ACCURATE ESTIMATION OF ELLIPTICAL SHAPE PARAMETERS FROM A GREY-LEVEL IMAGE

R. Safaee-Rad
Graduate Student

K.C. Smith
Professor

B. Benhabib
Assistant Professor

Computer Integrated Manufacturing Laboratory, Department of Mechanical Engineering
University of Toronto, 5 King's College Road, Toronto, Ontario, M5S 1A4 Canada

ABSTRACT

A new 3D-object-recognition technique under development at the University of Toronto is based on a pre-marking scheme. One of the main problems to be solved for the implementation of the proposed technique, is the accurate estimation of the five basic parameters of an elliptical shape based on its edge-point data. To achieve the desired degree of accuracy, a new error function is proposed in this paper. Also, to carry on a comparative study, an objective and independent measure for "goodness" of fit is proposed. The proposed new error function and two other previously developed error functions were applied to six different situations. The results of the comparative study show that when all the edge points are used for the fit, the performance of the proposed error function is better than the other two.

1. Introduction

Flexibility of robotic workcells can be significantly increased through the integration of visual sensors, reducing the need for special, and often complex, tooling. A research program currently underway at the University of Toronto focuses on developing an efficient and flexible 3D object-recognition method for vision-based robotic assembly. The main incentives in pursuing this research are the lack of generalized computer-vision theories applicable to robotic vision, and the inefficiency and inflexibility of existing 3D object-recognition methods for industrial environments [1].

Object recognition for assembly requires the identification of workpieces or sub-assemblies, determination of their locations (position and orientation), and extraction of their salient features for visual servoing. Volumetric methods based on the exact specification of an object, and multi-view feature representations based either on the characteristic views of an object or on the discrete view-sphere representation have been developed [2]. However, these techniques are complex and computationally expensive due to their need for a 3D matching process. A new 3D-object recognition technique is based on a pre-marking scheme, as a consequence of which the matching process can be performed in 2D space.

In the proposed 3D-object-recognition technique, an object is modeled using only a small number of 2D distinct perspective views, "**standard-views**", each with a corresponding **standard-view-axis**. For successful recognition purposes, the input image of an object should be one of its standard perspective views. Thus, a mobile camera is used, such that the camera's optical axis can be aligned with one of the standard-view-axes of the object in order to acquire a standard-view. Then, the matching process is performed between the acquired 2D standard-view of the object under consideration and the library of 2D standard-views of a set of objects.

To enable the vision system to acquire standard-views, the standard-view-axes must be pre-defined. This can be accomplished by defining a local surface normal for each distinct view of an object. These local surface normals can be defined by adding non-functional features to an object, preferably during the manufacturing process. This process is referred to as pre-marking. Thus, in the context of the proposed object-recognition method, pre-marking is used for defining local surface normals and the corresponding standard-view-axes.

Several mathematical algorithms for extracting 3D location from the marker's image have already been developed based on markers having circular geometry [3,4,5]. However, to implement these algorithms, the five basic parameters of an ellipse, (the perspective image of a circular marker), have to be accurately known, namely: the center location, length of minor and major radii, and orientation. This estimation is also required for the alignment of the camera's optical axis with the surface normal [4].

In the proposed 3-D object recognition scheme, both accurate camera calibration and camera-robot calibration are essential. To implement these calibration procedures [6,7], estimation of the calibration-point coordinates in the computer-image frame is required with a high degree of accuracy. In this work, the centers of a set of circles are used to define the calibration points. Thus, again, the problem of estimating two parameters of an elliptical shape, namely the ellipse center coordinates, arises.

In this paper, first, the main sources of errors that cause distortion in the shape of an ellipse are presented and methods for compensation of these distortion factors are formulated (section 2). In section 3, a method for estimation of ellipse center coordinates is proposed; a new method for the determination of an optimal elliptical fit to a set of input-edge-points, is presented; a new error function, based on a new geometrical interpretation of an existing error function, for estimation of all five parameters of an ellipse, is defined; and an objective and independent measure for "goodness" of various fits is defined. In section 4, the results of implementing the optimization process using various error functions and analyses of the data obtained, are presented. The conclusion is given in the final section.

2. Main Factors Leading to Distortion of 2D Shapes

Given a grey-level image of an ellipse, it is required to estimate accurately its five basic parameters. The typical level of accuracy desired is 0.01 mm (in about 30 cm) (1) for estimation of the surface normal; (2) for the alignment of the camera optical axis with the surface normal; (3) for camera calibration; and (4) for camera-robot calibration. A better understanding of this level of accuracy can be achieved by considering it in terms of pixels. Thus, in our situation for a 9.4x12.6 inch field of view, each pixel in a 640x480 computer image frame would correspond approximately to 0.5 mm. Thus, 0.01 mm would correspond approximately to 0.02 pixels.

In general, the image of a nominal elliptical shape acquired by a CCD camera is distorted by several factors. Thus, it must go through several compensation processes in order to obtain an accurate estimation of the ellipse's parameters. This sequential compensation procedure is as follows:

(a) *Subpixel-edge detection:* From a grey-level image of an ellipse and all the available information it contains, one must detect its edge with subpixel accuracy. Note that in practice one is faced with shadows around the edge of an ellipse in the image; thus it is impossible to identify edge points precisely by mere thresholding. Thus, a new sub-pixel edge detector based on the principle of the sample-moment-preserving transform (SMPT) was developed for accurate edge detection [8].

(b) *Transformation of computer-image coordinates to real image coordinates:* Computer-image coordinates are expressed in terms of pixel units. Our goal is to define the edge points in terms of absolute length units (in mm), and also to compensate for the timing mismatches which occur between image-acquisition hardware and camera scanning hardware (or even imprecision in the timing of the TV scan itself). For an uncertainty factor S_x, used to represent the mismatch, the following transformation is applied [6]:

$$X_d = \frac{d'_x}{S_x}(X_f - C_x)$$
$$Y_d = d_y(Y_f - C_y)$$

where, (X_d, Y_d) is the (lens-) distorted real image coordinate, (X_f, Y_f) is the computer-image coordinate with subpixel accuracy, (C_x, C_y) is the computer-image coordinate of the center of the computer frame memory, S_x is the uncertainty image scale factor, and d_y is the center-to-center distance between adjacent CCD sensor elements in the Y direction. Furthermore, d'_x is defined as,

CH2898-5/90/0000/0020$01.00 © 1990 IEEE

$$d'_x = d_x \frac{N_{cx}}{N_{fx}}$$

where, d_x is the center-to-center distance between adjacent CCD sensor elements in the X direction, N_{cx} is the number of sensor elements in the X direction, and N_{fx} is the number of pixels in a line as sampled by the computer. S_x, C_x, and C_y are estimated by accurate camera calibration, and the values for the other parameters (d_x, d_y, N_{cx}, and N_{fx}) are supplied by manufacturers of CCD cameras and digitizer boards.

(c) *Lens radial distortion:* The transformation between undistorted true-image coordinates (X_u, Y_u), and distorted real-image coordinates (X_d, Y_d) can be defined as [6],

$$X_u = X_d\,(1 + K\,r^2)$$
$$Y_u = Y_d\,(1 + K\,r^2)$$

where,

K = radial distortion factor

$$r = \sqrt{X_d^2 + Y_d^2}.$$

The K factor is determined through accurate camera calibration.

(d) *Interpolation:* Based on a set of true image coordinates (X_u, Y_n), one must estimate the five parameters of the imaged ellipse. The main body of this paper addresses this problem. In the next section this problem is stated in detail.

In the following section (section 3), the interpolation issue is addressed. Therein, a new error function is formulated for the optimal estimation of the five ellipse parameters. Furthermore, based on an objective performance measure developed here, this error function is compared to other error functions previously developed (section 4).

3. Elliptical-Shape Parameter Estimation

3.1 Problem Statement

As noted earlier, the need for the accurate estimation of the basic parameters of an elliptical shape arises in various machine-vision-related problems. In this section, we briefly review techniques which have been proposed by various research groups, and formulate the problem to be addressed in this paper.

One method for dealing with this estimation problem is based on the use of optimization techniques. From a purely mathematical point of view, the problem of fitting a conic or a conic section to a set of data has been addressed in various papers [9,10]. The same problem is also addressed in the applied literature: in dentistry, for the estimation of dental arch form [11], in biology, for automatic chromosome analysis [12], in the quality control of manufactured parts, for the estimation of an arc center and its radius [13,14]; in vision-based assembly, for polygonal and elliptical approximation of mechanical parts [15]; in object recognition, for detecting cylindrical parts and their orientation [16]; in shape recognition, for detecting ellipses and their orientation as one of selective confirmation items [17]; and, in pattern recognition and scene analysis, for the reconstruction problem [18,19].

Other methods for dealing with the same problem have been investigated: using the Hough transformation for detection of curves (e.g. circles) [20]; using a modified Hough transformation for detecting ellipses [21]; by decomposing the five-dimensional Hough transformation space into three sub-spaces based on the edge-vector-field properties of ellipses [22]; and estimating the parameters of an ellipse by combining transformation, projection and optimum approximation techniques [23].

Most of these papers are concerned more with ellipse *detection* while providing only a rough estimation of parameters. So, generally speaking, the accuracy is not an important aspect in their development. Even those who actually look for accurate estimation of ellipse parameters, confine themselves to finding the "best" fit to a final set of data points. As we have shown in the last section, for accurate estimation of the parameters, one must consider other factors as well.

It is important to emphasize that the quality of input data has a major impact on the accurate parameter estimation. No matter how good the fit may be, when the accuracy of input data is low, accurate estimation of parameters is not possible. The problem is one of quantization error. This important factor has been addressed in a number of studies [24,25]. Subpixel edge detection is closely related to this data-accuracy problem and has been addressed in [8].

In this paper, it will be assumed that the input data has high degree of accuracy, as if it is obtained through subpixel edge detection and following the removal of quantization error. This assumption leads to the following definition of the problem to be solved: "Given a set of 2D-image coordinates of edge points of an elliptical shape, it is required to determine the best ellipse fitting those points and subsequently estimate its five basic parameters, namely, location: (X_0, Y_0) (ellipse center), orientation: Θ_A (the angle between the major axis of the ellipse and the X-axis of the computer-image frame), and shape: A, B (major and minor radii)."

3.2 Parameter Estimation

There are several different methods for the solution of this problem, depending on the number of parameters that must be estimated, their expected level of accuracy, and the associated computational cost. In this paper, we present five different methods, the experimental data obtained for each, and a discussion on the comparative performance of each.

3.2.1 Location estimation of an ellipse

In camera calibration and camera-robot calibration, only an estimate of the center of a circular or elliptical shape is required. Thus, to estimate (X_0, Y_0) accurately without estimating other parameters, would be desirable. One simple and fast method is to use the first moments of the ellipse area. If m_k and n_k are the X-coordinates of edge points in each horizontal line, then the center coordinates can be expressed as follows:

$$X_0 = \sum_K \left\{ \frac{(n_k + m_k)}{2}\,[(n_k + 1) - m_k] \right\} / S \qquad (1)$$

$$Y_0 = \sum_K \left\{ K\,[(n_k + 1) - m_k] \right\} / S \qquad (2)$$

where,

$$S = \sum_K\,[(n_k + 1) - m_k]. \qquad (3)$$

The other possible method which one can use is an optimal fitting of a *circular* curve to an *elliptical* shape. This method is based on the existence of the single parameter held in common between an ellipse and a circle, namely, their center. To implement this method, one can use an analytical solution such as recently derived in [14]. This analytical solution is based on the optimization of an error function, defined as the difference between the constant area πR^2 and the area of the circle centered at (X_0,Y_0) with a radius $[(X_i - X_0)^2 + (Y_i - Y_0)^2]^{1/2}$.

3.2.2 Five-Parameter Estimation

Estimation of the five parameters of an ellipse based on a set of 2D coordinates, can be achieved by defining an error function and then minimizing it. However, the accuracy and cost of the estimation process depend on the geometrical nature of the error function, whether this function is linear or non-linear, as well on its form.

(1) The Error Function J_1 and its Geometrical Nature

Let

$$Q(X,Y) = aX^2 + bXY + cY^2 + dX + eY + f = 0 \qquad (4)$$

be the general equation of an ellipse. Let

$$(X_i,\, Y_i) \quad i = 1, N \qquad (5)$$

be a set of points to be fitted to an elliptical shape. Then, if one considers the value of $Q(X,Y)$ at a given point (X_i, Y_i), the quantity $Q(X_i, Y_i)$ vanishes if the point is on the ellipse, it is negative if the point is inside, and it is positive if the point is outside the ellipse. Accordingly, one technique to fit an elliptical shape to a given set of N points would be to use the least-squares error criterion which minimizes the following error function:

$$J_0 = \sum_{i=1}^{N} [Q(X_i,\, Y_i)\,]^2. \qquad (6)$$

Thus, the objective is to determine a parameter vector $W^T = (a,b,c,d,e,f)$. Although, this error function has been used by different researchers, each of whom has used different constraints and solution methods, most have not discussed the geometrical nature of the optimization process based on it. An exception is Cooper and Yalabik [19] who discussed this aspect of the problem while assuming f is equal to -1. As well, later, Bookstein [10] discussed it for the general case (eq. 4). With reference to Fig. 1, he proved that

$$Q(X_i,\, Y_i)\ \alpha \left[\frac{d_1}{d}\,(\frac{d_1}{d} + 2) \right] \qquad (7)$$

where (d_1 / d) is the ratio of the distances point-to-conic (PP') and center-to-conic ($P'O'$) boundary along the ray PO'. Since d is a maximum in the direction of the ellipse's major axis, and a minimum in the direction of the ellipse's minor axis, it can be seen that if two data points are equidistant from the ellipse, but one lies along the

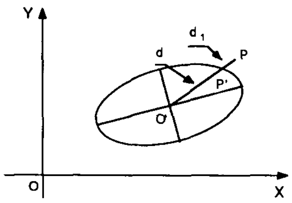

Figure 1. Distances d and d_1 of an ellipse.

major axis and the other along the minor axis, then, the contribution to (6) of the data point lying along the minor axis will be greater.

In this paper, we analyze a solution method [18] which is based on the constraint $f = 1$ (implying normalization with respect to f), for which equation (4) becomes,

$$Q(X,Y) = aX^2 + bXY + cY^2 + dX + eY + 1. \tag{8}$$

Using the least-squares error criterion, the error function is then defined as,

$$J_1 = \sum_{i=1}^{N} [Q(X_i, Y_i)]^2. \tag{9}$$

This method leads to a set of linear equations, for which the optimization process is non-iterative and therefore computationally fast. However, due to the geometrical nature of the error function, the contribution of data points is not uniform, a fact which affects the accuracy of the parameter estimates.

(2) A New Geometrical Interpretation for J_1

It is of interest to define an error function such that the contribution of the data points to the error function is uniform. In order to achieve this, first one must attempt to clarify the geometrical nature of the error function J_1 defined in (9), and then, to normalize the contributions of data points to the error function.

Let (X_i, Y_i) $(i=1,N)$ be a set of data points, and let the optimal parameters of the ellipse be (A,B,Θ,X_0,Y_0). Furthermore, let (A',B',Θ,X_0,Y_0) be the parameters of the ellipse that passes through the data point (X_i,Y_i). The two ellipses are concentric, and have the same orientation (Fig. 2). An error can be now defined as the difference between the two ellipses' areas as,

$$e_i = S - S'_i. \tag{10}$$

In order to calculate S'_i, first consider a line that passes through a data point $P(X_i,Y_i)$ and the center $O'(X_0,Y_0)$; the intersection point of this line with the optimal ellipse is $P'(X'_i,Y'_i)$. By defining $d_i = P'O'$, $d'_i = PO'$, and $\delta_i = d'_i - d_i$, the following relation can be obtained, based on the similarity of two ellipses,

$$e_i = \pi(AB)\left(1 - \frac{d_i'^2}{d_i^2}\right). \tag{11}$$

Using the definition of δ_i, (11) can be re-written in the following form:

$$e_i = \pi(AB)\left[\frac{\delta_i}{d_i}\left(\frac{\delta_i}{d_i} + 2\right)\right]. \tag{12}$$

Comparison of the two relations, (7) and (12), shows that they both represent the same geometrical quantity. To further clarify this issue, an explicit expression for the term $\left(1 - \frac{d_i'^2}{d_i^2}\right)$ in (11) can be derived as follows,

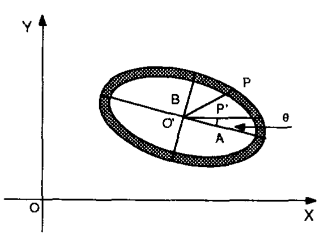

Figure 2. The area difference of two concentric ellipses.

$$1 - \frac{d_i'^2}{d_i^2} = 1 - \frac{[(X_i - X_0)\cos\Theta + (Y'_i - Y_0)\sin\Theta]^2}{A^2} - \frac{[-(X'_i - X_0)\sin\Theta + (Y_i - Y_0)\cos\Theta]^2}{B^2} \tag{13}$$

where (X'_i, Y'_i) are the coordinates of point P'. The expression on the right hand side of (13) is the equation of an ellipse when written explicitly in terms of its five parameters. This expression can be written in the form that was given in (1) as follows:

$$1 - \frac{d_i'^2}{d_i^2} = -(a X_i^2 + b X_i Y_i + c Y_i^2 + d X_i + e Y_i + f) \tag{14}$$

where,

$$a = (A^2 \sin^2\Theta + B^2\cos^2\Theta)/A^2B^2$$
$$b = 2(B^2 - A^2)\sin\Theta\cos\Theta/A^2B^2$$
$$c = (A^2\cos^2\Theta + B^2\sin^2\Theta)/A^2B^2$$
$$d = -2(a X_0 + \frac{b}{2} Y_0)$$
$$e = -2(\frac{b}{2} X_0 + c Y_0)$$
$$f = (a X_0^2 + b X_0 Y_0 + c Y_0^2 - 1).$$

Thus, one can conclude that if the constant term in the equation of an ellipse is normalized such that it would be equal to F, defined as

$$F \equiv (aX_0^2 + bX_0Y_0 + cY_0^2 - 1) \tag{15}$$

the expression for the error defined by (11) can be written as,

$$e_i = -\pi(AB)\,[Q(X_i, Y_i)]_{f \equiv F} \tag{16}$$

where $Q(X_i,Y_i)$ is the equation of an ellipse, in which the constant term is normalized with respect to F.

Based on (16), one can define an error function as,

$$J_2 = \sum_{i=1}^{N}\left[\frac{1}{\pi AB}(S - S'_i)\right]^2 = \sum_{i=1}^{N}\left[\frac{e_i}{\pi AB}\right]^2 = \sum_{i=1}^{N}\left[Q(X_i, Y_i)\right]_{f \equiv F} \tag{17}$$

The comparison of J_2 with J_1 defined by (9) shows that the only difference between the two is that in J_1, $f \equiv 1$ while in J_2, $f \equiv F$; which is just a constant multiplication factor. Thus, one can conclude that the two error functions are equivalent. However, in the process of deriving an expression for J_2, a new and clearer geometrical interpretation of the error function J_1 is obtained. In fact, it has been shown that the optimization of this error function minimizes the error generated due to the difference in *area* as defined in (8). Based on this new interpretation of the error function, we can define a new error function with an improved performance.

(3) An Improved Error Function, J_3

In order to improve the performance of the optimization based on the error function J_1, the contribution of each data point to the error function should be normalized. This can be achieved by defining a weighting factor which is a function of the position of an individual data point. Let δ_i be the distance of a particular data point to the optimal ellipse (PP' in Fig. 2). Then the amount of error due to this data point is,

$$e_1 = \pi AB\left[\frac{\delta_i^2 + 2\,d_i\,\delta_i}{d_i^2}\right]. \tag{18}$$

If this point, with the same δ_i, was on the major axis of the optimal ellipse, the error would be:

$$e_2 = \pi AB\left[\frac{\delta_i^2 + 2A\delta_i}{A^2}\right]. \tag{19}$$

Using (18) and (19), the following weighting factor can be defined for a data point (X_i,Y_i)

$$w_i = \frac{e_2}{e_1} = \left(\frac{d_i}{A}\right)\left[\frac{1 + \frac{\delta_i}{2A}}{1 + \frac{\delta_i}{2d_i}}\right]. \tag{20}$$

Based on the weighting factor defined above, the proposed new error function is defined as follows,

$$J_3 = \sum_{i=1}^{N}\left[w_i\frac{1}{\pi AB}(S - S'_i)\right]^2 = \sum_{i=1}^{N}\left[w_i\,Q(X_i, Y_i)\right]^2_{f \equiv F} \tag{21}$$

$$J_3 = \sum_{i=1}^{N}\left[w_i\,Q(X_i, Y_i)\right]^2_{f \equiv 1}. \tag{22}$$

In order to avoid an *iterative* process, the coefficient w_i must be first estimated for each data point. To achieve this, one must have an *initial* "optimal" ellipse from which δ_i and d_i can be estimated for each data point. Thereafter, the optimization (minimization) of the error function J_3 would proceed by taking the first derivatives with respect to the five unknowns ($a,b,c,d,$ and e), to yield a set of five linear equations with five unknowns, the solution of which is the vector $W^T = (a,b,c,d,e)$. The five parameters of the final optimal ellipse can then be estimated using the following equations,

$$X_0 = \frac{2cd - be}{b^2 - 4ac} \quad , \quad Y_0 = \frac{2ae - bd}{b^2 - 4ac} \; ,$$

$$\Theta_A = atan\left[\frac{(c-a) + \sqrt{(c-a)^2 + b^2}}{b}\right] ,$$

$$A^2 = \left[\frac{2(1-F_S)}{b^2 - 4ac}\right]\left[(c+a) + \sqrt{(c-a)^2 + b^2}\right] , \tag{23}$$

$$B^2 = \left[\frac{2(1-F_S)}{b^2 - 4ac}\right]\left[(c+a) - \sqrt{(c-a)^2 + b^2}\right] ,$$

where,

$$F_S = \frac{bde - ae^2 - cd^2}{b^2 - 4ac} .$$

(4) Error Function J_5

Another existing error function, based on distance measurements, which has been claimed to have a better performance than J_1, is briefly reviewed in this section for its comparison with the newly proposed J_3 function. Let the distance H_i of a data point to an ellipse (Fig. 3) be estimated using the following equation [18].

$$H_i^2 = \frac{\Delta X_i^2 \; \Delta Y_i^2}{\Delta X_i^2 + \Delta Y_i^2} . \tag{24}$$

Then, the error function can be defined as,

$$J_4 = \sum_{i=1}^{N} H_i^2 . \tag{25}$$

However, since the expression for distance H_i s very involved, an alternative, but less involved, distance measure proposed in [15] is considered in this paper, namely

$$J_5 = \sum_{i=1}^{N} d_i \tag{26}$$

where d_i is the distance $P_i \, P_i'$ as defined in Fig. 4, for the data point (X_i, Y_i).

As can be derived, the error function is nonlinear; thus the optimization process must be iterative. Moreover, the form of the error function J_5 is such that, in each iteration for its estimation, the contribution of each individual data point must be evaluated separately. As a result, the optimization process based on the J_5 error function can be very costly.

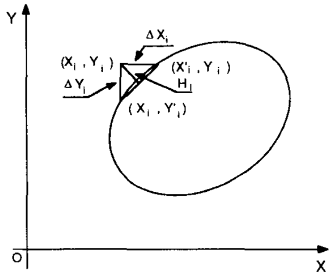

Figure 3. Definition of the Hi distance for an ellipse.

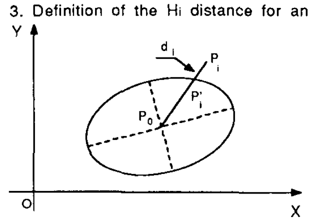

Figure 4. Definition of the di distance for an ellipse.

A final point that should be mentioned here is concerned with the idea of defining an error function based on the normal distance of a data point to an ellipse. Though, as defined, it would yield a best fit, this error function is highly involved, complex and computationally expensive. The reason is due to the nonlinearity of its error function requiring an iterative process of optimization. More importantly, in each iteration for every single point, a fourth-order equation has to be solved in order to find the normal distance.

3.3 An Objective Measure for "Goodness" of Fit

In order to carry out a comparative study of the three error functions J_1, J_3, and J_5, an objective and independent measure should be formulated to determine the level of "goodness" of fit. In this section such an objective measure is defined as "the total normal distance of all data points to the optimal ellipse", and a procedure for its evaluation is presented.

This measure can be obtained following the optimization process of each error function. However, the coordinates of the intersection point of a normal line which passes through a particular data point and the optimal ellipse must be first determined.

Thus, the initial problem can be defined as follows: Given a point (X_i, Y_i) and an ellipse with parameters (A,B,Θ,X_0,Y_0), find the normal line to the ellipse which passes through point (X_i,Y_i). In order to simplify the estimation process, and obtain more manageable equations, a transformation consisting of a translation and then a rotation, will be applied to all the data points and the optimal ellipse, as follows:

$$\begin{aligned} X_{i1} &= X_i - X_0 \\ Y_{i1} &= Y_i - Y_0 \\ X_{iS} &= X_{i1} \cos\Theta + Y_{i1} \sin\Theta \\ Y_{iS} &= -X_{i1} \sin\Theta + Y_{i1} \cos\Theta \; . \end{aligned} \tag{27}$$

Following this, the transformed equation of the optimal ellipse can be written in the following form:

$$\frac{X^2}{A^2} + \frac{Y^2}{B^2} = 1. \tag{28}$$

Let P be a data point, and PP'' be the normal to the ellipse (Fig. 5). Thus, if m_1 is the slope of the tangent line at point P'', and m_2 is the slope of line PP'' , then one of the constraints on the coordinates of point P'' is:

$$m_1 \, m_2 = -1. \tag{29}$$

Let the coordinates of P'' be (X_i'', Y_i''). Then

$$m_2 = \frac{Y_i'' - Y_i}{X_i'' - X_i} \; . \tag{30}$$

It can be shown that the equation of a tangent line to an ellipse at point (X_i'', Y_i'') is given by:

$$\frac{X_i \, X_i''}{A^2} + \frac{Y_i \, Y_i''}{B^2} = 1. \tag{31}$$

Then

$$m_1 = -\frac{B^2}{A^2}\left[\frac{X_i''}{Y_i''}\right]. \tag{32}$$

Thus, the first constraint on point P'' can be expressed as follows (using (29)),

$$\left[\frac{Y_i'' - Y_i}{X_i'' - X_i}\right]\left[\frac{X_i''}{Y_i''}\right] = \frac{A^2}{B^2} . \tag{33}$$

The second constraint on point P'' is that it must satisfy the equation of the optimal ellipse, i.e.,

$$\frac{X_i''^2}{A^2} + \frac{Y_i''^2}{B^2} = 1. \tag{34}$$

From (33), an expression for Y_i'' can be determined. Substituting this expression into (34) and simplifying it, a quartic equation is obtained:

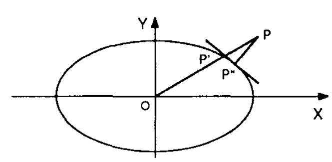

Figure 5. Normal and "central" lines to an ellipse.

$$X_i''^4 + a_1 X_i''^3 + a_2 X_i''^2 + a_3 X_i'' + a_4 = 0 \qquad (35)$$

where,

$$a_1 = \left[2A^2 \left(1 - \frac{A^2}{B^2}\right) X_i \right] / a_5$$

$$a_2 = \left[A^2 Y_i^2 - A^2 B^2 \left(1 - \frac{A^2}{B^2}\right)^2 + \frac{A^4}{B^2} X_i^2 \right] / a_5$$

$$a_3 = \left[-2A^4 \left(1 - \frac{A^2}{B^2}\right) X_i \right] / a_5$$

$$a_4 = \left[-\left(\frac{A^6}{B^2}\right) X_i^2 \right] / a_5$$

$$a_5 = \left[B^2 \left(1 - \frac{A^2}{B^2}\right)^2 \right] .$$

Equation (35) can have a maximum of four real solutions, corresponding to which there would be at most four points satisfying the two constraints. Correspondingly, it can be shown that a data point (X_i, Y_i) can have either two or four normals to an ellipse. Furthermore, it can be shown that one of the normals yields the minimum distance to an ellipse, and this normal is the one that we want to determine. To solve equation (35) the method of solution of a quartic equation based on the resolvent cubic equation can be employed [26].

Once the coordinates of point P'' have been estimated for all data points, $\delta_i' = PP''$ can be calculated for each point. The "goodness" measure is then defined as:

$$G = \sum_{i=1}^{N} \delta_i' , \qquad (36)$$

where an error function that yields a smaller value for G, is a better fit.

4. Experimental Data Analysis

4.1 Standardization of input data

In order to reduce the influence of numerical errors, the coordinates of the input data points obtained from the boundary of an ellipse can be standardized by first calculating the approximate center of ellipse (x_0, y_0), and then carrying out the following translation and scaling:

$$\begin{aligned} x_i &= X_i - x_0 \\ y_i &= Y_i - y_0 \\ x_{si} &= x_i / R \\ y_{si} &= y_i / R \end{aligned} \qquad (37)$$

where,

$$R = \frac{1}{N} \sum_{i=1}^{N} [x_i^2 + y_i^2]^{1/2}. \qquad (38)$$

This process generally reduces the size of the numbers involved in the numerical calculation; thus, reducing the computation time. It should be noted, however, that the corresponding optimal values obtained for the center coordinates (X_0, Y_0), major radius (A), and minor radius (B) of the ellipse must be scaled and translated as well.

4.2 Initial Values for Nonlinear Optimization

As was indicated before, if an error function is nonlinear, the optimization process is generally an iterative one. For such a process a "good" initial set of unknown parameter values can reduce the number of iterations substantially, and reduce the chances of reaching a local minimum due to the nonlinearity of the optimization process.

The use of the J_1 error function is proposed for establishing the initial values, since J_1 is linear, and the estimation process of the five ellipse parameters based on it, is non-iterative. Using the results of the optimization of J_1, as a set of initial values, the other two error functions, J_3 and J_5, can be optimized.

The iterative optimization method used in this study is the simplex method developed in [27].

4.3 Case 1: Simulated Ellipse - No Noise

A "perfect" ellipse is generated using digitizer-board graphic commands, the boundary of which is digitized according to a mathematical procedure that minimizes the digitization error of a continuous boundary of an ellipse (Image 1) [28]. The results of the optimization processes are shown in Tables 1 to 3.

In order to better appreciate the significance of the results, the following premises must be emphasized: (1) There is no noise in the image of the generated ellipse; the coordinates of the edge points are exact in the context of the digitized boundary; thus, all the edge points are symmetric with respect to the center of the ellipse. This implies that the coordinate of the ellipse center ($X_0 = 150.0, Y_0 = 329.0$) and its orientation ($\Theta = -90.0$) are exact, and thus, can be used as a reference for comparison. (2) Due to digitization, no matter how accurate the fitting technique is, there will always be a minimum amount of error referred to as the *dc* error. This error reflects itself in the values of A and B. Thus $A = 60.0$ and $B = 30.0$ can-not be used as reference for comparison purposes. (3) Based on the size of the generated ellipse, the image-scaling ratio in (*pixel / mm*) is about (0.023 / 0.01); that is, if an accuracy, in length measurement, better than 0.01 *mm* is required, the accuracy, in pixel units, must be better than 0.023 pixel.

Review of results (Table 1) for Image 1, N=242

Based on the above premises, and the fact that the edge-point coordinates of the simulated ellipse are very close to the true coordinates, the differences between the results of the various error-function optimizations are very small. As expected, though, in Table 1 some differences are still visible. Nevertheless, the center coordinates of the ellipse are estimated accurately by the first two methods, under the given conditions.

The other three methods, which are based on the optimization of various error functions, provide all five ellipse parameters. J_1 and J_2 both estimate ellipse-center coordinates accurately, but J_5 has an error of -0.027 pixel which is slightly more than desired. Note that one can get more accurate results with J_5 if the number of iterations is increased. With respect to angle Θ (ellipse orientation), J_1, J_3 and J_5 have equal error (considering only the first three decimal points). Unfortunately, for parameters A and B, there does not exist an exact reference, and thus one cannot compare the results directly. However, using G, one can construct a reliability measure to compare the "totality" of the performance of various error functions, one, that is, where *all* five parameters are optimized at the same time. Thus, because we have an exact reference for X_0, Y_0 and Θ, the measure G *indirectly* shows the degree of reliability of the estimated values for A and B, corresponding to each error function.

Comparison of G-values shows the following: (1) The G-value for the generated ellipse is much higher than the corresponding values for the three optimization methods (-59 compared to -55). This means that the values $A = 60.0$ and $B = 30.0$ are less accurate than the other sets of values. (2) The results for the three optimizations are very close, though the G-value for the error function J_3 is less than the other two. On the whole, we can conclude that under the given condition, J_3 performs better than either J_1 and J_5.

Review of results (TABLE 2) for Image 2, N=50

In this case, to evaluate the sensitivity of the process to the number of edge points used, we sample edge points in 5-pixel steps along the Y-axis. Furthermore, the sampled points are all symmetric with respect to the ellipse center $X_0 = 150.0$ and $Y_0 = 329.0$. Under the given conditions we see a similar performance to that in the earlier case (N = 242) by all five methods with respect to X_0, Y_0, and Θ; though, error ΔL and error Θ for J_5 is less than before.

Image1. A simulated ellipse - no noise.

TABLE 1. Experimental results for a simulated ellipse (N=242)

METHOD	X0	Y0	A	B	Θ	# iter.	G	ΔL
GENERATED ELLIPSE	150.000000	329.000000	60.000000	30.000000	-90.000000	--	59.075732	---
FIRST MOMENT	150.000000	329.000000	---	---	---	--	---	0.0
CIRCULAR APPROX.	150.000000	329.000000	---	---	---	--	---	0.0
J1	150.000000	329.000000	60.181337	29.788122	-89.999999	--	55.485145	0.0
J3	150.000000	329.000000	60.227620	29.767604	-89.999999	--	55.240819	0.0
J5	150.000000	329.026941	60.206829	29.789570	-89.999991	124	55.291314	0.026941

Notes: All edge points are used (N=242).
Lengths in pixels, angles in degrees.
$\Delta L = [(X_0 - 150.0)^2 + (Y_0 - 329.0)^2]^{1/2}$.
$G = \Sigma \delta'_i$, δ'_i: The normal distance from a data point to the optimal ellipse.

As a whole, J_5 performs better than the other two, though J_3 is very close to J_5. Thus, it can be concluded that using a subset of data points reduces the "goodness" of the overall performance of J_3. The only reason for this behavior is that reducing the number of points reduces the effect of the weighting factors and thus relatively increases the G-value for J_3 compared to the one for J_5.

Review of results (Table 3) for Image 3, N=20

In this case, only a segment of the ellipse boundary is used. As a result, the input data has symmetry with respect to the Y-axis but it does not have symmetry with respect to the X-axis. The objective is to evaluate the error-function performance when one has data points from a partial ellipse.

Due to the symmetry with respect to the Y-axis, the first two methods estimate X_0 exactly, but both fail to estimate Y_0 accurately. Thus, it can be concluded that both methods are effective only if data that represents the whole ellipse boundary is available and, moreover, symmetricity with respect to ellipse center is maintained. In other words, the level of accuracy obtained depends on the degree of symmetry of the input data with respect to the ellipse center.

The three error functions, though based on 20 data points that cover only a segment of the ellipse boundary, estimated the five ellipse parameters with relatively small errors. J_3 estimated Y_0 more accurately than J_5 or J_1. On the other hand J_3 and J_1 provides more accurate values for Θ than J_5.

As a whole, J_5 performs better than the other two, though J_3 performs better than J_1. Here again, the same effect of reducing the number of data points on the total performance of J_3 is observed.

4.4 Case 2: An Imaged Ellipse

In this case, an "imperfect" ellipse image was used to simulate possible external distortions, such as the thresholding effect on a grey-level image, (Image 4). The shape of this ellipse is further distorted by its passage through the image-acquisition system (that is, the camera and the digitizer, for which all the distortion factors that were earlier-mentioned apply). The results of the analyses are given in Tables 4 to 6 (Images 4 to 6).

In this case, the ellipse parameters are not known exactly in advance as in the simulated ellipse case. Thus, the only objective measure of comparison is the G value. J_3 results in the least G value, and J_1 results in the greatest G value when all data points are used. On the other hand, when the number of input data points is reduced, the relative G-value of J_3 increases compared to the one for J_5. That is, here again, reduction of data points reduces the level of total performance of J_3.

The first two methods, for only ellipse-center-coordinate estimation, show the same behavior, that is, if one uses all edge points, they give more accurate estimates. On the other hand, if a set of edge points from a segment of an ellipse boundary is used, there exists a large amount of error in the estimated values, especially for the first method (see Table 6). As a whole, circular-approximation method results in a more accurate estimation of ellipse center coordinates, under varying conditions, than the first-moment method.

TABLE 2. Experimental results for a simulated ellipse (N=50)

METHOD	X0	Y0	A	B	Θ	# iter.	G	ΔL
GENERATED ELLIPSE	150.000000	329.000000	60.000000	30.000000	-90.000000	--	10.957015	---
FIRST MOMENT	150.000000	329.000000	---	---	---	--	---	0.0
CIRCULAR APPROX.	150.000000	329.000000	---	---	---	--	---	0.0
J1	150.000000	329.000000	60.277751	29.791679	-89.999999	--	9.624126	0.0
J3	150.000000	329.000000	60.310573	29.775033	-89.999999	--	9.551414	0.0
J5	150.000000	328.999998	60.307152	29.762277	-90.000003	120	9.523424	0.000002

Note: Sampled edge points are used (N=50).

Image2. Sampled edge points of a simulated ellipse. White lines are used to indicate the sampled edge points.

TABLE 3. Experimental results for a simulated ellipse (N=20)

METHOD	X0	Y0	A	B	Θ	# iter.	G	ΔL
GENERATED ELLIPSE	150.000000	329.000000	60.000000	30.000000	-90.000000	--	4.113339	---
FIRST MOMENT	150.000000	361.617371	---	---	---	--	---	32.617371
CIRCULAR APPROX.	150.000000	357.355758	---	---	---	--	---	28.355758
J1	150.000000	327.151639	62.265487	30.018486	-89.999999	--	3.316382	1.848361
J3	150.000000	328.127738	61.234822	29.885090	-89.999999	--	3.237098	0.872262
J5	150.000000	327.064834	62.248258	29.948910	-90.000005	145	2.933231	1.935166

Note: Sampled edge points of a segment of an ellipse are used (N=20).

Image 3. Sampled edge points of a segment of an ellipse. White lines are used to indicate the sampled edge points.

TABLE 4. Experimental results for an imaged ellipse (N=160)

METHOD	X0	Y0	A	B	Θ	# iter.	G
GENERATED ELLIPSE	?	?	?	?	?	--	?
FIRST MOMENT	98.384061	415.710963	---	---	---	--	---
CIRCULAR APPROX.	98.260115	416.028898	---	---	---	--	---
J1	98.288630	415.916523	57.935092	30.011632	-30.753048	--	65.647547
J3	98.190024	415.982907	57.823128	30.067143	-30.618907	--	64.605385
J5	98.288699	415.899130	57.930050	29.999341	-30.749076	180	65.525370

Note: All edge points are used (N=160).

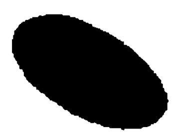

Image 4. An imaged ellipse. A binary image of an elliptical shape obtained by thresholding a grey-level image.

TABLE 5. Experimental results for an imaged ellipse (N=32)

METHOD	X0	Y0	A	B	Θ	# iter.	G
GENERATED ELLIPSE	?	?	?	?	?	--	?
FIRST MOMENT	98.226619	416.000899	---	---	---	--	---
CIRCULAR APPROX.	98.228790	416.224052	---	---	---	--	---
J1	98.211804	416.056472	57.623615	30.045967	-30.543799	--	10.416757
J3	98.147285	416.111461	57.537374	30.070007	-30.379317	--	10.329010
J5	98.216410	416.036456	57.619913	30.034260	-30.279920	138	10.107135

Note: Sampled edge points are used (N=32).

Image 5. Sampled edge points of an imaged ellipse. White lines are used to indicate the sampled edge points.

Conclusions

For camera calibration, just the center coordinates of an ellipse are needed. Thus, the first moments of the ellipse area, and another proposed method based on circular approximation of an ellipse, were tested. The results show that the later method has a better performance. The degree of accuracy is approximately 0.1 mm (in 30 cm) based on the use of all edge points. However, since this is less than desired accuracy a more accurate method, that is an optimization method, is shown to be needed.

For the estimation of all the five ellipse parameters, three different optimization methods based on different error functions (J_1, J_3 and J_5) were presented. J_3 is the new error function proposed in this paper. It is a modified form of the error function J_1, where a weighting factor is applied to each data point. To define this weighting factor, a clearer geometrical interpretation of the error function J_1 was obtained. Furthermore, for objective and independent comparison of the performance of various error functions, an independent and reliable measure was proposed in this paper.

All three error functions were applied to six different cases. The first three cases are based on an ideal simulated ellipse, while the second three involved an imaged ellipse. The results show that the performance of the proposed error function J_3 is better than the other two if all the edge points are used, which is generally true in the proposed 3D-object-recognition technique. When the number of edge points is reduced, or when a partial ellipse is considered, the *total* performance of J_5 is better than J_3. However, a general problem with J_5 is the iterative nature of its optimization process, and the need for the estimation of the function value for each individual data point, in each iteration.

References

[1] Pavlidis, T., "A Critical Survey of Image Analysis Methods", *Proc., Int. Conference on Pattern Recognition,* pp. 502-511, 1986.

[2] Basel, P.J. and Jane, R.C., "Three-Dimensional Object Recognition", *Computing Surveys,* Vol. 17, No. 1, pp. 75-145, March 1985.

[3] Zhou, Z., Smith, K.C., Benhabib, B., and Safaee-Rad, R., "Morphological Skeleton Transforms for Determining Position and Orientation of Pre-Marked Objects", *IEEE, Proc., Pacific Rim Conference on Communications, Computers and Signal Processing,* pp. 301-305, Victoria, B.C., June 1989.

[4] Safaee-Rad, R., Benhabib, B., Smith, K.C., and Zhou, Z., "Pre-Marking Methods for 3D Object Recognition", *IEEE, Proc., Int. Conference on Systems, Man, and Cybernetics,* Boston, MA, Nov. 1989.

[5] Safaee-Rad, R., Smith, K.C., Benhabib, B., and Tchoukanov, I., "A New Technique for 3D-Location estimation of Pre-Marked Objects for An Active Vision System", *IEEE, Int. Conference on Systems, Man, and Cybernetics,* Nov. 1990. (Submitted)

[6] Tsai, R.Y., "A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses", *IEEE, J. of Robotics and Automation,* Vol. RA-3, No. 4, pp. 323-344, Aug. 1987.

[7] Shiu, Y.C., "Calibration of Wrist-Mounted Robotic Sensors by Solving Homogeneous Transform Equations of the Form AX=XB", *IEEE, Trans. on Robotics and Automation,* Vol. 5, No. 1, pp. 16-29, Feb. 1989.

[8] Tchoukanov, I., Safaee-Rad, R., Benhabib, B., and Smith, K.C., "Sub-Pixel Edge Detection for Accurate Estimation of Elliptical Shape Parameters", *Proc., CSME Mechanical Engineering Forum 1990,* Toronto, ON, June 1990. (In Print)

[9] Gnanadesikan. R., **Methods for Statistical Data Analysis of Multivariate Observations,** John Wiley & Sons, 1977.

[10] Bookstein, F.L., "Fitting Conic Sections to Scattered Data", *Computer Graphics and Image Processing,* Vol. 9, pp. 59-71, 1979.

[11] Biggerstaff, R.H., "Three Variations in Dental Arch Form Estimated by a Quadratic Equation", *Dent. Res.,* Vol. 51, No. 5, pp. 1509, Sept.-Oct. 1972.

[12] Paton, K.A., "Conic Sections in Automatic Chromosome Analysis", *Machine Intelligence,* Vol. 5, pp. 411-434, 1970.

[13] Landau, U.M., "Estimation of a Circular Arc Center and its Radius", *Computer Vision, Graphics, and Image Processing,* Vol. 38, pp. 317-326, 1987.

[14] Thomas, S.M., and Chan, Y.T., "A Simple Approach for the Estimation of Circular Arc Center and Its Radius", *Computer Vision, Graphics, and Image Processing,* Vol. 45, pp. 362-370, 1989.

[15] Nakagawa, Y., and Rosenfeld, A., "A Note on Polygonal and Elliptical Approximation of Mechanical parts", *Pattern Recognition,* Vol. 11, pp. 133-142, 1979.

[16] Nagata, T., Tamura, H., and Ishihashi, K., "Detection of an Ellipse by use of a Recursive Least-Square Estimator", *J. of Robotic Systems,* Vol. 2, pp. 163-177, 1985.

[17] Lee, H.C., and Fu, K.S., "3-D Shape From Contour and Selective Conformation", *Computer Vision, Graphics, and Image Processing,* Vol. 22, pp. 177-193, 1983.

[18] Albano, A., "Representation of Digitized Contours in Terms of Conic Arcs and Straight-Line Segments", *Computer Graphics and Image Processing,* Vol. 3, pp. 23-33, 1974.

[19] Cooper, D.B., and Yalabik, N., "On the Computational Cost of Approximating and Recognizing Noise-Perturbed Straight Lines and Quadratic Arcs in the Plane", *IEEE, Trans. on Computers,* Vol. C-25, No. 10, pp. 1020-1032, 1976.

[20] Duda, R.O., and Hart, P.E., "Use of the Hough Transformation to Detect Lines and Curves in Pictures", *Communications of the ACM,* Vol. 15, No. 1, pp. 11-15, Jan. 1972.

[21] Tsuji, S., and Matsumoto, F., "Detection of Ellipses by a Modified Hough Transformation", *IEEE, Trans. on Computers,* Vol. C-27, No. 8, pp. 777-781, 1978.

[22] Tsukune, H., Goto, K., "Extracting Elliptical Figures From an Edge Vector Field", *IEEE, Proc., Conference on Computer Vision and Pattern Recognition,* pp. 138-141, Jan. 1983.

[23] Wang, R., Hansen, A.R., and Riseman, E.M., "Fast Extraction of Ellipses", *Proc., 9th Int. Conference on Pattern Recognition,* Vol. 1, pp. 508-510, 1988.

[24] Ho, C.-S., "Precision of Digital Vision Systems", *IEEE, Trans. On Pattern Analysis and Machine Intelligence,* Vol. PAMI-5, No. 6, pp. 593-601, Nov. 1983.

[25] Kamgar-Parsi, B., and Kamgar-Parsi, B., "Evaluation of Quantization Error in Computer Vision", *IEEE, Trans. on Pattern Analysis and Machine Intelligence,* Vol. PAMI-11, No. 9, pp. 929-940, Sept. 1989.

[26] Beyer, W.H., **Standard Mathematical Tables,** CRC Press Inc., Florida, 1981.

[27] Nelder, J.A., and Mead, R., "A Simple Method for Function Minimization", *Computer Journal,* Vol. 7, pp. 308-313, 1964.

[28] Maxwell, P.C., and Baker, P.W., "The Generation of Polygons Representing Circles, Ellipses and Hyperbolas", *Computer Graphics and Image Processing,* Vol. 10, pp. 84-93, 1979.

Acknowledgement

This work was partially supported by the Manufacturing Research Corporation of Ontario (MRCO).

TABLE 6. Experimental results for an imaged ellipse (N=20)

METHOD	X0	Y0	A	B	θ	# iter.	G
GENERATED ELLIPSE	?	?	?	?	?	--	?
FIRST MOMENT	90.092414	427.827586	---	---	---	--	---
CIRCULAR APPROX.	91.607475	414.117651	---	---	---	--	---
J1	97.788037	416.850320	57.094373	29.533836	-29.445986	--	5.777911
J3	97.612431	417.087419	57.746596	29.475077	-29.063033	--	5.514233
J5	97.795913	416.895087	57.041361	29.639491	-28.994755	150	5.358229

Note: Sampled edge points of a segment of an ellipse are used (N=20).

Image 6. Sampled edge points of a segment of an ellipse. White lines are used to indicate the sampled edge points.