Calibration of a Video Rasterstereographic System

E. Hierholzer

Abstract

Rasterstereography requires special methods of calibration in order to make use of the well-known photogrammetric procedures. The basic methods have already been published in previous papers. In the particular case of video rasterstereography, it is possible to collect the calibration data from a sequence of several video frames, provided that pixel-synchronous image digitization is used. By this technique, automatic image processing is considerably facilitated. In the present paper the procedures for recording and analyzing of the calibration images (using a control point system with changeable control planes) are described. Some results for a prototype system are presented.

Introduction

Video rasterstereography is a modification of rasterstereography (Frobin and Hierholzer, 1981), using a high-resolution solid-state video camera for recording of the projected raster lines. Such a system for measurement of the human back surface has been described in a publication by Frobin and Hierholzer (1991). The basic procedures for automatic processing of the video image are quite similar to those used in conventional rasterstereography (using film recording and scanning (Frobin and Hierholzer, 1983a; 1983b). The same holds true for the calibration procedure (Frobin and Hierholzer, 1982a; 1982b); however, various improvements are possible by using the video technique. A detailed discussion of video technology in photogrammetry is given by El-Hakim *et al.* (1989).

Although rasterstereography is basically equivalent to common stereophotogrammetry, there are considerable differences in the respective methods of calibration, in particular, in the initial steps of image data analysis. Some preliminary considerations and the special design of a control point system suitable for video rasterstereography have already been outlined in Frobin and Hierholzer (1991). Using a video camera not only facilitates and accelerates the surface measurement, but also has considerable advantages for the calibration procedure.

A very important aspect is that in a solid-state video camera no fiducial marks are necessary, because the pixel matrix itself defines a fixed image coordinate system. To take advantage of this fact, it is necessary that the video signal be pixel-synchronously digitized, and that the lens be mounted in a fixed position relative to the image sensor. If such a camera is used, several video images, which have been sequentially recorded and digitized, may thereafter be exactly superimposed by computation. This makes it possible to split up the recording of the control point system into a sequence of partial images, each of them containing only a fraction of the necessary calibration information. Thus, image processing is considerably simplified. In contrast to this, in conventional film rasterstereography a rather complex image has to be analyzed (see, e.g., Figure 1 in Frobin and Hierholzer (1983a)).

In rasterstereography, the projector is considered as an inverse camera. Thus, the recording system is equivalent to a stereo camera pair, and the image pair consists of the video image and the raster diapositive. Conventional photogrammetric methods of calibration and model reconstruction can, therefore, be applied. In the calibration procedure, however, a difficulty arises from the fact that the images of the control points in the projector image plane (i.e., in the raster diapositive) cannot be measured. As has been shown by Frobin and Hierholzer (1982b), these images can be calculated by interpolation between the raster line images in the camera image plane. In other words, the projector image of the control point system can be determined from data measured solely in the camera image. By this procedure, image data are converted into a conventional stereo image pair, and standard calibration procedures such as bundle adjustment can be applied.

In the present work, recording and analysis of the calibration images is described. In addition to the 2 by 9 elements of interior and exterior orientation of the camera and the projector, the lens distortions and the x and y pixel spacings of the sensor (more exactly, the ratio of the x and y spacings) are determined. The raster diapositive and the image sensor are assumed to be free of distortion.

Recording of the Control Point System

Figure 1 shows the control point system mounted in the proper position and orientation relative to the video camera and the projector. It consists of a solid base of cast aluminum onto which control planes can be mounted at six different positions: i.e., at two different height levels ($z = 0$ mm and $z = 120$ mm) and at three lateral positions ($x = 0$ mm and $x = \pm 200$ mm). On each plane seven control points are located at $y = 0$ mm, ± 125 mm, ± 225 mm, and ± 325 mm. Thus, a total of 42 control point locations is distributed in a volume of 400 by 650 by 120 mm^3 (W by H by D) which approximately corresponds to the lateral, longitudinal, and sagittal dimensions of the human back surface. Using a camera with a standard TV lens (f:1.4/16 mm), the control point sys-

Institut für Experimentelle Biomechanik, Universität Münster, Domagkstrasse 11, D-4400 Münster, Germany.

Photogrammetric Engineering & Remote Sensing, Vol. 60, No. 6, June 1994, pp. 745–750.

0099-1112/94/6006–745$03.00/0

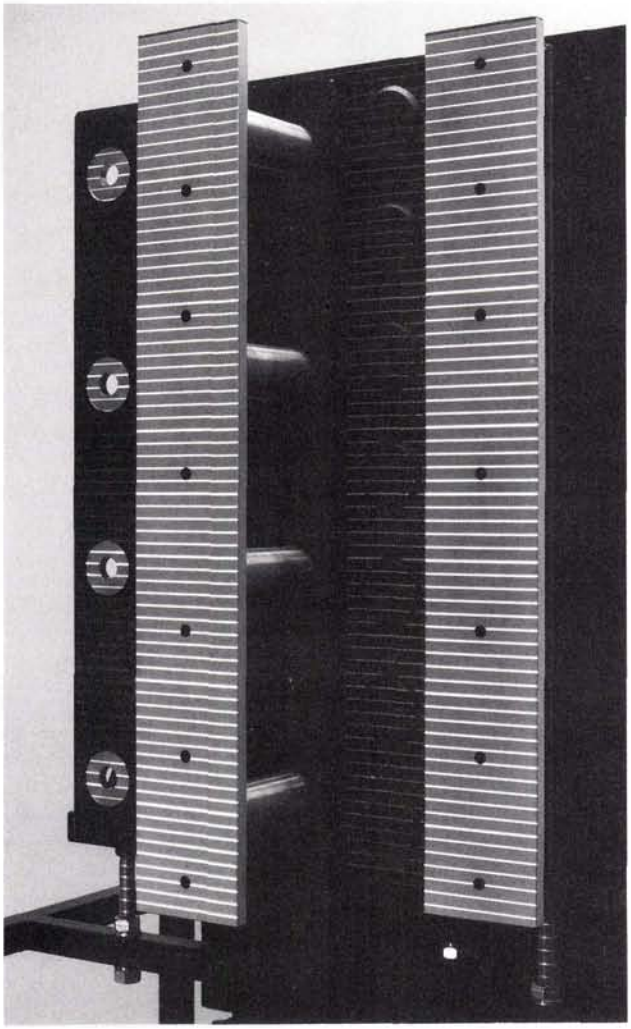

Figure 1. Control point system for video rasterstereography. Two control planes with seven control points are mounted in two (of six possible) positions.

tem is imaged on a Valvo NXA 1011 sensor with 576 (*H*) by 604 (*V*) pixels. The camera is used in upright position (portrait format) which is appropriate for the human back. According to the specifications of the manufacturer, the pixel spacing is 7.8 μm (*H*) by 10.0 μm (*V*). The accurate spacings are, however, determined as additional parameters in the calibration procedure.

In principle, 12 video images of the control point system are necessary: for any of the six possible control plane positions, two images must be recorded with the raster projector switched on and off, respectively. Because two control planes (for the high and low position) have been made and can be mounted at the same time (Figure 1), this reduces the number of frames to capture to six recordings of three different control plane constellations. By interactive image manipulation (see next section), the respective image region for each control plane position is isolated, resulting in a total of 12 partial images to be analyzed. During the whole recording procedure, special care must be taken not to move the base of the control point system relative to the camera and projector. The fittings of the control planes are prepared in a way to allow an easy and very exact interchange.

The six images recorded with the projector switched off (but with diffuse environmental illumination) enable calibration of the camera in the usual way. However, in a preliminary step the image coordinates of the control point centers must be determined and combined for all the 6 by 7 control points. From the remaining six images recorded with the projector switched on, additional information for calibration of the projector system is obtained. The basic principle of this method has been described in Frobin and Hierholzer (1982b). The video method allows, however, an improvement by fully utilizing the information content of the projected raster lines.

Image Processing

Measurement of the Control Points in the Camera Image

The image coordinates of the control points in the sensor plane of the video camera are determined from the video images with the raster projector switched off. In these images the control points are identified and their approximate locations are marked interactively by moving a cursor in the monitor image (Figure 2). Alternatively, an automatic search-

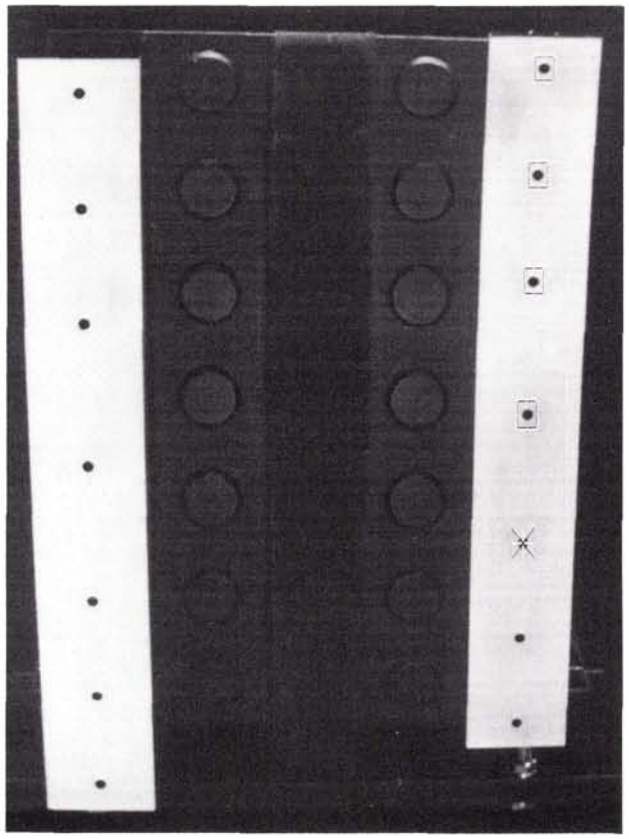

Figure 2. Video image of control points (raster projector off). The control point locations are interactively marked for image processing (right control plane, low position).

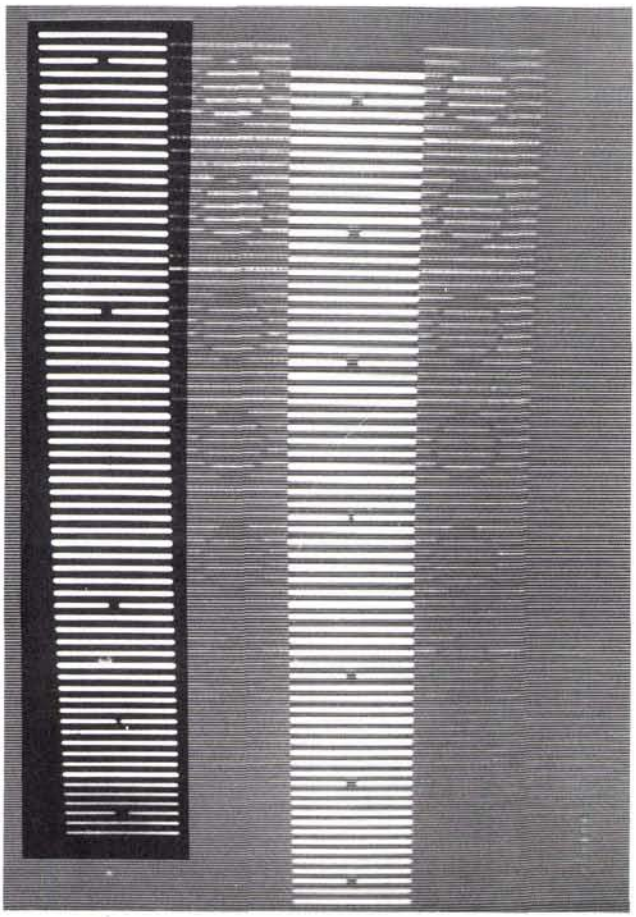

Figure 3. Video image of control planes (raster projector on). The left control plane (low position) has been selected for image processing (unshaded region).

ing algorithm could be implemented. However, this might require an unreasonable programming expense if arbitrary positions of the camera relative to the control point system are admitted.

The exact location of each control point is determined by placing a window of appropriate size around it. Using standard image processing algorithms such as erosion and dilatation, the black control point areas in the windows are isolated. After verification of the circular shape (testing of the second central moments of the area), the camera image coordinates are calculated from the first moments (centroid) of the control point area. The method has been described in detail in Frobin and Hierholzer (1983a).

The procedure is repeated for all control plane positions, resulting in 42 image coordinate pairs (x_k,y_k). The primary coordinate data are given in units of the pixel spacing, which may be different in the x and y direction. If the pixel spacing is exactly known, the calculation of true image coordinates (in millimetres) is straightforward. Otherwise, the x and y pixel spacings must be introduced as additional parameters in the calibration procedure or determined independently (El-Hakim *et al.*, 1989). The critical parameter is, however, the ratio γ of the x and y pixel spacings because a change in the principal distance c_k can compensate for an error in overall scale. As discussed in the last section, an independent determination of the spacing ratio γ is easily possible in the present application.

Calculation of the Control Point Images in the Projector

In rasterstereography, the calibration of the projector system cannot be based on material images of the control points. Although the control points are imaged by the projector lens into the raster diapositive plane, they cannot be measured there. Nevertheless, their coordinates (x_p,y_p) in the diapositive plane can be indirectly determined from data measured solely in the camera image. The basic principle of the method has been outlined in Frobin and Hierholzer (1982b). It relies on interpolation between the raster lines, which are projected onto the control planes in the surroundings of the control points (e.g., lines i and $i + 1$ in Figure 4). For this purpose, the six video images with the raster projector switched on are analyzed. The video technique and the present design of the control point system enable a simplified and improved procedure which is described in the following.

In the first step, the raster lines in the camera image are measured, and they are identified by their ordinal numbers i. This task is carried out using the common methods of rasterstereographic image processing (Frobin and Hierholzer, 1983a; 1983b). For the sake of simplicity, the raster lines are measured separately for each of the six control plane locations. The video images are, therefore, clipped in order to isolate the respective regions of interest. For example, in Figure 3 the two control planes are located in the low position (left) and in the high position (center). The image region containing the low control plane has been interactively cut out and the shaded area is not considered in the subsequent image processing. This is repeated for all control plane positions. The data of the six partial images are subsequently merged in a single data file. Owing to pixel-synchronous image digitization, this is possible without introducing any additional error.

Next, straight lines are fitted to the data points belonging to any particular raster line. In the camera image coordinate system (x_k,y_k), these lines are represented by

$$y_k = a_0(i) + a_1(i)\, x_k \qquad (1)$$

where the intercepts $a_0(i)$ and the slopes $a_1(i)$ are dependent on the ordinal number i of the respective raster line.

The coordinates (x_p,y_p) of the control point images in the projector image plane can now be calculated in the following way. Owing to the high manufacturing accuracy of the control point system, the raster lines can be considered to be projected on only two (although interrupted) planes corresponding to the low and high position of the control planes. The projection is governed by the rules of projective mapping. The projected lines are, in turn, projected onto the video sensor. Because the result of two consecutive projections is another projective mapping, the raster line images in the video sensor plane may likewise be described by projective equations.

In particular, the original system of parallel raster lines is transformed into a pencil of lines passing through a common vanishing point V in the video sensor plane as shown in Figure 4 (separately for either control plane). The raster lines in the diapositive plane may likewise be considered as a pencil of lines with a vanishing point at infinity.

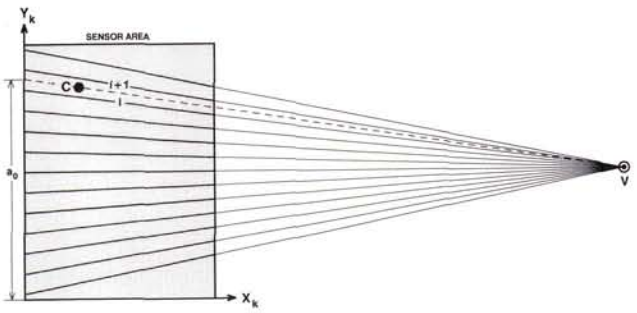

Figure 4. Determination of intercept a_0 (for Equation 5) of control point C using the vanishing point V (camera image plane). The heavy lines within the sensor area are the raster line images.

The lines of both pencils are completely described by their vanishing points and their intersections with the respective y axes. In the camera image plane the vanishing point V can easily be determined from the straight lines fitted to the raster line points (Equation 1). The intercepts $a_0(i)$ are functions of the ordinal number i: i.e.,

$$a_0(i) = f(i) \tag{2}$$

In the raster diapositive the intercepts b_0 are simply given by

$$b_0(i) = g \cdot i = y_p(i) \tag{3}$$

where g is the grid constant of the raster diapositive and $y_p(i)$ is the vertical coordinate of raster line i in the projector coordinate system (the x_p coordinate is assumed to run in the direction of the raster lines).

We now consider the y_p axis in the diapositive plane (i.e., the line $x_p = 0$) together with the intercept points $b_0(i)$ on it (Equation 3). This line is mapped onto the y_k axis in the camera image plane (Figure 4) with intercepts $a_0(i)$ according to Equation 2. The mapping function f is described by a linear projectivity which is mathematically represented by a fractional linear transformation: i.e.,

$$a_0(i) = \frac{A \cdot b_0(i) + B}{C \cdot b_0(i) + 1}. \tag{4}$$

Because the intercepts $a_0(i)$ can be measured in the video image (Equation 1) and the intercepts $b_0(i)$ are *a priori* known from the design of the raster diapositive, the unknown coefficients A, B, and C can be fitted in a least-squares procedure. For either control plane, a separate set of coefficients A, B, and C is obtained.

For any arbitrary line in the camera image passing through the vanishing point V, the corresponding line $y_p = b_0$ (parallel to the raster lines) in the raster diapositive can be calculated from the intercept a_0 in the camera image plane by solving Equation 4 for b_0: i.e.,

$$b_0 = \frac{B - a_0}{C \cdot a_0 - A}. \tag{5}$$

We now consider a control point C imaged onto the video sensor (Figure 4). A line passing through C and the vanishing point V (dashed line) may have the intercept a_0. The corresponding line in the projector image plane passes through the projector image of the control point. Hence, the desired projector image coordinate y_p of the control point is given by the intercept b_0 according to Equation 5.

This procedure allows only the determination of the coordinate perpendicular to the raster lines, i.e, the coordinate y_p, if a line raster oriented in the x_p direction is used. The coordinate x_p in the direction of the raster lines remains indeterminate. However, a sufficiently accurate estimation is possible from the camera image coordinate x_k. In the usual optical arrangement of rasterstereography, the raster lines and the x_p and x_k axes are oriented perpendicularly to the stereo base. In such a configuration the projector coordinate x_p is approximately proportional to the camera coordinate x_k: i.e.,

$$x_p \approx \alpha \cdot x_k. \tag{6}$$

The constant of proportionality α is given by the scale of the imaging of the video sensor plane into the diapositive plane (via the control planes): i.e.,

$$\alpha \approx (c_p/c_k) \cdot (\langle e_k \rangle / \langle e_p \rangle) \tag{7}$$

where c_p and c_k are the principal distances of the projector and the camera. $\langle e_p \rangle$ and $\langle e_k \rangle$ are the respective average distances to the control point system. In the standard geometry of rasterstereography, $\langle e_p \rangle$ and $\langle e_k \rangle$ are roughly equal; then, $\alpha \approx c_p/c_k$. The accuracy of the estimated x_p coordinate depends on the depth modulation of the object (in this case, the z range of the control planes). The rather low accuracy can be accounted for by an appropriately low weight factor in the bundle equations. This procedure is justified because one of the four stereo image pair coordinates $(x_k,y_k; x_p,y_p)$ is redundant in calibration and model reconstruction.

The above method of calculating the projector image coordinate y_p is superior to a simple local interpolation, as described in Frobin and Hierholzer (1982b), because it utilizes the information of the complete line pencil together with the exact projective Equation 4 instead of a local linear interpolation (e.g., between lines i and $i + 1$ in Figure 4). A minor disadvantage is that it is difficult to account for radial lens distortion, at least for the projector. This introduces an additional non-linearity in Equations 4 and 5. In the present application, however, a noticeable effect on the final results could not be observed when neglecting lens distortions at this stage.

In Figure 5, an integrated display of the results (in the video image plane) is given for the lower control plane. The control points are represented by black dots. About 6,500 raster line points are plotted together with their regression lines (these are actually smoothed data calculated from about 20,000 original raster points). The RMS deviation from the regression lines (averaged over all raster lines and control planes) is 0.56 μm, that is, 5.6 percent of the vertical pixel spacing. This figure confirms the importance of subpixel interpolation enabled by pixel-synchronous image digitization.

Calibration

Using the method described in the preceding section, stereo image pairs $(x_k,y_k; x_p,y_p)$ can be determined from the video images where x_k and y_k are measured directly, x_p is estimated according to Equation 6, and y_p is calculated from Equation 5. These coordinates are associated with weight factors reflecting the respective accuracies. According to common experience with subpixel interpolation (Raynor *et al.*, 1990) the localization accuracy of the control point im-

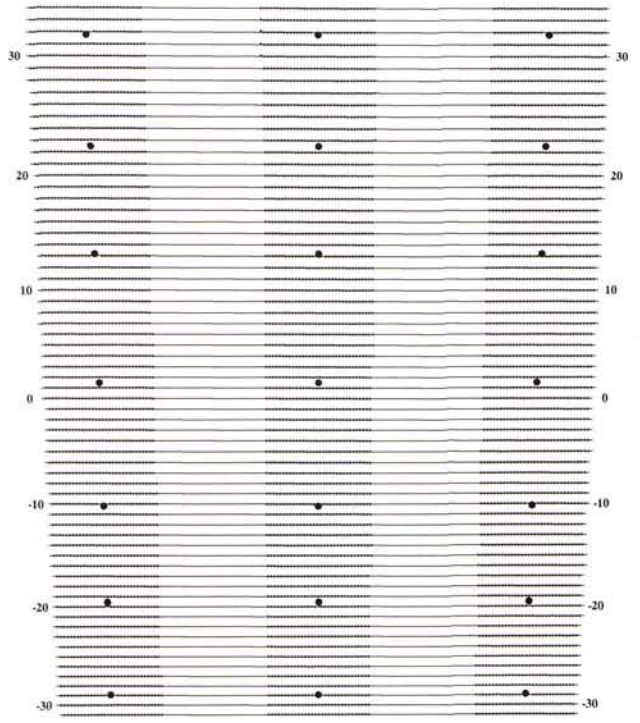

Figure 5. Results of automatic image processing (low control plane only); integrated data of six different video images.

big dots = control points
small dots = raster line points
solid lines = regression lines
numbers = ordinal numbers i of the raster lines

ages is (conservatively) estimated to 0.2 pixel, i.e., 1.5 μm in the x direction and 2 μm in the y direction for the Valvo NXA 1011 image sensor. The accuracy of the raster lines in the diapositive (coordinate y_p, Equation 3), as given by the manufacturer, is 5 μm. In contrast to these figures, the uncertainty of the coordinate x_p is in the order of millimetres, because only a rather inaccurate estimation based on the average scaling factor α is available, as discussed in the preceding section. The accuracy of the three-dimensional coordinates (x_c,y_c,z_c) of the control points is given by the mechanical precision of the construction. Is is estimated to be 0.03 mm for all three directions. Thus, the following linear weight factors (inversely proportional to the scatter) have been chosen for the calibration procedure:

$g(x_k) = 2/3 \qquad g(y_k) = 1/2$
$g(x_p) = 1/1000 \qquad g(y_p) = 1/5$
$g(x_c) = 1/30 \qquad g(y_c) = 1/30 \qquad g(z_c) = 1/30$

Effectively, the squares of these weight factors are used in the bundle equations.

The 2 by 9 orientation elements consist of the 2 by 3 elements of interior orientation (principal point coordinates and principal distance) and the 2 by 6 elements of exterior orientation (three orientation angles and three coordinates) of the camera and the projector, respectively. Three additional parameters have been taken into account: radial lens distortions β_k of the camera and β_p of the projector and the ratio γ of x and y pixel spacings of the image sensor. Although data for β_p and γ were supplied by the manufacturers, these parameters have been adjusted by minimization of the residual errors in the bundle adjustment. For the sake of simplicity, the additional parameters have not been introduced into the imaging equations themselves (see, e.g., Fryer (1989)). Instead of that, they have been adjusted independently by running the bundle adjustment program with different values β_k, β_p, and γ and searching for minimum residual errors. This procedure is justified by the fact that the parameters proved to be sufficiently independent from each other. No correlation of the optimized values was observed.

As a result, the parameters β_p and γ (as given by the manufacturers) were nearly exactly reproduced. No value for the camera distortion β_k was available, which is, however, less important because only a very narrow field of view of the camera lens (4.5 by 6 mm) is actually used. In addition, a somewhat irregular distortion - probably due to lens centering errors - was observed which could not be corrected for. The errors given below may therefore be improved by using a better camera lens.

With these data, the bundle method using the implementation of Frobin and Hierholzer (1982a; 1982b) has been applied. The apparatus had been adjusted in order to obtain a simple geometry (e.g., raster lines and projector axis horizontal, stereo base vertical, camera axis intersecting the projector axis, and so on). Thus, the necessary initial values for the orientation elements could be estimated rather accurately. About ten iterations were needed for final optimization (computing time less than 15 seconds on a 68030 computer).

The residual errors of the calibration procedure (depending of course on the choice of the relative weight factors) are listed in Table 1. All residual errors are within the expected order of magnitude. As discussed in the preceding section, the error of x_p is rather large because x_p can only roughly be estimated from x_k. The camera errors are about one-tenth of the pixel spacing. The control point errors refer to the actual three-dimensional coordinates and reflect the mechanical accuracy of the system. The larger error of the control point coordinate y_c is probably due to minor vertical movements of the control point system when interchanging the control planes. The very small residual error of z_c may result from the fact that the distance between the control planes is maintained by columns (see Figure 1, center) with particularly precise fittings. From the rather small control point errors, it might, however, be suspected that the assumed mechanical accuracy of 0.03 mm underestimates the actual errors and overemphasizes the respective weight factors in the bundle equations.

TABLE 1. RESIDUAL ERRORS OF THE BUNDLE ADJUSTMENT

device	coordinate	residual RMS error (μm)
projector	x_p	381.1
projector	y_p	4.0
camera	x_k	1.0
camera	y_k	0.9
control point	x_c	3.3
control point	y_c	8.0
control point	z_c	1.0

TABLE 2. ERRORS OF THREE-DIMENSIONAL MODEL RECONSTRUCTION

coordinate	RMS error (mm)	maximum error (mm)	mean deviation (mm)
x_c	0.138	0.305	−0.010
y_c	0.091	0.211	−0.003
z_c	0.225	0.532	−0.009

The actual accuracy of the surface measurement is best estimated by model reconstruction of the control points. The three-dimensional errors (i.e., the deviations from the nominal values) are listed in Table 2. The larger z_c model error is expected from the geometry of the system with a convergence angle of about 22°, resulting in a depth resolution which is (theoretically) 2.5 times lower than the lateral resolution. The mean deviation (last column) is equivalent to a linear displacement of the centroid of the reconstructed model with respect to the expected nominal value (0,0,0). It is listed only to show that no systematic deviation is present (e.g., in consequence of uncompensated lens distortions, etc.). The model errors are in the same order of magnitude as found in other applications of rasterstereography.

Conclusion

Calibration of a video rasterstereographic system is possible by using conventional methods of stereophotogrammetry, although only one half image (produced by the video camera) is available. The other half image (projector image) can be constructed using only data obtained from the video image. The results prove that an accuracy comparable to that of conventional close-range photogrammetry is attainable.

References

El-Hakim, S. F., A. W. Burner, and R. R. Real, 1989. Video technology and real-time photogrammetry, in *Non-Topographic Photogrammetry, Second Edition* (H. M. Karara, editor), American Society for Photogrammetry and Remote Sensing, Falls Church, Virginia, pp. 279–304.

Frobin, W., and E. Hierholzer, 1981. Rasterstereography: A photogrammetric method for measurement of body surfaces, *Photogrammetric Engineering & Remote Sensing* 47:1717–1724.

———, 1982a. Calibration and model reconstruction in analytical close range stereophotogrammetry, Part I, *Photogrammetric Engineering & Remote Sensing* 48:67-72.

———, 1982b. Calibration and model reconstruction in analytical close range stereophotogrammetry, Part II, *Photogrammetric Engineering & Remote Sensing* 48:215-220.

———, 1983a. Automatic measurement of body surfaces using rasterstereography, Part I, *Photogrammetric Engineering & Remote Sensing* 49:377–384.

———, 1983b. Automatic measurement of body surfaces using rasterstereography, Part II, *Photogrammetric Engineering & Remote Sensing* 49:1443–1452.

———, 1991. Video rasterstereography: A method for on-line measurement of body surfaces, *Photogrammetric Engineering & Remote Sensing* 57:1341–1345.

Fryer, J. G., 1989. Camera calibration in non-topographic photogrammetry, *Non-Topographic Photogrammetry, Second Edition* (H.M. Karara, editor), American Society for Photogrammetry and Remote Sensing, Falls Church, Virginia, pp. 59–69.

McGlone, J. C., 1989. Analytic data-reduction schemes in non-topographic photogrammetry, *Non-Topographic Photogrammetry, Second Edition* (H.M. Karara, editor), American Society for Photogrammetry and Remote Sensing, Falls Church, Virginia, pp. 37–57.

Raynor, J. M., P. Seitz, and D. Wanner, 1990. A universal pixel-synchronous data acquisition system for high-resolution CCD image sensors, *Proc. SPIE 1265*, Society of Photo-Optical Instrumentation Engineers, Bellingham, Washington.

(Received 22 May 1992; revised and accepted 26 February 1993)

Eberhard Hierholzer

Dr. Hierholzer began his undergraduate studies in physics at the Universities of München, Berlin, and Freiburg, which he completed in 1966. He then went on to work in high energy physics at the DESY in Hamburg. He finished his doctoral Thesis in low energy atomic and molecular collisions at the University of Freiburg in 1972 and, since that time, has been at the University of Münster where he has been working on the biomechanics of joints, x-ray photogrammetry, development of rasterstereography, and special calibration problems in close-range photogrammetry.