Capturing Facial Surface Information

Michael Keefe
Department of Mechanical and Aerospace Engineering, University of Delaware, Newark, DE 19716
Donald R. Riley
University of Minnesota, Minneapolis, MN 55455

ABSTRACT: Generating range information using stereophotogrammetry traditionally relies on human operators to correlate the multiple images. Automation of that procedure has proven difficult or highly specific to certain scenes. Active (single-viewpoint) techniques can automatically produce range information, but must rely on restricted domains. This paper presents a passive (multi-viewpoint) approach which uses a structured light technique we developed to semi-automatically generate surface information. The stereometric technique of Direct Linear Transformation (DLT) is utilized to eliminate elaborate recording instrumentation. By photographing a series of laser-generated dots projected on the surface, the correlation necessary between the stereo pair is simplified, making automatic correlation realistic. Automatic digitizing using an image analyzer requires no subjective interpretation (for most cases). Finally, the system's computer base provides graphic manipulation of the resulting surface data plus the ability to generate a database. This can provide a cost-effective tool useful to those interested in studying facial form.

INTRODUCTION

THE HUMAN FACE is often modified surgically not for physiological reasons but rather for aesthetic reasons. Because facial beauty is very important in our society for its social and psychological effects, it is not surprising that facial description has been the subject of many quantifying techniques.

Yet facial treatment is still largely an art. Qualitative judgments are made based primarily on clinical experience. The need to develop a more objective approach has long been apparent (Herron, 1972). A large body of facial data has been collected based on two-dimensional data traced from roentgenograms. Cephalometric (face measuring) procedures have been defined and have proven quite useful for studying the soft tissue of the face and the underlying bones. Unfortunately, such data are limited to the mid-sagittal profile. Besides being incomplete, two-dimensional data are ambiguous when used to define a three-dimensional object; therefore, two-dimensional methods can lead to errors (Moyers and Bookstein, 1979).

However, the three-dimensional surface has largely been avoided in the clinical analysis of facial form, primarily because an accurate, economical, and rapid method for three-dimensional surface mapping has not yet been achieved (Young and Altschuler, 1981). Facial researchers have utilized plaster moulages, or three-dimensional casts made from impressions of the surface. Although qualitatively pleasing, the moulages do not solve the problem of extracting the three-dimensional data from the surface, and they are time consuming and difficult to obtain. Also, this technique does not measure the facial surface directly.

Clearly, a major objective would be to develop a system capable of creating a database of three-dimensional facial profiles which could then be used to quantify what is meant by "normal" features and determine the range of normal for those features. That knowledge could then form the basis for research in objective treatment planning. Because of the wide variety and complex nature of human facial form, the database must statistically encompass a large and random population sample. This will only be feasible if the tool to gather the data can be applied clinically, i.e., accurate, safe, reliable, economical, simple, easy to use, fast, and automatic.

There are many approaches to gathering three-dimensional information about an object. Table 1 indicates those techniques presently available along with their major deficiencies when used to clinically gather data. A blank box represents no particular problem. From this, the two most promising techniques for a clinical tool are Moiré analysis and stereophotogrammetry.

Moiré analysis is a common interferometric technique for generating range information. One obvious advantage is that a Moiré pattern provides a direct means of observing and recording contour lines. However, one cannot judge from a single picture of Moiré contours whether a concentric fringe represents a hill or a valley; i.e., there is an ambiguity in the fringe order. Many different methods can be utilized to resolve this ambiguity but they usually require more equipment and/or more processing (Manual of Photogrammetry, 1980).

Furthermore, quantitative analysis becomes tedious and very time consuming when one attempts to handle the entire data from a Moiré topogram (Soares, 1983). Although many clinical Moiré applications rely on manual digitizing of the contour (e.g.,

PHOTOGRAMMETRIC ENGINEERING AND REMOTE SENSING,
Vol. 52, No. 9, September 1986, pp. 1539–1548.

0099-1112/86/5209-1539$02.25/0

TABLE 1. THREE-DIMENSIONAL DATA GATHERING METHODS

	Manual	Computed Tomography			Holography	Moiré	Stereo Techniques	
		CT Scan	NMR	Ultrasound			Active	Passive
Quantitative: Error ±1mm Resolution 2mm	Time consuming	Slice Thickness		Theoretical				
Development Status			Biological Tissue	Rays Bend Biological	Plate Handling			
Automation	Unwieldy					Ambiguous Fringe		Match Views
Data Acquisition Speed		Seconds	Minutes	Seconds			Depends on Set Up	
Clinic: Human Subjects	Distortion	Radiation		Sensors	Motion Distortion		Restricted Domain	
Equipment	Bulky	Expensive	Expensive	Expensive	Bulky Isolation Mounts		Costly	

Saunders, 1983), an automatic method of data processing is desired. There are techniques being investigated to automate Moiré image digitizing (Yatagai *et al*, 1982). Because the fringe number is needed to determine depth, it is difficult to simplify the contour data without limiting the depth being represented.

Stereophotogrammetry has been the most prominent method for contour mapping. The best known application is probably the compilation of topographic maps and surveys. The stereo procedure involves viewing an object from two (or more) distinct points. Stereo analysis allows one to use geometry to recover information about the three-dimensional nature of the object. The main problem with stereo analysis is locating corresponding points in each of the views. This becomes especially difficult on relatively featureless surfaces. Although this problem can be extremely difficult for raw images such as terrain analysis (Panton 1978), recent systems for robotic vision and industrial inspection have shown that structured light can simplify the image processing (Agin and Highnam, 1982; Duncan *et al.*, 1984). A discussion of different range finding techniques for computer vision can be found in the article by Jarvis (1983). Commercial inspection systems utilizing stereo techniques are in operation (Burgam, 1984).

Of these two approaches, stereophotogrammetric techniques have been demonstrated, through their use in commercial systems, to provide a simpler solution to the problems associated with automating the data processing.

SYSTEM

The basic assumption behind stereophotogrammetry is that the image, which is assumed to lie on a perfect plane, is formed by a projection through a single point of the object space. This means that there is a straight line which passes through any object point (O), its corresponding image point (o), and a common projection center (P (Figure 1).

Given an arbitrary X, Y, Z object space coordinate system, one can then solve for the x, y image space coordinates using basic linear transformations (Rogers and Adams, 1976).

One can transform between any arbitrary coordinate systems through a finite series of rotations and translations that include perspective using

$$[X\ Y\ Z\ 1] \begin{bmatrix} L_1 & L_5 & R_1 & L_9 \\ L_2 & L_6 & R_2 & L_{10} \\ L_3 & L_7 & R_3 & L_{11} \\ L_4 & L_8 & R_4 & 1 \end{bmatrix} = s^*[x\ y\ z\ 1]$$

where L_9, L_{10}, L_{11} contain the information for s, the perspective scaling; L_4, L_8, R_4 represent translations; and the remaining 3 by 3 matrix contains the concatenated rotations. However, because we are assuming that the image space is a perfect plane, z =

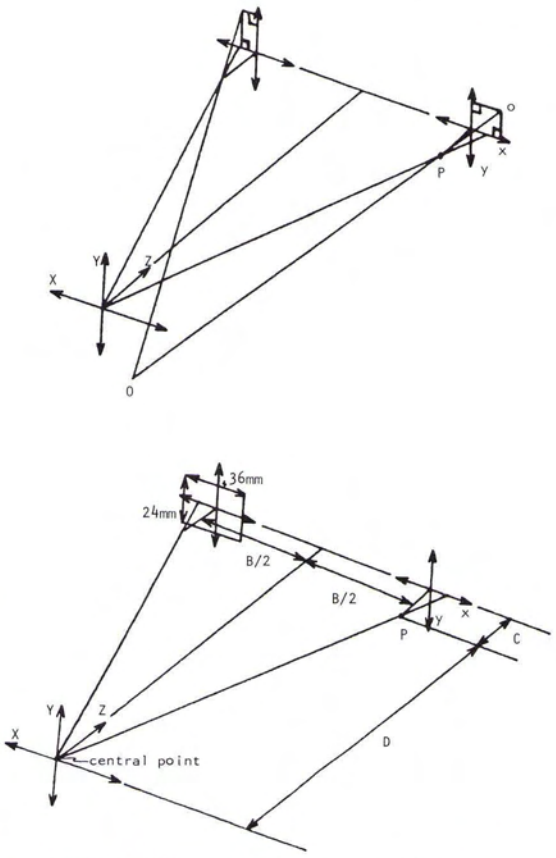

FIG. 1. Stereophotogrammetric setup and variables.

0 for all X, Y, Z. Therefore, $R_1 = R_2 = R_3 = R_4 = 0$. One can now write the following equations:

$$x = (L_1{*}X + L_2{*}Y + L_3{*}Z + L_4)\ /s$$
$$y = (L_5{*}X + L_6{*}Y + L_7{*}Z + L_8)\ /s$$
$$s = L_9{*}X + L_{10}{*}Y + L_{11}{*}Z + 1.0$$

where x,y are the planar image coordinates, and $L_1,\ldots,L_{11}$ are 11 coefficients.

These equations were formulated for stereophotogrammetry by Abdel-Aziz and Karara in 1971 and are called Direct Linear Transformation (DLT) because they represent a linear transformation and can be solved directly without needing initial approximation for the unknowns. An advantage of DLT over other methods is that it requires no prior information about the recording cameras and thus allows the use of relatively inexpensive non-metric cameras.

One problem with a clinical application of stereophotogrammetry had been the special equipment required for stereophotogrammetry, for example, metric cameras. Metric cameras have interior orientation parameters which are well-defined. According to Karara (1972), compared to metric cameras, non-metric cameras tend to have a smaller image format, have no film flattening plate, lack fiducial marks, lack level bubbles or other orientation devices, have an unstable principal distance, and have more and sometimes irregular lens distortion.

However, if one can work around the problems, non-metric cameras are more suitable for a clinical tool by offering lower cost, better availability, a wider range of focus, interchangeable lenses, power drives, and orientation mobility. Therefore, our system was based on using 35-mm cameras, at present the most common non-metric camera in clinical use.

To use DLT, one must think of the equations in two ways. First, the equations must have a minimum of six known points (x,y and corresponding X, Y, Z locations) to solve for the 11 coefficients and, thus, indirectly to solve for the interior orientation parameters of each camera. Second, once the 11 coefficients have been found, one can get two equations for any film coordinate point (x, y) containing the three unknown object space coordinates (X,Y,Z). With a minimum of two views, a least-squares approach can be used to solve for the unknown object space coordinates.

In order to apply DLT clinically, the following procedure was developed (a more detailed description of the equipment is given in Keefe (1985)). Two 35-mm SLR cameras are fixed onto a plate mounted on a tripod. Near the subject is a laser with a collimator and a cylindrical lens.

A calibration frame, containing eight light-emitting diodes (LEDs) whose locations have been precisely measured, is first placed in the view and photographed simultaneously from two views. This known object is then removed, and the subject is placed in the field of view with a lightweight headframe which contains LEDs that provide the information necessary to orient the facial information. By being separate from the calibration, the system's headframe provides the necessary reference plane for the generated three-dimensional coordinates while still being mobile and patient oriented. However, because the calibration is not viewed in every shot, one must compensate for possible film movement in the image analyzer's film transport mechanism. This is accomplished with a fixed pair of reference diodes visible in every exposure and providing a line of reference for the image data. Also, the calibration must be re-photographed at the end of a data-gathering session to insure that the geometry was not accidentally altered.

For any stereophotogrammetric method, the points from the two stereo views must be correlated. This is easy when single spots of light, for example LEDs, are used as in the calibration or headframe. However, the facial surface is much more difficult. As mentioned, general correlation methods are clinically impractical. Because this is a close-range application, a simpler method is to place visible landmarks on the surface. In order to minimize patient discomfort, ease application, and limit changes to the facial surface caused by the marks, this sys-

tem produces the landmarks by projecting structured light onto the face. This method works well with the use of cameras to record the information and is applicable to any visible surface. Potential problems needing consideration are (1) sharp changes or large surface angles may cause the reflected light to be blocked or indistinguishable in a view, i.e., the point must be visible to both cameras; and (2) variations in reflected intensity may be caused by the dot generating method or surface characteristics such as skin pigmentation.

One method would be to project an entire grid at once (Altschuler *et al.*, 1981). Achieving the necessary resolution at the face would require projecting about 10,000 points. For a passive system, this would make automatic correlation of the images very difficult due to the intricacy of the large number of points reflected off an uneven surface. As the number of points increases, this method becomes very similar to correlating the raw images directly. Another problem is in creating that large a number of points. Using multiple sources can quickly become bulky and impractical. However, exposure times necessary to record the information increase as more points are generated from a single source.

Alternatively, a single point of light can be projected. This would prove impractical in generating a complete surface. More commonly, a line segment of light is projected (e.g., Hierholzer and Frobin, 1983). However, unlike the active industrial systems, a solid line segment is not desirable in a passive system when viewing a complex surface. This is because a passive system requires the correlation of multiple images. Therefore, a solid line would be useful only if the end points of the segment were visible in the views. Guaranteeing that the end points were always visible would be a stringent limitation to clinical usefulness when applied to the face.

In order to address this problem of being able to match corresponding points from a stereo pair, the line segment projected is composed of a series of dots. In this manner, the image becomes a sequence of distinct features.

To create the line of dots, a single low-power laser beam is used as the light source. The laser beam is focused and spread into a line using a collimator and cylindrical lens. This laser line is then passed through an opaque barrier containing small holes. The laser provides a sharp intensity peak for correlating the dots on the images. This was chosen due to the requirements that the line of dots be long enough to cover the area of interest, that all the dots be bright enough to be distinguished, and that the dots produce the desired resolution at the face.

Note that the closer the barrier is to the subject, the finer is the resolution and the clearer the dots are defined. However, the dot-generating equipment must not block the facial surface. Therefore, a front-faced mirror is used to change the direction of the laster beam. This allows the laser to be placed out of the cameras' views, and the mirror and barrier are made such that their width does not interfere with the cameras' field of view. Using this technique, a line segment 140-mm long consisting of 80 dots is generated on the X-Y plane (Figure 1) of the object space.

The laser line of dots is projected onto the subject and photographed. Now, every photograph consists of a series of light spots on a dark background; each spot represents either a diode or a point on the face. Each stereo pair of photographs provides enough information to re-create a profile slice of the subject. Multiple photographs with the beam scanning the object are necessary to create surface information.

The film is developed and then digitized. The digitizing is done by an automatic scanning image analyzer. To resolve the data dots, the analyzer used has 256 gray levels and a resolution of 1024 picture elements (pixels) across the 35-mm slide (Keefe, 1985).

Each dot must be given a unique location in the image coordinate system. The spots are initially found by searching for edges between light and dark areas. One cannot use a global threshold filter and dark areas to distinguish spots; thus, local intensity minimums are used to separate spots. Because we are projecting laster light that has been spread by a cylindrical lens into a line in the y-direction, the intensity distribution projected, although generally symmetrical over the entire y range, would be asymmetrical for a particular spot. Therefore, in order to assign a single x,y value to a spot, a weighted intensity center rather than a geometrical center is calculated.

In this manner, each dot on the slide film is given a single x,y coordinate. Also, based on spot size and intensity, a preliminary separation between diodes and laser spots is performed.

The data are referenced to the digitizer reference bar in that view's calibration shot. Then, the left and right view of the stereo pair must be correlated. This is possible with the calibration and headframe because the number and order of the diodes are known. The laser dots, however, are not as easily correlated.

The intensity of the reflected light varies due to how uneven the surface is; sharp angles spread the energy over a larger area. Also, it is possible that a data point may be hidden from one view.

In order to correlate the data from a stereo pair, the film is analyzed by comparing the spacing, or difference in vertical coordinate, between adjacent data spots. This spacing should be approximately uniform because the holes in the data-generating barrier are spaced uniformly. If the difference between any two adjoining spots is too large, then it is assumed that data are missing. The number of points missing is determined based on the average spacing. This missing area is flagged in both views and not used in subsequent calculations.

The data point closest to the diodes representing the top of the headframe is then found in both views.

The distance on the film between that data point and the headframe, along with the distance between the headframe and the digitizer reference bar, are used to indicate a common start of the data line in each of the two views. The dots from a stereo pair are then matched one to one, to to bottom from this point. No extrapolation is performed.

In applying the DLT procedure to the set of stereo pairs, a minimum of two sets of stereo pairs is required: left and right calibration, and left and right data. The *X, Y, Z* coordinates of the headframe are then used to orient the calculated three-dimensional coordinates of the line(s) of facial-data dots. Any number of lines of facial-data dots can be stored together, thus generating the three-dimensional surface information of the subject.

At this point, computer graphics is employed to transform the *X, Y, Z* coordinates into a "picture" of the surface. The data stored are composed of "vertical" profile lines where horizontal is parallel to the ear-ear segment. Cubic splines can be generated for each line and the raw data, as photographed, can be drawn. To create a surface mesh, horizontal profiles are generated using the existing vertical profiles. Evenly spaced planes, perpendicular to the average vertical profile plane, are generated. These planes are then intersected with the cubic spline curves from the vertical profiles, thus generating a set of horizontal profile lines. The same technique can then be applied to this new horizontal profile set to generate another vertical profile set. A horizontal and vertical set of profile lines can be drawn together using cubic splines to generate a bicubic surface.

To better visualize the surface information, a software link was provided between the system-generated three-dimensional coordinates and a grapics display system. This gives the user a powerful tool for manipulating and viewing the surface information.

The described data gathering equipment should not inhibit the widespread clinical use of the system, thus encouraging the development of a large database of facial form. However, this equipment still provides the desired common orientation plane and the possibility of automatically solving the stereo correspondence problem. Computer graphics will then provide the capability of graphically accessing the digital facial data.

ERROR

As with any photogrammetric set up, the ratio of object size to image size determines system scaling (ratio *D/C*, Figure 1). Camera separation (*B*) is then a function of required image overlap. To minimize the error, the object distance (*D*) should be as small as possible. However, for stereo correspondence, a point must be visible in both the views. To increase this certainty, for a fixed camera separation (*B*), the object distance (*D*) should be increased. This system compromises by choosing the largest value for the object distance (*D*) which still gives acceptable theoretical error.

To calculate this error, let a point be located at the origin of the *X, Y, Z* reference of Figure 1, i.e., at the central point, and assume symmetry. A relationship between *X, Y, Z* and *x, y* and the system parameters *D, B, C* can be geometrically derived. Object-space error and the image-space error are calculated as (Hallert, 1960)

$$mX = (D/C)*mx,$$
$$mY = (D/C)*my, \text{ and}$$
$$mZ = (2*(D/C)/(B/D))*mx$$

where mx, my = image space coordinate error, and mX, mY, mZ = object space error.

Due to the discrete nature of the image analyzer, one can get an idea of the image error (mx,my). The analyzer will not "notice" that a spot has moved until the energy has changed enough to cause one of the discrete outputs. This makes sensitivity to change a function of how rapidly the actual distribution changes and the size of the spot (plus concerns about how fast the image is traversed and the recovery time of the photo-sensitive element). A spot with infinitely small width could be anywhere inside the pixel and still generate the same output. Thus, one could assume that error would be at most half a pixel in x and y. Given knowledge of image-to-digital resolution plus acceptable error, system geometry (and, thus, camera parameters) can be defined.

Using that approach, for this system, the values for *D*, *B*, and *C* are 1.5 m, 0.30 m, and 0.090 m. Image error equals one-half the pixel size or (0.5)*(36mm/1024), about 0.018 mm.

Note that, from the error equations, a stereophotogrammetric system is most sensitive to image errors in a direction parallel to the line joining the camera image centers (x-axis of Figure 1). Thus, to minimize the effect of correlation errors, the line of dots is projected perpendicular to that axis.

Because the major component of the previous error analysis was in the *Z* direction, that error should be examined more closely. By assuming symmetry, one developed a maximum *Z* error, but because digitizer error should be random, that assumption is too restrictive. By allowing asymmetry in error about the central point, the previous results can be modified (Karara, 1972).

From the geometry in Figure 2, using similar triangles, one can say

$$(X_2/(x_2\text{-}x_0)) = (X_1/(x_1\text{-}x_0)) = D/C$$

where

$$H/a = C/x_1,\ H/b = C/x_2, \text{ and } x_0/C = (B/2)/D.$$

Rearranging the two equations involving H and using the relation

$a + b = X_1 + X_2$ yields

$$X_1 + X_2 = (H/C)*(x_1 + x_2).$$

Substituting for X_1, X_2 and then for x_0 yields

$$H = D*(x_1 + x_2 - (B*C/D))/(x_1 + x_2).$$

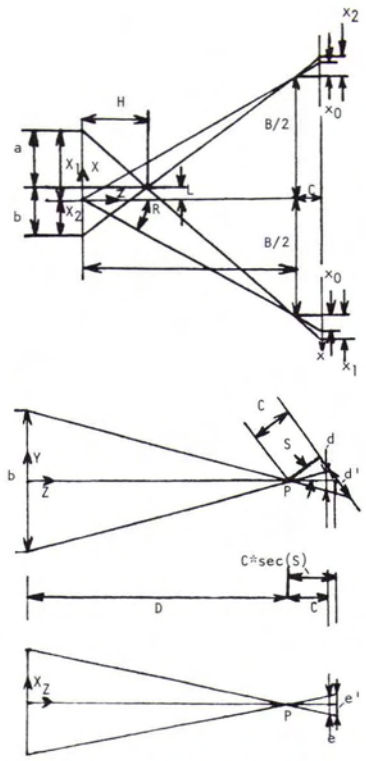

FIG. 2. Error variables and rotation effect.

Taking the derivative,

$dH/dx_2 = dH/dx_1 = (D/C) / (B/D)$.

Applying propagation of error yields

$mH\ (=mZ) = ((D/C)/\ (B/D))$*SQRT$((mx_1)$**2 + (mx_2)**2).

If one assumes that $mx_1 = mx_2 = mx$ (the usual case), then

mZ = SQRT(2)*$((D/C)/(B/D))$*mx.

Similarly, for the X direction in Figure 2, one also has

$L = X_1 - a$.

Applying values for X_1 and a, then for x_0 and H, gives

$L = (D/C)$*$(x_1 - (B$*$C)/(2$*$D))$

$-(D$*$x_1)$*$(x_1 + x_2 - ((B$*$C)/D))/(C$*$(x_1 + x_2))$,

which simplifies to

$L = (B/2)$*$(x_1 - x_2)/(x_1 + x_2)$.

Taking the derivatives

$dL/dx_1 = (B$*$x_2)/((x_1 + x_2)$**2) and

$dL/dx_2 = -(B$*$x_1)/((X_1 + x_2)$**2).

Assuming the error is small, both x_1 and x_2 are approximately x_0 or $(B$*$C)/(2$*$D)$:i.e.,

$dL/dx_1 = D/(2$*$C)$ and $dL/dx_2 = -D/(2$*$C)$.

Applying propagation of error yields

$mL\ (=mX) = (D/(2$*$C))$*SQRT$((mx_1)$**2+(mx_2)**2).

Again, if $mx_1 = mx_2 = mx$, one gets

$mX = (D/C)$*mx*(1/SQRT(2)).

For the Y direction, one gets two different y values, producing two distinct Y's, i.e., we have more equations than unknowns. Because we are using a least-squares technique, the Y chosen would be the one minimizing the distance to the two distinct computed values.

Therefore,

$Y_1 = (D/C)$*y_1 and $Y_2 = (D/C)$*y_2

where $Y = (Y_1 + Y_2)/2$.

Using the previous approach will then give

$mY = (D/(2$*$C))$*SQRT$((my_1)$**2+(my_2)**2),

or, if $my_1 = my_2 = my$,

$mY = (D/C)$*my*(1/SQRT(2)).

If we denote our earlier results as $mX[s]$, $mY[s]$, and $mZ[s]$ ([s] for symmetric), then one can see that

$mX[s]$ = SQRT(2)*mx,

$mY[s]$ = SQRT(2)*mY,

$mZ[s]$ = SQRT(2)*mZ, and

$mT[s]$ = SQRT(2)*mT

(where mT = SQRT$((mX)$**2+(mY)**2+$(mZ$**2)).

The expected error changed as a function of the square root of the number of images being processed to obtain the data (assuming the mx, my of each image are equal). Also,

mZ/mT = SQRT(2/(2 + (B/D)**2)).

For this system, $D/B = 5$. Therefore, mZ accounts for 99 percent of the total error.

Abdel-Aziz and Karara (1974) expanded the error equations to include a rotation of the camera about the projection center. This was accomplished by calculating the relationship between the rotated and original camera coordinate systems and applying the previous error analysis results.

Abdel-Aziz and Karara (1974) ignore the multicamera effect on X error and Y error by assuming the worst case. This turns out to have only a small effect on the total error, as in most photogrammetric setups mZ dominates the error. Also, the geometry of this particular setup makes the system relatively insensitive to small camera rotation prior to calibration.

This theoretical error can be checked experimentally in the following manner. After a complete data gathering session, the calibration frame is re-photographed as a system check. The pair of calibration slides photographed prior to the data gathering are used to calculate the DLT coefficients. The pair of calibration frame slides taken at the end can then be used with those DLT coefficients to yield the calculated, X, Y, Z coordinates of the calibration frame. The eight calibration diodes (taken two at a time) will yield 28 possible inter-diode distances. These interdiode distances allow comparison between the calculated and the measured calibration frame. Table 2 shows the mean, standard deviation, and maximum deviation for the inter-diode distance differences for a number of experiments.

After re-photographing the calibration frame in Test 1, the cameras were deliberately moved to see the effect. Remember that the calibration photographs at the beginning are used to determine the geometry of the setup. Should the geometry change, i.e., the cameras move, how does that change the

TABLE 2. ERROR: EXPERIMENTAL - CALIBRATION

Test	Mean (mm)	Standard Deviation (mm)	Maximum Deviation (mm)
0	0.446	0.336	1.179
1	0.292	0.287	0.990
1 (Y-Z)	1.769	1.696	5.677
1 (X-Z)	3.619	3.755	12.195
2	1.406	1.279	4.295
3	0.361	0.298	1.066
4	0.762	0.646	2.634

system error? First, the cameras were moved together by tipping the plate onto which they were mounted, i.e., rotating the cameras in a plane parallel to the Y-Z plane. After that, the cameras were brought back to near their original position, and one camera was tilted, that is, rotated in the X-Z plane.

By rotating the cameras, one is distorting the geometry of the calibration setup. Figure 2 also shows a rotation of one camera in the Y-Z plane. We are assuming, for simplicity, that the rotation is about the projection center, P. Because we use the digitizer reference bar for image orientation, it is the change in the distance between two points (which the rotation creates) that generates the error.

To calculate the error, let the camera be looking directly at an object of height b centered at point O. Let C be the camera's principal distance, d be the original image height, and d' the distorted image distance. By trigonometry, one can show (with $\delta = (d/2)$ that

$$d' = d^*((1+(\text{TAN}(S))^{**}2)/(1-(\delta^*\text{TAN}(S)/C)^{**}2)).$$

Therefore, one has introduced a y error at the image of

$$d'-d = d^*(1+(\delta/C)^{**}2)^*(\text{TAN}(S))^{**}2/(1-(\delta^*\text{TAN}(S)/C)^{**}2).$$

One can use similar triangles to calculate the x error at the image by letting e be the original image width and e' the distorted width:i.e.,

$$e' = e^*(\text{SEC}(S)), \text{ and } e'-e = e^*(\text{SEC}(S)-1).$$

This gives the distortion to the distances from rotation.

Because the camera setup was chosen to fill up most of the image in the y-direction, and in order to keep the necessary information on the slide, only a two-degree rotation could be allowed. Now, looking at the slide containing the calibration information, and including the digitizer reference bar, the maximum y image distance is about 15 mm and the maximum x image distance is about 10 mm. Putting these values in the above equations yields

$d'-d = 0.0184$ mm, and $e'-e = 0.0061$ mm.

Assuming the digitizer error would still be half of a pixel, one can add the above distortion in to get

mx = 0.0237 mm, and my = 0.0360 mm.

Putting these values into the theoretical error equations gives a theoretical error mT of 2.84 mm for one point, or possibly 5.69 mm for the distance between two points.

A similar analysis can be used for a rotation in the X-Z plane. For a 4.5 degree rotation of the one camera, the x error is affected the most and becomes $d'-d = 0.0621$ mm, and the y error is $e'-e = 0.0464$ mm.

Adding the digitizer error now gives

$mx = 0.0797$ mm, and $my = 0.640$ mm.

However, only one camera was rotated in the X-Y plane. That means the one view was distorted while the other was essentially not distorted. Therefore, mx and my for the non-rotated camera would still be about 0.0176 mm. Using the equations for mX, mY, mZ that contain mx_1, mx_2 and my_1, my_2 yielded a theoretical error, mT, of 6.87 mm for one point, or possibly 13.74 mm for the distance between two points.

Table 2 confirms a marked effect in the error should the cameras be moved after the calibration frame is photographed, with maximum distance error close to the expected value. By using the post-session calibration photographs, the researcher can determine if the data gathered are good. For example, one might suspect that the tripod was possibly bumped during Test 2, and, until the system is recalibrated, the data gathered would be suspect.

Therefore, a person gathering data should intersperse the data gathering with calibration shots, especially during a long session. This will allow the system to be recalibrated if necessary and minimize bad data due to the cameras being moved by accident.

The second technique to measure experimental error was to use the system on known geometric objects and compare the calculated results to the known description of the object. The following profiles were generated: a line (projection on a plane), two lines (projection on two planes intersecting at 89.93 degrees), and a circle (projection on a hemisphere with a 50.84 mm radius). Note that all the projections are planar. The procedure was to take the system-generated data and use a least-squares approach to find the best plane that fit the data. Then, in that plane, the data were compared again, in a least-squares sense, to either a line or a circle. Both the angled object and the circle were photographed twice.

Table 3a shows the mean, standard deviation, and maximum deviation of the distances from the data points to the best-fit plane for the data representing the known objects.

This is extremely good, but one would expect good results. Due to the way the laser-dots are projected, the profile planes generated are parallel to the Y-Z plane of the calibration. Therefore, deviation from a plane would be an X error. One can see from the DLT error equations that most of the error would be expected in the Z dimension.

Now, one can calculate the mean, standard deviation, and maximum deviation of the distances from the data to the best-fit of what the data should represent. These results are shown in Table 3b.

These results are comparable to the experimental

TABLE 3. (A) ERROR: EXPERIMENTAL - PLANAR (X,Y)

Object	Mean (mm)	Standard Deviation (mm)	Maximum Deviation (mm)
Line	0.0417	0.0250	0.107
angle (1)	0.0451	0.0317	0.179
angle (2)	0.0385	0.0331	0.143
circle (1)	0.0334	0.0229	0.0969
circle (2)	0.0334	0.0221	0.0952

(B) ERROR: EXPERIMENTAL KNOWN OBJECTIVES

Object	Mean (mm)	Standard Deviation (mm)	Maximum Deviation (mm)	Angle or Radius	Correlation to a Line
Line	0.190	0.213	0.842	na	0.997
angle (1)				88.95°	
1st line	0.250	0.214	0.900		−0.9997
2nd line	0.198	0.183	0.776		0.9998
angle (2)				88.35°	
1st line	0.213	0.146	0.589		−0.9998
2nd line	0.147	0.120	0.530		0.9999
circle (1)	0.474	0.207	0.997	50.25mm	na
circle (2)	0.826	0.414	1.727	50.08mm	na

error calculations of Table 2. In all these experiment, 80 dots were projected that covered about 140 mm on the object. For the circle, only about 50 dots could be resolved. Not only was the diameter of the circle less than 140 mm, but also the surface bends away from the camera. This has two effects. The first is that the size of the dots and spaces increases, thus decreasing resolution. Second, the angle of reflection increases, making it difficult to locate the dots. By analyzing the data from the experiment involving the hemispherical surface, acceptable results were achieved until the dots could no longer be separated from the background noise. That occurred when the surface normal was off horizontal by about 60 degrees.

One can use the "good" sessions from Table 2 and Table 3b to calculate an average experimental error. (The error from Table 2 is divided equally between two points as the calibration distances involve the distance between two calculated positions.) Expecting most of the errors to fall within two standard deviations of the mean gives an experimental error of 0.71 mm.

Error is a function of geometric scale, inaccuracy in the object space definition, and inaccuracy in the image space coordinates. Also, one must deal with the stereometric assumptions of a perfect film plane and of a central projection, i.e., film flatness and lens distortion. Because we are looking at a relatively large object, when compared to the film format, this system's geometrical scale makes the most significant source of error the inaccuracy in the image space coordinates: specifically, the error in the x-direction. For example, given the previously developed error equations, the X, Y, Z errors in defining the object space (i.e., measuring the calibration frame) could produce an image error of at most 0.0015 mm, which is approximately the film's grain size and an order of magnitude less than the pixel size. Also, solving the non-linear equations for 12 coefficients, i.e., including a model for lens distortion (Marzan and Karara, 1975), improved the error mean less than 0.04 mm. If one were to look at smaller object spaces, then it might be useful to incorporate more complex models to include film flatness or lens distortion.

The error in the image coordinate is a function of the data intensity distribution and the sensitivity of the digitizer. Our system gave an experimental error of mT equals 0.71 mm. Using the equation for mT, and assuming $mx = my$, this corresponds to a digitizer error, mx, of 0.006 mm, about one-sixth of a pixel. The maximum error, mT, of 1.73 mm, corresponds to a digitizer error of 0.014 mm, or 80 percent of the expected half-pixel maximum digitizer error.

Vertical resolution is the dot spacing along a projected line of data. The resolution in the other dimension is the nominal distance between projection lines. This is an arbitrary number depending on the number of profiles used by the researcher to generate the surface. Figure 3 shows photographs of a plaster facial moulage. For the same moulage, Figure 4 shows, from upper left, the system generated mid-line profile next to an actual tracing of that profile, the nine vertical profile lines used to generate the surface, a wire-frame of the bi-cubic surface, a contour map of the surface (about 2.5 mm between contours), and a shaded surface created from the system-generated data. One can see the loss of detail going from the nose to the ear (nine profiles) compared to the surface from the forehad to the chin (eighty dots). The ability of the computer-generated output to portray the face can be clearly seen however. Also, the surface is now stored in a digital manner and can be readily manipulated.

CONCLUSION

The automated system described can, based on the experimental data analyzed, generate three-dimensional surface information. The experimental accuracy of the data had a 0.29-mm mean, a 0.21-mm standard deviation, and a 1.7-mm maximum deviation. The results can be theoretically accounted for primarily by error in the image space. Present resolution at the face is, at best, 1.75 mm. The object must be visible to the recording instrumentation, and there is some evidence that the reflectivity of the object has an effect on the system. Preliminary investigation indicates that, to be visible, the reflection angle (measured from the line connecting the principal point to the projection center) must be less than 60 degrees. Further experiments would be necessary to better study the effects of reflectivity and surface orientation.

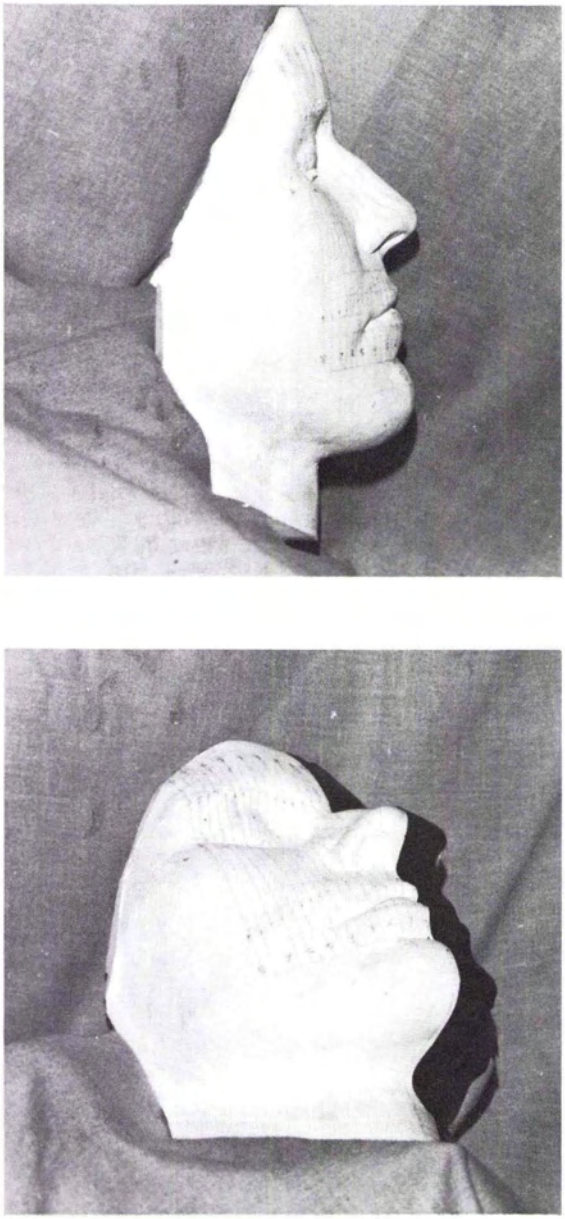

FIG. 3. Plaster moulage, actual surface (two views).

The system generates a true digital facial surface. By photographing a series of dots in a line, automatic correlation becomes realistic. By projecting a line and using multiple profile slices to create the surface, this system provides a bridge between the large amount of historical two-dimensional studies and the development of true three-dimensional studies. Computer graphics gives the researcher a powerful tool for manipulating and viewing the surface information.

It is hoped that this system provides a methodology useful toward the development of a large facial database and thus permits creation of needed three-dimensional cephalometric theories. The key is the automatic clinical acquisition of the three-dimensional data.

More generally, a passive stereo technique has been created which uses structured light to permit a relatively simple solution of the stereo correspondence problem without using subjective human interpretation of the stereo pair.

REFERENCES

Abdel-Aziz, Y. I., and H. M. Karara, 1974. *Photogrammetric Potential of Non-Metric Cameras*, Civil Engineering Studies, Photogrammetry Series No. 36, Urbana, University of Illinois, Department of Civil Engineering.

Agin, G. J., and P. T. Highnam, 1982. Movable Light-Stripe Sensor for Obtaining Three-Dimensional Coordinate Measurements, *Robotics and Industrial Inspection*, D.P. Casasent (ed.), SPIE, Vol. 360, pp. 326–333.

Altschuler, M.D., J. L. Posdamer, G. Frieder, B. R. Altschuler, and J. Taboada, 1981. The Numerical Stereo Camera, *3-D Machine Perception*, SPIE, Vol. 283, pp. 15–24.

Burgam, P.M., 1984. Chrysler Canada Adds Vision for Minivan Body Fit, *Society of Manufacturing Engineers*, Vol. 92, No. 2, pp. 49–52.

Duncan, J. P., P. M. Wild, and P. M. Hoemberg, 1984. *Automatic Metrology and Machining of Arbitrary Surfaces*, SME Paper No. MS84-945 1984.

Hallert, B., 1960. *Photogrammetry*, McGraw Hill, New York.

Herron, R. E., 1972. Biostereometric Measurement of Body Form, *Yearbook of Physical Anthropology*, Vol. 16, pp. 80–121.

Hierholzer, E., and W. Frobin, 1983. Automated Measurement of Body Shapes using Rasterstereography, *Biostereometrics 82*, R. E. Herron (ed.), SPIE, Vol. 361, pp. 125–131.

Jarvis, R. A., 1983. A Perspective on Range Finding Techniques for Computing Vision, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 5, No. 2, pp. 122–139.

Karara, H. M., 1972. Simple Cameras for Close-Range Application, *Photogrammetric Engineering*, Vol. 38, No. 5.

Keefe, M., 1985. *A Computer-Assisted Tool to Semi-Automatically Generate Three-Dimensional Facial Soft-Tissue Information*, Ph.D. Thesis, University of Minnesota.

Manual of Photogrammetry, 4th edition, Chester C Slama, editor, American Society of Photogrammetry, Falls Church, Virginia, 1980.

Marzan, G. T., and H. M. Karara, 1975. A Computer Program for Direct Linear Transformation Solution of the Collinearity Condition, and Some Applications of It, *Proceedings of the Symposium on Close-Range Photogrammetric Systems*, Fall Curch, Virginia, American Society of Photogrammetry, July 28-August 1.

Moyers, R. E., and F. L. Bookstein, 1979. The Inapprpriateness of Conventional Cephalometrics, *American Journal of Orthodontics*, Vol. 75, No. 6, pp. 599–617.

Panton, D. J., 1978. A Flexible Approach to Digital Stereo Mapping, *Photogrammetric Engineering and Remote Sensing*, Vol. 44, No. 12, pp. 1499–1512.

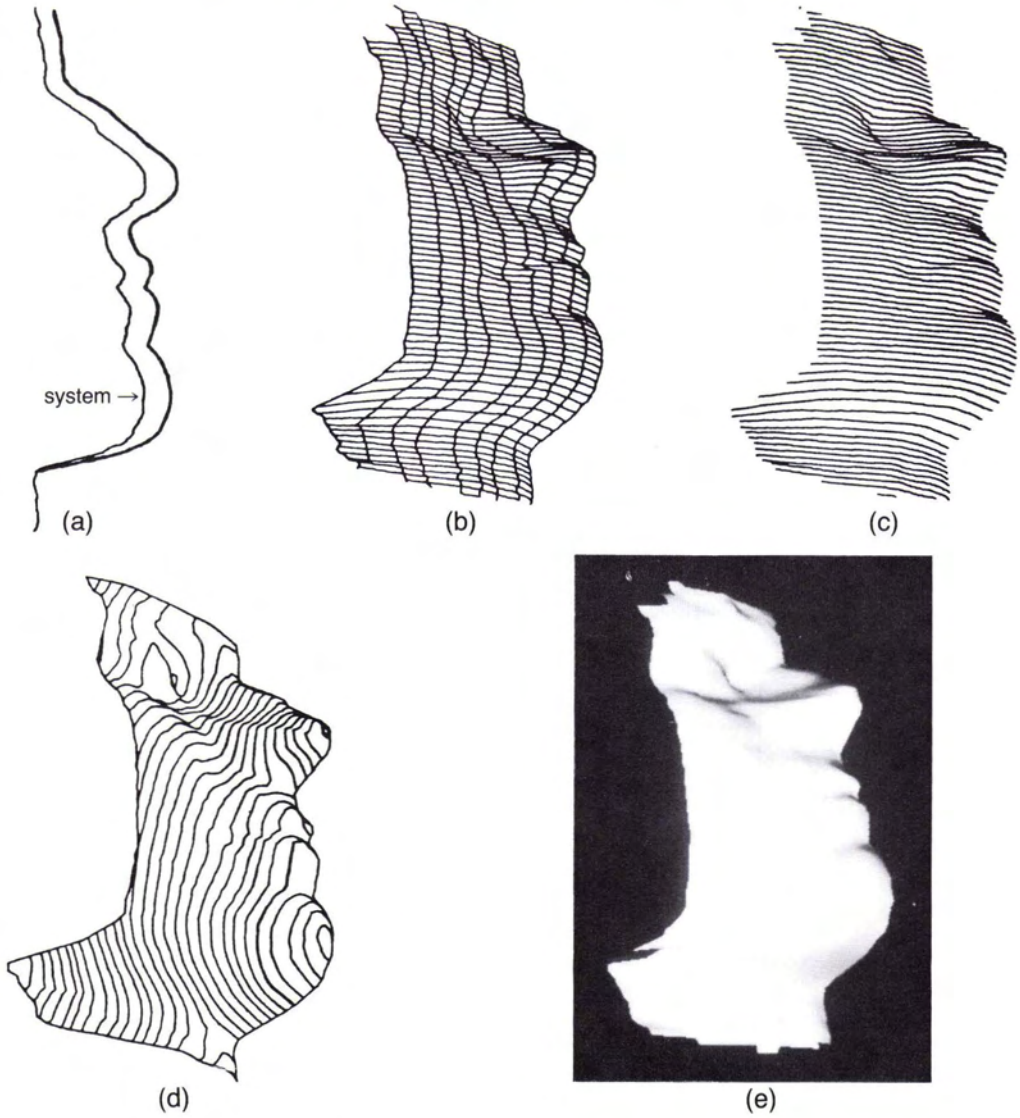

FIG. 4. Plaster moulage, system generated surfaces: (a) System tracing, (b) nine vertical profiles, (c) bi-cubic surface, (d) contour surface, (e) shaded surface.

Rogers, D. F., and J. A. Adams, 1976. *Mathematical Elements for Computer Graphics*, McGraw Hill, New York.

Saunders, C. G., 1983. Reconstruction of Anatomical Shapes from Moire Contourographs, *Biostereometrics 82*, R. E. Herron (ed), SPIE, Vol. 361, pp. 99–106.

Soares, O. D. D., 1983. Moiré Techniques in Biomechanical Case Studies, *Biostereometrics 82*, R. E. Herron (ed.), SPIE, Vol. 361, pp. 92–98.

Yatagai, T., S. Nakadate, M. Idesawa, and H. Saito, 1982. Automatic Fringe Analysis using Digital Image Processing Techniques, *Optical Engineering*, Vol. 21, No. 3, pp. 432–435.

Young, J. M., and B. R. Altschuler, 1981. Topographic Mapping of Oral Structures Problems and Applications in Prosthodontics, *3-D Machine Perception*, SPIE, Vol. 283, pp. 70–77.

(Received 8 November 1985; accepted 29 January 1986; revised 26 February 1986)