable: Error function					
8	P1	O+	P1	D	0.05
9	P2	O+	P2	D	1.27
10	P3	O+	P3	D	1.40
11	P4	O+	P4	D	7.12
12	P4s	O+	P4s	D	9.70
13	P5	O+	P5	D	1.93
14	P5s	O+	P5s	D	1.83
Constant: Error function, starting point Variable: Dataset					
15	P4	O+	P4s	O+	9.77
16	P4	D	P4s	D	9.70
17	P5	O+	P5s	O+	8.85
18	P5	D	P5s	D	11.65
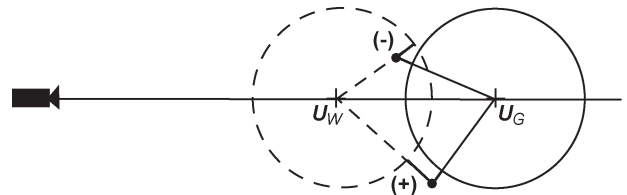
Fig. 15. Schema explaining the existence of two minima for the orthogonal error function: the solid circle centered at U_G is the correct solution; the dashed circle centered at U_W is the wrong one.

data set (Run 1 in Table I), while $U_c(2)$ is the fitted sphere center where one of the three input parameters used for $U_c(1)$ is changed (Run 2 in Table I). In the table, an O+ indicates that the minimization located the correct minimum U_G from the orthogonal error function. An O− indicates that the wrong minimum U_W from the orthogonal error function was located, and a D indicates that the directional error function located the minimum (see Fig. 15).

VI. Discussion

The most striking difference between the two error functions is the number of minima: one for the directional error function and two for the orthogonal error function. For the orthogonal error function, the correct minimum U_G and the wrong one at U_W are separated by an approximate distance of $1.5R$ (Table I, rows 1–7). The presence of a wrong minimum U_W is not surprising because the orthogonal error function does not differentiate between fitting experimental points to the front and fitting experimental points to the back of the sphere (alternatively, one may say that the orthogonal error does not differentiate between points on a convex and those on a concave surface, i.e., outer and inner parts of a sphere surface). Fig. 15 illustrates that, as the center of the dashed circle moves toward the instrument, an increasing number of experimental points will have a distance to this center greater than R. Similarly, as the center of the dashed circle moves away from the instrument, an increasing number of points will have a distance less than R. Due to the continuity of the error function, somewhere between the two extremes there will be a point U_W where contributions from the negative and positive terms in (5) are balanced, yielding a zero gradient. Thus, the result of fitting a sphere to range data depends on the location of the starting point when the orthogonal error function is used.

The directional error function, in contrast, is different when fitting the front or back of a sphere (i.e., concave or convex parts of a sphere surface). It is based on the distance of the measured points to the sphere along the scanning direction, and therefore, the function has only one minimum. The results for the variety of data sets used in this paper support the conclusion that the observed pattern is general and does not depend on the data noise level or sphere radius.

As stated in the introduction, the current trend is to automate sphere fitting for target-based registration of range images; therefore, the selection of the starting point for minimization should automatically be performed, such as by using the centroid of the points U_0. Plots shown in Figs. 11(a) and 14(a) indicate that U_0 may be a wrong starting point, which leads to an incorrect minimum when the orthogonal error function is used. Centroid coordinates depend on the experimental points U_j, and therefore, the segmentation of data for sphere fitting should be done with great caution. This requirement is particularly important for sparse and noisy data sets. Fig. 11(a) shows that U_0 is located very close to the local maximum where the gradient magnitude is almost zero. This makes the optimization sensitive to numerical error: by moving the starting point to a new location equal to $1.0005U_0$, orthogonal fitting yields a different solution (to the right of U_0). Figs. 9–14 clearly demonstrate that discarding too many points may lead to the wrong result when the orthogonal error function is used. Another automatic method to select the starting point is to use the sphere center estimated by the algebraic fit (3), which is believed to give a better starting point; however, this may also lead to the wrong minima for sparse and noisy data sets. Because the directional error function has only one minimum, it is not as sensitive to the segmentation and choice of starting point.

Rows 8–14 in Table I show that both orthogonal and directional error functions give similar results when applied to the same data set (assuming that the orthogonal fitting converges to the right minimum). Fitting a sphere to a full data set and then to its subset may lead to slightly different results, even for the same error function. The distance between the two locations provides a rough estimate of how sensitive the fitted sphere center is to variations in a data set. In this context, there is no clear advantage of one error function over the other (assuming that the right minimum is reached in the orthogonal fitting). Rows 15 and 16 in Table I show that a shift in the fitted sphere centers is smaller for directional fitting than for orthogonal fitting for one pair of data sets, and the opposite is true for another pair of data sets, as shown in the two bottom rows of Table I.

The overall behavior of the directional gradient is not affected by the singular behavior of the jth term in (11) when $q_j \rightarrow R_{(-)}$. The surface plots in Fig. 16 show that the generally smooth changes in the gradient magnitude are only locally disturbed by spikes, corresponding to the locations of the sphere center U for which $q_j \approx R$, and thus, an arbitrarily large value of an individual jth term in (13) is expected for $q_j < R$. As mentioned earlier, these local discontinuities do not hamper the convergence of the minimization. Fig. 16(a) and the enlargement of neighborhood of the fitted sphere center U_c shown in Fig. 16(b) reveal two important features. First, the magnitudes of the spikes are relatively small compared to the rest of the surface; the value at a spike is only a few times larger than the gradient magnitude in the immediate neighborhood of the spike. This supports our earlier observation that, for most data sets, the large terms in (11) cancel each other and that the resulting sum remains small. Second, in the magnified region around the fitted sphere center U_c in Fig. 16(b), the regions where the directional gradient is discontinuous are less frequent. In the vicinity close to U_c, the directional gradient is smooth. This means that, when a minimization process converges to the final solution U_c, the exit condition based on a small value of the gradient magnitude remains valid.

It is worth noting that, in addition to two minima, the orthogonal error function also has one local maximum and two saddle points [see Fig. 7(a) for an example]. In all, there are five stationary points where the gradient of the orthogonal function has a zero length, all of them lying within less than R from the centroid U_0 (see Appendix A). Robust minimization methods will generally avoid local maxima, but saddle points are always potential traps. The directional error function is not prone to this problem because it has only one stationary point corresponding to the global minimum.

VII. Conclusion

Two different error functions used in fitting a sphere to 3-D imaging data have been investigated. The directional error function has only one minimum, and therefore, the choice of the starting point for minimization is not critical. This finding is based on the analysis of 47 experimental data sets that were collected under a range of experimental conditions. Although there is no analytical proof that the directional error function
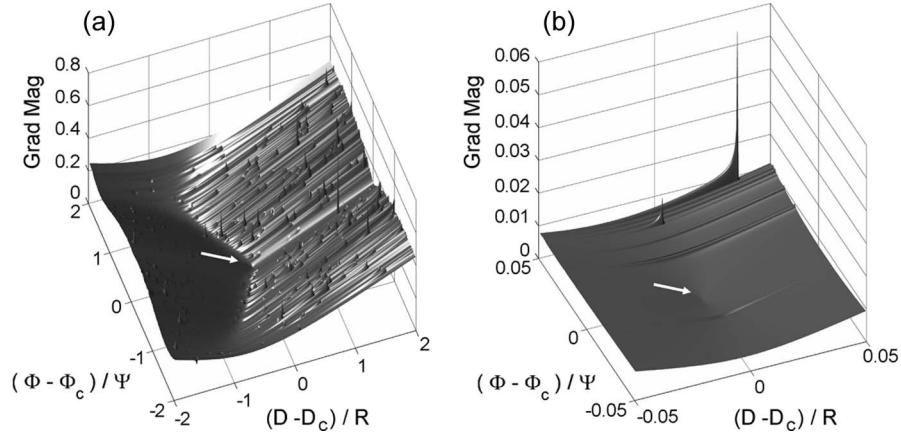
Fig. 16. Gradient magnitude of the directional error function versus range and azimuth for data set $P2$. (a) Neighborhood around a minimum U_c (indicated by the arrow). (b) Zoomed-in view around the minimum.

has only one minimum for an arbitrary data set, the theoretical arguments provided in Appendix B for certain situations (large distances between the sphere and the instrument) support our finding. The orthogonal error function has two minima, and the result of the minimization depends on the choice of the starting point. The common practice of using the centroid U_0 of a data set as the starting point may lead to the wrong solution when using the orthogonal error function. For sphere fitting, a wrong minimum may sometimes be avoided if the starting point is appropriately selected by choosing $U_{\rm ini} = (1 + 2R/\|U_0\|)U_0$; however, for partially occluded spheres (where U_0 is not parallel to U_c) or for other objects modeled by a general analytical surface (particularly those with a large radius of curvature—e.g., parabolic satellite dishes), the right choice of starting point remains an open problem when the orthogonal error function is used. The directional error function is dependent on the distances of the measured points from the front side of a scanned object along the instrument's line of sight, and therefore, this function has only one minimum. Its location is very close to the location of the correct minimum from the orthogonal error function, and both error functions result in equally fast convergence of optimization; therefore, the use of the directional error function for fitting a sphere to 3-D imaging data is more advantageous as it eliminates the problem of an incorrectly selected starting point.

APPENDIX A

A proposition referred to in the main text concerning orthogonal fitting based on LS will be proved. The error function (4) can be written as

$$ErO(X,Y,Z) = (1/N) \sum_{j=1}^{N} (d_j - R)^2 \qquad (A1)$$

where $d_j = \|U_j - U\|$ is the Euclidean distance between data point U_j and the sphere center U

$$d_j = \sqrt{(X - X_j)^2 + (Y - Y_j)^2 + (Z - Z_j)^2}. \qquad (A2)$$

Proposition: All stationary points and, particularly, all minima of the error function $ErO(X,Y,Z)$ are within a distance less than the radius R from the centroid $U_0 = [X_0, Y_0, Z_0]$ of the data points U_j.

Proof: Recall that $d_j \neq 0$, and from expressions for gradient (5) and (6)

$$\frac{\partial ErO}{\partial X} = X - X_0 - (R/N) \sum_{j=1}^{N} \frac{X - X_j}{d_j}$$

$$\frac{\partial ErO}{\partial Y} = Y - Y_0 - (R/N) \sum_{j=1}^{N} \frac{Y - Y_j}{d_j}$$

$$\frac{\partial ErO}{\partial Z} = Z - Z_0 - (R/N) \sum_{j=1}^{N} \frac{Z - Z_j}{d_j} \qquad (A3)$$

are the conditions for a stationary point U of ErO. For such a point, it follows that

$$\|U_0 - U\|^2 = (R^2/N^2) \left[\left(\sum_{j=1}^{N} \frac{X - X_j}{d_j} \right)^2 \right.$$

$$\left. + \left(\sum_{j=1}^{N} \frac{Y - Y_j}{d_j} \right)^2 + \left(\sum_{j=1}^{N} \frac{Z - Z_j}{d_j} \right)^2 \right] \qquad (A4)$$

where $\|U_0 - U\|$ is the Euclidean distance between the centroid U_0 and a stationary point U. The right side of (A4) can be written as a double sum

$$\|U_0 - U\|^2$$

$$= (R^2/N^2) \sum_{i=1}^{N} \sum_{j=1}^{N} \left[\left(\frac{X - X_i}{d_i} \right) \left(\frac{X - X_j}{d_j} \right) \right.$$

$$+ \left(\frac{Y - Y_i}{d_i} \right) \left(\frac{Y - Y_j}{d_j} \right)$$

$$\left. + \left(\frac{Z - Z_i}{d_i} \right) \left(\frac{Z - Z_j}{d_j} \right) \right]. \qquad (A5)$$

Each term in the square brackets is the dot product of two unit vectors $\boldsymbol{V}_i \circ \boldsymbol{V}_j$, where $\boldsymbol{V}_i = (\boldsymbol{U}_i - \boldsymbol{U})/\|\boldsymbol{U}_i - \boldsymbol{U}\|$, and that for $\boldsymbol{V}_j$ is given in a similar manner. Therefore, the right side of (A5) represents the double sum of N^2 cosines

$$\|\boldsymbol{U}_0 - \boldsymbol{U}\|^2 = (R^2/N^2) \sum_{i=1}^{N} \sum_{j=1}^{N} \cos \tau_{i,j} \qquad \text{(A6)}$$

where $\tau_{i,j}$ is the angle between $\boldsymbol{V}_i$ and $\boldsymbol{V}_j$. Since only those terms for which $i = j$ actually reach the value 1, this establishes the strict bound of N^2 on the double sum

$$\sum_{i=1}^{N} \sum_{j=1}^{N} \cos \tau_{i,j} < N^2. \qquad \text{(A7)}$$

Thus

$$\|\boldsymbol{U}_0 - \boldsymbol{U}\| < R \qquad \text{(A8)}$$

which proves the proposition for all stationary points $\boldsymbol{U}$, including those that are minima. On the other hand, a minimum may not occur at an actual data location, that is, if $d_j = 0$. This is because the corresponding error term $ErO_j(\boldsymbol{U}) = (d_j - R)^2/N$ represents an upward-oriented cusp centered at $\boldsymbol{U}_j$, which forces the descent direction—incompatible with a minimum—of the full error function ErO.

APPENDIX B

A sphere used as a target for registration is usually placed in a scanned scene in such a way that the distance D_c between the sphere center and an instrument is much larger than radius R. Then, the ratio R/D_c and the angle Ψ, which is depicted in Fig. 1(b), are very small, and all vectors $\boldsymbol{U}_j$ can be considered as parallel. (For all data sets in this paper, the largest ratio R/D_c and the corresponding Ψ were less than 0.0163). In such approximation, without lost of generality, the Z-axis can be aligned with the direction of the centroid $\boldsymbol{U}_0$ (calculated in the instrument's coordinates frame), while X and Y are coordinates on the plane passing through the instrument and perpendicular to $\boldsymbol{U}_0$. Then, the directional error function $ErD(X, Y, Z)$ defined by (9) and (10) can be rewritten as

$$ErD_p(X, Y, Z) = (1/N) \left\{ \sum_{j=1}^{J} (Z - r_j - \sqrt{R^2 - z_j^2})^2 \right.$$
$$\left. + \sum_{k=1}^{K} \left[(Z - r_k)^2 + (Q_k - R)^2 \right] \right\} \qquad \text{(B1)}$$

where J is a number of points for which $Q_j < R$, and K is the number of points for which $Q_j \geq R$. The actual values of K and J depend on X and Y: for some (X, Y), either K or J (but never both) can be zero. Both numbers satisfy the normalization condition $J + K = N$. The function Q_j does not depend on Z

$$Q_j(X, Y) = \sqrt{(X - X_j)^2 + (Y - Y_j)^2} \qquad \text{(B2)}$$

where X_j and Y_j are coordinates of experimental points expressed in the rotated coordinate frame, with the Z-axis aligned to $\boldsymbol{U}_0$. From (B1), the partial derivative of ErD_p with respect to Z can be calculated as

$$\frac{\partial ErD_p}{\partial Z}$$
$$= (2/N) \left[\sum_{j=1}^{J} \left(Z - r_j - \sqrt{R^2 - Q_j^2} \right) + \sum_{k=1}^{K} (Z - r_k) \right].$$
$$\text{(B3)}$$

At any stationary point, including all minima, the gradient of the directional error function has to be zero. Applying this condition to the Z-coordinate of the gradient given by (B3), the following holds for any X and Y:

$$NZ - \sum_{n=1}^{N} r_n - \sum_{j=1}^{J} \sqrt{R^2 - Q_j^2} = 0. \qquad \text{(B4)}$$

Since this equation is linear in Z, its explicit solution $Z^* = Z^*(X, Y)$ can be obtained

$$Z^*(X, Y) = R_0 + (1/N) \sum_{j=1}^{J} \sqrt{R^2 - Q_j^2} \qquad \text{(B5)}$$

where

$$R_0 = (1/N) \sum_{n=1}^{N} r_n \qquad \text{(B6)}$$

is the mean range, which does not depend on (X, Y). Thus, any stationary point of the directional error function must be located at $U^* = [X^*, Y^*, Z^*(X^*, Y^*)]$, where

$$\frac{\partial ErD_p}{\partial X}(X^*, Y^*, Z^*) = 0 \quad \frac{\partial ErD_p}{\partial Y}(X^*, Y^*, Z^*) = 0. \quad \text{(B7)}$$

It should be noted that the search space for the directional error function is also confined (similar to the orthogonal error function discussed in Appendix A). Every Q_j in (B5) is restricted to the $[0, R]$ interval, and therefore, the sum over index j is also bounded. This leads to the following restriction on Z^*:

$$R_0 \leq Z^*(X, Y) \leq R_0 + (J/N)R. \qquad \text{(B8)}$$

The number $J = J(X, Y)$ can vary between two limits: $J = 0$ (none of the experimental points can be projected on a sphere centered at $[X, Y, Z^*(X, Y)]$) and $J = N$ (where all points can be projected on a sphere). This gives the upper bound on $Z^*(X, Y)$ independent of X and Y

$$R_0 \leq Z^*(X, Y) \leq R_0 + R. \qquad \text{(B9)}$$

In the case when $J = 0$, $Z^*(X, Y) = R_0$ and $ErD_p(X, Y, Z^*)$ can be written as

$$ErD_p(X, Y, Z^*) = (1/N) \sum_{n=1}^{N} \left[(R_0 - r_n)^2 + (Q_n - R)^2 \right].$$

(B10)

Then, the conditions (B7) for a vanishing gradient are given by

$$\sum_{n=1}^{N} (Q_n - R) \frac{X^* - X_n}{Q_n} = 0 \quad \sum_{n=1}^{N} (Q_n - R) \frac{Y^* - Y_n}{Q_n} = 0.$$

(B11)

Noting that $-1 \le (X^* - X_n)/Q_n \le 1$ and $-1 \le (Y^* - Y_n)/Q_n \le 1$, the following restrictions have to apply for the case when $J = 0$:

$$|X^* - X_0| \le R \quad |Y^* - Y_0| \le R$$

(B12)

where X_0 and Y_0 are coordinates of the centroid of all data points

$$X_0 = (1/N) \sum_{n=1}^{N} X_n \quad Y_0 = (1/N) \sum_{n=1}^{N} Y_n.$$

(B13)

However, (B12) contradicts the assumption that $J = 0$ (which requires that, for a given (X, Y), all $Q_n(X, Y) \ge R$ for $n = 1, \ldots, N$). Therefore, the directional error function cannot have a minimum at $[X', Y', Z'(X', Y')]$ if $J(X', Y') = 0$, and the search space in the minimization of ErD_p is limited to a cylinder of radius $2R$ and height R, with its center located at $[X_0, Y_0, R_0 + R/2]$ and its axis parallel to the Z-axis of the coordinate frame.

If data points U_j equally cover the entire hemisphere of a target facing the instrument, then one may reasonably assume that the only minimum of ErD_p is located at $X^* \approx X_0$, $Y^* \approx Y_0$, and $Z^* \approx Z^*(X_0, Y_0)$. For other data sets, where a target is partially occluded and the data points are not evenly distributed over the whole hemisphere, the above assumption does not hold. However, contour plots of ErD for data sets covering only half of a hemisphere (manually generated from the original data sets investigated in this paper) have only one minimum, similar to the plots of ErD included in Figs. 6–8.

REFERENCES

[1] W. C. Stone, M. Juberts, N. Dagalakis, J. Stone, and J. Gorman, *Performance Analysis of Next-Generation LADAR for Manufacturing, Construction, and Mobility*, 2004. NISTIR 7117.
[2] E. Golub and C. F. V. Loan, *Matrix Computation*. Baltimore, MD: The Johns Hopkins Univ. Press, 1996.
[3] C. Witzgall, G. S. Cheok, and A. J. Kearsley, "Recovering circles and spheres from point data," in *Perspectives in Operations Research*, F. B. Alt, M. C. Fu, and B. L. Golden, Eds. New York: Springer-Verlag, 2006, pp. 393–413.
[4] P. Gilmore and C. T. Kelley, "An implicit filtering algorithm for optimization of functions with many local minima," *SIAM J. Optim.*, vol. 5, pp. 269–285, 1995.
[5] R. Glowinski and A. J. Kearsley, "On the simulation and control of some friction constrained motions," *SIAM J. Optim.*, vol. 5, pp. 681–694, 1995.
[6] R. M. Lewis and V. Torczon, "Pattern search algorithms for bound constrained optimization," *SIAM J. Optim.*, vol. 9, pp. 1082–1099, 1999.
[7] S. J. Ahn, W. Rauh, and H. J. Warnecke, "Least-squares orthogonal distances fitting of circle, sphere, ellipse, hyperbola, and parabola," *Pattern Recognit.*, vol. 34, no. 12, pp. 2283–2303, Dec. 2001.
[8] I. D. Coope, "Circle fitting by linear and nonlinear least squares," *J. Optim. Theory Appl.*, vol. 76, no. 2, pp. 381–388, Feb. 1993.
[9] D. Umbach and K. N. Jones, "A few methods for fitting circles to data," *IEEE Trans. Instrum. Meas.*, vol. 52, no. 6, pp. 1881–1885, Dec. 2003.
[10] W. Gander, G. H. Golub, and R. Strebel, "Least-squares fitting of circles and ellipses," *BIT Numer. Math.*, vol. 34, no. 4, pp. 558–578, Dec. 1994.
[11] J. Garcia-Lopez, P. A. Ramos, and J. Snoeyink, "Fitting a set of points by a circle," *Discrete Comput. Geom.*, vol. 20, no. 3, pp. 389–402, 1998.
[12] I. Kasa, "A circle fitting procedure and its error analysis," *IEEE Trans. Instrum. Meas.*, vol. IM-25, no. 1, pp. 8–14, Mar. 1976.
[13] Y. Nievergelt, "Computing circles and spheres of arithmetic least squares," *Comput. Phys. Commun.*, vol. 81, no. 3, pp. 343–350, 1994.
[14] M. Renault, *Fitting Circles and Ellipses to Data Using the Least-Squares Method*. [Online]. Available: http://www.math.temple.edu/~renault/ellipses.html
[15] C. M. Shakarji, "Least-squares fitting algorithms of the NIST algorithm testing system," *J. Res. NIST*, vol. 103, pp. 633–640, 1998.
[16] H. Spath, "Least-square fitting with spheres," *J. Optim. Theory Appl.*, vol. 96, no. 1, pp. 191–199, Jan. 1998.
[17] B. A. Jones and R. B. Schnabel, "A comparison of two sphere fitting methods," in *Proc. IEEE Instrum. Meas. Technol.*, 1986, pp. 104–109.
[18] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 1995.

Marek Franaszek received the M.Sc degree in physics from Jagiellonian University in Krakow, Poland, and the Ph.D. degree in physics from the Institute of Molecular Physics, Polish Academy of Science, Poznan, Poland, where he worked on deterministic chaos and noise, in 1986.

He is currently with the National Institute of Standards and Technology (NIST), Gaithersburg, MD. Prior to joining NIST in 2007, he was with the National Institutes of Health, Bethesda, MD, where he worked on medical image processing. His current research interests include mathematical modeling and algorithm development for processing data acquired with 3D imaging systems.

Geraldine S. Cheok received the B.S. degree (*cum laude*) and the M.S. degree from the University of Maryland, College Park, in 1983 and 1986, respectively, both in civil engineering.

In 1984, she joined the National Institute of Standards and Technology, Gaithersburg, MD. Her current research interests include the use of 3D imaging systems for construction and the development of standards for these systems.

Kamel S. Saidi received the Ph.D. degree in civil engineering from the University of Texas, Austin, in 2002, where he focused on information technology at construction sites and construction automation systems.

He is currently with the National Institute of Standards and Technology (NIST), Gaithersburg, MD. His current research includes the development of the NIST Construction Automation Testbed, particularly the design and implementation of a robotic crane and various advanced sensing systems, and the establishment of standards for the performance evaluation of 3D imaging systems.

Christoph Witzgall received the Ph.D. degree in mathematics from the University of Munich, Germany, in 1958.

In 1962, he joined the National Institute of Standards and Technology, Gaithersburg, MD, where he serves as an Emeritus. His current research interests include the geometrical, statistical, and computational aspects of large coordinate data sets, such as those produced by 3D imaging systems.