Automated Procedures with Coded Targets in Industrial Vision Metrology

Susumu Hattori, Keiichi Akimoto, Clive Fraser, and Harutaka Imoto

Abstract

In industrial photogrammetric measurement with CCD cameras, the number of target images may amount to several hundred when large and complex objects are involved. It is therefore impractical to use solely manual means to identify these image points. This paper reports on a further design for coded targets, which can be automatically located and identified, and also discusses special arrangements of reference targets called exterior orientation devices. These can be employed in conjunction with new computational schemes for sensor orientation to facilitate automated data processing. Procedures adopted to ensure sufficient speed and reliability in automated off-line digital photogrammetric measurements employing these developments, which have been found in practice to reduce data processing time by as much as 93 percent, are discussed and experimental results are reported.

Introduction

Off-line industrial photogrammetric measurement with CCD cameras typically utilizes retro-reflective targets to signalize feature points, which are recorded in multiple images. This targeting approach has long been adopted, for reasons of both cost and precision, for large and complex objects such as ship sections and bridge blocks (Fraser, 1993; Fraser et al., 1998). With increasing object size and an emphasis on surface contour determination, the number of required target points can become very large, with several hundred being not at all uncommon for curved ship blocks. Coupled with the fact that between 30 and 100 images might be used within the photogrammetric network, this can add significantly to the burden of photogrammetric measurement if manual identification is relied upon in the data processing phase, which will involve thousands of image point measurements. As a further complication to manual identification processes, retro-reflective targets generally appear as simple dots on a significantly underexposed background.

Prior to bundle adjustment, approximations of all photogrammetric parameters (interior and exterior orientation parameters, and object coordinates) are necessary. Once again, manual involvement in this determination of preliminary parameter values is quite undesirable from a productivity point of view, and also from a reliability standpoint given that labeling errors are to be expected in underexposed imagery with dense target fields. In order to overcome these problems, measurement automation based on coded targets (CTS) and exterior orientation (EO) devices has been developed over the last half decade (e.g., Fraser, 1997; Ganci and Handley, 1998; Hattori et al., 1999; Hattori et al., 2000).

Beyond the basic provision of CTS and an EO device, there is more than one approach to determining sensor EO, which provides a preliminary image-to-image connection, and solving the image point correspondence problem necessary for object point triangulation and subsequent final bundle adjustment. In this paper, the authors present a further alternative process for off-line vision metrology automation. This utilizes a computational process whereby resection for EO determination is initially achieved using a closed-form 3D similarity transformation algorithm. Sequential image connection follows, after which image point correspondences for un-coded targets are determined and a bundle adjustment is performed. An experimental application of the process, which has proven to be fast, robust, and reliable, is then discussed. The presented process is intended fundamentally for full automation. However, the procedures involved have been designed to allow operator interruption and interaction at any stage to accommodate hindrances found in practical measurement operations, e.g., image noise and occlusions, and on-site constraints to network configurations.

Measurement Process

In general, coded targets are too large to be employed for all objects points (Niederoest and Maas, 1996; Hattori et al., 1999), and indeed there is no need to consider such a course of action because CTS are primarily used for initial EO determination, often in conjunction with EO devices. Moreover, for sequential connection of images, they need only be imaged from selected camera stations, preferably more than three, whereas standard retro-reflective targets for measurement (MTS) are placed at all interest points on the object. Keeping this in mind, the following vision metrology measurement process has been adopted:

- For network design evaluation, prior plans of camera station and target configurations are initially examined using a CAD-based network simulator. This phase is useful for complex objects having associated CAD data, although network simulation is often dispensed with for straightforward measurement tasks.
- Following image recording, all CTS in all images are recognized and identified to check the connectivity of images.

S. Hattori is with the Department of Information Processing Engineering, Fukuyama University, Fukuyama, Japan (hattori@fup.fukuyama-u.ac.jp).
K. Akimoto is with the Department of Control Engineering, Shikoku Polytechnic College, Marugame, Japan.
C. Fraser is with the Department of Geomatics, University of Melbourne, Melbourne, Australia.
H. Imoto is with the Department of Production Engineering, Ishikawajima-Harima Heavy Industries Co. Ltd., Yokohama, Japan.

Photogrammetric Engineering & Remote Sensing
0099-1112/02/6805-441S3.00/0
© 2002 American Society for Photogrammetry and Remote Sensing

May 2002 441
The imaging network may then be optimized through image addition (or subtraction) if the degree of connectivity is not sufficient.

The images are connected one by one for coarse, preliminary orientation, thus yielding an approximate EO for the bundle adjustment.

A bundle adjustment is performed to obtain an initial refinement of EO parameters and XYZ coordinates of cm, after which MT~ are iteratively identified and their object space coordinates estimated by multi-ray intersection. A second bundle adjustment is performed to compute the final EO and camera parameters, and XYZ object point coordinates for the multi-image measurement network.

Coded Target Design

The first element upon which the outlined process is contingent is the provision of CTS. To date, there have been two main types of CTS developed, concentric rings (van den Heuvel et al., 1992; Niederoest and Maas, 1996) and dot distributions (Ganci and Handley, 1998). The former, shown in Figure 1, features a relatively limited number of code possibilities, but displays simple and stable recognition and decoding performance (Hattori et al., 1999; Hattori et al., 2000). The latter requires a more complex identification algorithm, but allows for a greater number of code numbers/labels. The design discussed here, which produces 420 code numbers, belongs to the dot distribution group.

The developed CT, as shown in Figure 2, consists of six retro-reflective circular dots, three of which, A, O, and B, define the origin and code coordinate axes, while the pattern of the remaining dots represents the code. Dots for coding are allocated to three positions in 21 circular regions, designated as shown in Figure 2b so as to separate dots by at least one row and column of circles to prevent the occurrence of image blobs in very close proximity. An image of a CT is deformed by affine projection when viewed from an oblique incidence angle. A one-to-one correspondence is maintained under affine transformations by allocating one of three code dots to the opposite side of the line segment AB from the origin O, which leads to 420 possible codes. The size of the code arrangement will depend upon imaging scale. The dimensions shown in Figure 2a are appropriate for the experimental testing reported in the section on Experimental Application, which involved a relatively large imaging scale of around 1:50 (object distance of 1m). The size of each reflective circular dot is the same as that for the adopted MT shown in Figure 3.

The reading process for the CT proceeds in the following way. First, the image is binarized, i.e., zero (0) for a dark pixel and one (1) for a light pixel, through adaptive thresholding, and the blobs forming the code are then extracted by an image dilation technique (Gonzalez and Woods, 1992). If six reflective targets merge to one segment, the group is regarded as a CT. A check by eye is afforded for the operator in cases of sub-optimal image quality or low incidence angle to the target. Second, the three dots (image blobs) A, O, and B, which define the code xy coordinate axes, are searched for. Emerging patterns of CT images are sorted into appearance order for this purpose. Dark and small images of CTS are initially rejected as undeterminable, and every combination of axis candidates from a CT image is then compared to a model pattern. For this purpose, the distances from the centroid of an extracted dot group to these dots are calculated. There is more than one possible combination of axes which can be derived from A, O, and B, as shown in Figure 4. These candidates are tested by projecting the remaining
three dots to the axes and ascertaining whether they fall within a parallelogram defined by A, O, and B. Finally, the ID number/label of the code is uniquely determined. Experience to date has been that, once a valid CT is detected, it can be successfully decoded. A practical concern that has arisen on occasion, however, is false recognition of a cluster of MTS as a CT. In the rare instances where a label is wrongly assigned to these false CTs, they must be detected and rejected within the image orientation phase or eliminated in the worse case by operator intervention.

Image Connection and Exterior Orientation Determination

In the proposed strategy for the use of CTs to support automated measurement, alternative network configuration plans warrant examination with regard to the disposition of CTs and MTS, for a given overall network design, and especially for a given camera station geometry. This “network simulation” process is quite effective in measurement planning for complex objects because the first-order design (Fraser, 1984) also allows the image connections to be ascertained, and therefore suitable CT distributions to be established. Taking into account that a significant percentage of large manufactured components nowadays have an associated design model in CAD, the authors have developed a network simulator, indicated in Figure 5, which works in conjunction with an off-the-shelf CAD system.

By means of the simulator, there can be a check as to whether all proposed images will include at least three well-recognizable CTs, for example. Camera station positions need only be approximated to support this operation. Where there is a deficiency in CT distribution, either extra images can be added or additional CTs employed. It is, of course, always possible to use selected MTS in the role of CTs, but the assignment of labels must be manual in this case. These MTS could, for example, be definition points for a desired reference coordinate system, or even control points on a scale bar.

The image connection process proceeds sequentially, with images being successively connected to a chosen first image pair. The initial two (and possibly more) images are oriented using an EO device. In this case, the EO device is the reference cross shown in Figure 6, which is a portable 2D target array comprising six retro-reflective dots whose coordinates are known accurately in a device-specific 3D coordinate system. The absolute scale of the EO device is not critical; indeed, very robust sensor orientation can be obtained with relatively small target separation. The dimensions shown in Figure 6 are once again appropriate to the object used in the reported experiment. The initial two images are so taken that the EO device lies near the frame centers, and the approximate orientation of the two initial images is obtained by a direct linear transformation (DLT). The resulting 11 parameters of the DLT are converted to six EO parameters, while the estimated interior orientation parameters are disregarded. A bundle adjustment is then carried out using all available CTs.

The third and ensuing images are approximately oriented by single image orientation using three or more CTs whose object space coordinates have already been estimated. Though other alternatives for this resection can be employed, a closed-form of the 3D similarity transformation, which is discussed in the following section, has proven to be a fast and robust approach, which has produced favorable performance in experiments conducted to date. Other possible approaches include those based on a conventional collinearity model, or on the DLT.

With the addition of each new image in the sequential network connection process, the object space coordinates of newly available CTs are first estimated by intersection and then refined by a bundle adjustment “update.” The order in which the images are connected is an important factor for efficient processing under the proposed process. It is not always practical to take images in the pre-planned connection sequence for complex-shaped objects, but it is important to treat the images in the order that ensures continuity in the propagation of the known XYZ reference system to newly added CTs. It is always feasible after image-capture to determine the best connection order, considering the distribution of CTs within the images. Once all the images are connected and oriented approximately, all system parameters are updated by bundle adjustment in order to support efficient correspondence determination of MTS.

Closed Form Solution for Similarity Transformation

In Figure 7a, let the three dots A, O, and B of a CT be projected to a, o, and b on an image plane with the coordinates of o being...
The lines OA and OB define the local object coordinate system \((X', Y')\) with origin O. Then, let a plane \(\Omega\) and the coordinate axes \(u\) and \(v\) on \(\Omega\) be defined in the following way: The plane \(\Omega\) is set to be perpendicular to line \(OO\) (and accordingly \(OE\)) passing through point \(o\) as the origin, and the \(u\) axis is taken as the intersection between the image plane and \(\Omega\), while the \(v\) axis is taken to be orthogonal to the \(u\) axis.

Suppose a point \(p(x, y)\) is projected to \(p'(u, v)\) in \(\Omega\). Figure 7b shows a side view of the image plane projected to the plane including the line \(OE\) and the \(v\) axis, while Figure 7c shows a top view looking downwards to the image plane. The coordinates \(u\) and \(v\) are computed from

\[
\begin{align*}
x' &= x \cos \alpha - y \sin \alpha, \\
y' &= x \sin \alpha + y \cos \alpha
\end{align*}
\]

\[
u = (\sqrt{x_0^2 + y_0^2} - y') \cos \beta
\]

where

\[
\tan \alpha = y_0 / x_0, \quad \cos \beta = c / \sqrt{x_0^2 + y_0^2} + c^2
\]

with \(c\) the camera principal distance.

Because CT images are small, parallel projection from object space to image space is assumed. By referring to Figure 7d, it can be seen that the following transformation equation can be formed:

\[
\begin{bmatrix} u \\ v \end{bmatrix} = \text{scale} \begin{bmatrix} \cos \phi & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} X' \\ Y' \end{bmatrix}.
\]

Here, \(\text{scale}\) is a scale ratio of the \(uv\) space to the \(X'Y'\) space, \(\theta\) is an angle between the \(X'\) axis and the perpendicular to the \(u\) axis, and \(\phi\) is an angle between the perpendicular and the \(v\) axis. Let points \(A(5,0)\) and \(B(0,5)\) in the CT be projected to points \(a'\) and \(b'\) in the image space, respectively. Their coordinates are then obtained from Equation 1 as \(a'(u_a, v_a)\) and \(b'(u_b, v_b)\). The \(\text{scale}\) is then given from Equation 3 by

\[
\text{scale} = \frac{S}{\sqrt{v_a^2 + v_b^2}}.
\]

For \(n\) CTs (\(n\) greater than or equal to 3), of which the global object coordinates are already known and given by \(X_i = [X_i, Y_i, Z_i]^T\) \((i = 1, 2, \ldots, n)\), the collinearity conditions are expressed by

\[
x_i = \text{scale} \cdot \mathbf{M}(X_i - X_0)
\]

where \(x_i = [x_i, y_i - c]^T\) are image coordinates expressed in the object space, \(\mathbf{M}\) is a rotation matrix, and \(X_0 = [X_0, Y_0, Z_0]^T\) are the unknown object space coordinates of a camera station. Equation 5 is normalized by scaling to yield

\[
u_i = \mathbf{M}(X_i - X_0)
\]

where \(\nu_i = [x_i, y_i - c]^T / \text{scale}\). Let \(X_G\) be the center of gravity of the object space coordinates and \(\nu_i\) the corresponding point for the normalized image coordinates. Then,

\[
u_i - \nu_g = \mathbf{M}(X_i - X_0) - \mathbf{M}X_g.
\]

The rotation matrix \(\mathbf{M}\) is estimated so as to minimize the squared sum of discrepancies arising from Equation 7: i.e.,

\[
\sum_i \nu_i^2 = \sum_i (\nu_i - \mathbf{M}X_i)^T (\nu_i - \mathbf{M}X_i) = -2 \sum_i \overline{X}_i^T \mathbf{M}^T \overline{\nu}_i + \text{const}.
\]

Equation 8 reaches a minimum when the right-hand side reaches a maximum. Singular-value decomposition of the correlation matrix

\[
\mathbf{C} = \sum_i \overline{\nu}_i \overline{X}_i^T
\]

leads to

\[
\mathbf{C} = \mathbf{V}^T \mathbf{A} \mathbf{W}
\]

where \(\mathbf{V}\) and \(\mathbf{W}\) are orthogonal matrices and \(\mathbf{A}\) is a diagonal matrix with elements of positive singular values in ascending order. The optimal \(\mathbf{M}\) that minimizes Equation 6 is given by (Kanatani, 1993) as

\[
\mathbf{M} = \mathbf{V}^T \mathbf{W}.
\]

The camera station coordinates are finally given employing a rearrangement of Equation 7: i.e.,

\[
X_0 = X_G - \mathbf{M}^T \overline{\nu}_g
\]

When only three CT images are available, one of the singular values becomes zero, which leads to two solutions for \(\mathbf{M}\), and hence two values of \(X_0\). One solution is false, this being easily singled out by the connection order of images.

**Identification of MTs**

Once all the images are connected and a unified network is formed, all MTs are identified by multi-ray intersection from neighboring images. To eliminate improbable candidate image points for intersection, only MTs included in image groups
which share common CTs are examined. For the multi-ray intersection, a threshold value of \( Th = \sqrt{2} \sigma_g \) for each two rays is employed, where \( \sigma_g \) is the root-mean-square (RMS) value of the standard errors of camera station coordinates. The final gross error detection is performed within the subsequent bundle adjustment, which includes all CTs and identified MTS. This bundle adjustment is executed by free-network adjustment with a Moore-Penrose generalized inverse; thus, it is referred neither to the coordinate datum of the reference cross nor to a designated object space reference system. Instead, the datum is implicitly assigned within the process of minimizing the mean variance of all parameters. The rank deficiency in observation equations is compensated by adding seven basis vectors in the complement of the null space (Granshaw, 1980; Akimoto et al., 1998).

The identification process for MTS is iterative. The first iteration identifies non-ambiguous targets only and the remainder are kept for later processing. Exterior orientation parameters are then updated. The second and subsequent iterations are employed to identify the remaining targets, which continues until no further improvement is achieved. Operator intervention and checking are allowed at every stage. The overall final bundle triangulation with all validated MTS is executed again by a free-network adjustment, but this time with the constraint of minimal mean variance applying to the object space coordinates of MTS only, the result of which determines the precision of the triangulated object points.

**Experimental Application**

In order to evaluate the performance of the proposed approach, which integrates CTs and an EO device with a new computational scheme, a number of experimental measurements were conducted. One of these is summarized here to exemplify the results obtained. The test object was the iron box structure (600 by 400 by 300 mm) shown in Figure 8. On the five faces (four sides and top), 144 MTS and 32 CTs were placed, along with two scale bars (1,000 mm long) and the EO device. For metric verification, object-space coordinates of MTS were measured with a five-axis Numerical Control (NC) machine (Mitsui Precision Machinery model HS5A) to an estimated measurement accuracy of 0.01 mm. The camera used was a Kodak DCS460 B/W (3K- by 2K-pixel array) with 20-mm lens, which had its CCD chip stabilized within the camera housing.

Figure 8. Test box structure showing EO device, CTs, and MTS.

Figure 9 illustrates the exposure station configuration, which comprised 34 camera stations, including two nadir-looking and eight from different elevations. This network geometry inevitably arose because the object was placed on a table. Images were recorded with a range of orthogonal camera roll angles, and the camera-to-object distance was set so as to image the box such that it best filled the frame format. The task of image recording consumed only about 15 minutes. In the resulting 34-station network there were 2,037 target images, including 352 images of CTs. In the initial image scanning involving automatic recognition by binarization, a small number of valid targets were missed and some false targets were found. These errors were manually rectified by the operator. Four to eight well recognizable CTs were found in every image, and all were correctly identified with the adopted decoding algorithm.

After DLT orientation of the first two images, the remaining images were connected in the order dictated by the occurrence of common CTs. According to the authors’ experience, when only three CT images per image are available, adjustment calculations tend to perturb largely at first, but soon converge correctly. All the connected images were bundle adjusted with only CT image points, which yielded a mean camera station standard error of \( \sigma_g = 0.7 \) mm. Recall that this is an interim orientation/triangulation only. Of the MT images, 98.6 percent (1,659 points out of 1,682) were labeled in the first iterative spatial intersection of the image point correspondence process. The total processing time to this point was about four minutes, including image connection on a PC with a 130 Mhz Pentium II processor. The remaining 23 MT points not automatically identified and matched were manually labeled. These tended to be grossly underexposed targets near the extremeties of the image format. Within the second intersection iteration \( (\sigma_g = 0.1 \) mm), no additional points to the original 98.6 percent were automatically located. This phenomenon that the points which could not be labeled in the first identification are rarely labeled in the second was also observed in other experiments.

By way of comparison, a full manual measurement operation was carried out, including identification of MTS with automatic measurement of centroids and input of approximations of EO parameters. This took 60 minutes, which meant that the new computational process with CTs saved 93 percent of the total mensuration and data processing time. As regards measurement accuracy, the RMS discrepancy value between the XYZ object-point coordinates measured by vision metrology and those measured by the NC machine was 0.04 mm (1/25 pixels), which was consistent with the NC machine was 0.04 mm (1/25 pixels), which was consistent with the NC machine was 0.04 mm (1/25 pixels), which was consistent with the NC machine was 0.04 mm (1/25 pixels), which was consistent with the NC machine was 0.04 mm (1/25 pixels).
Conclusion
In an effort to enhance the performance of the off-line vision metrology measurement technique, a further automated image measurement and data processing scheme has been developed and tested. This process utilizes CTS, an EO device, an alternative algorithm for preliminary sensor orientation based upon closed-form similarity transformation, and a process for sequential image connection using CTS to form a network for initial bundle triangulation. Iterative determination of image point correspondences of targets through spatial intersection follows to provide a full network of unlabeled object points for final bundle adjustment and sensor self-calibration. The computational process has been illustrated through the test measurement of an iron box structure, which yielded 0.04-mm accuracy. In comparison to manual measurement, it also produced a time-saving of close to 95 percent. The results obtained show the method described to be a very encouraging development for practical implementation in industrial vision metrology.

References


(Received 27 March 2001; accepted 19 July 2001; revised 24 August 2001)

Forthcoming Articles


B. Guindon and C.M. Edmones. Large-Area Land-Cover Mapping through Scene-Based Classification Compositing.

Yong Hu and C. Vincent Tao. Updating Solutions of the Rational Function Model Using Additional Control Information.


Zhulin Li, Xiuqiao Yuan, and Kent W.K. Lam. Effects of JPEG Compression on the Accuracy of Photogrammetric Point Determination.


Boniface O. Oindo. Predicting Mammal Species Richness and Abundance Using Multitemporal NDVI.


C. Vincent Tao and Yong Hu. 3D Reconstructin Methods Based on the Rational Function Model.


J. Wu, M.D. Ransom, M.D. Nellis, G.J. Kluitenberg, H.L. Seyler, and B.C. Rundquist. Using GIS to Assess and Manage the Conservation Reserve Program in Finney County, Kansas.