Available online at www.sciencedirect.com

SCIENCE DIRECT®

Image and Vision Computing 23 (2005) 59–67

image AND vision COMPUTING

www.elsevier.com/locate/imavis

Complete calibration of a structured light stripe vision sensor through planar target of unknown orientations

Fuqiang Zhou*, Guangjun Zhang

School of Instrument Science and Opto-electronics Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100083, China

Received 30 January 2004; accepted 7 July 2004

Abstract

Structured light 3D vision inspection is a commonly used method for various 3D surface profiling techniques. In this paper, the mathematical model of the structured light stripe vision sensor is established. We propose a flexible new approach to easily determine all primitive parameters of a structured light stripe vision sensor. It is well suited for use without specialized knowledge of 3D geometry. The technique only requires the sensor to observe a planar target shown at a few (at least two) different orientations. Either the sensor or the planar target can be freely moved. The motion need not be known. A novel approach is proposed to generate sufficient non-collinear control points for structured light stripe vision sensor calibration. Real data has been used to test the proposed technique, and very good result has been obtained. Compared with classical techniques, which use expensive equipment such as two or three orthogonal planes, the proposed technique is easy to use and flexible. It advances structured light vision one step from laboratory environments to real engineering 3D metrology applications.

Keywords: Structured light; Calibration; Control points; Planar target; Vision inspection

1. Introduction

With the development of electronics, photo-electronics, image processing and computer vision technique, vision inspection has made giant strides in development. It is becoming increasingly relevant in industry for dimensional analysis, on-line inspection, component quality control and solid modeling, such as the 3D contouring and gauging of large-surface car parts and small-size microelectronic components, and the fast-dimensional analysis of object in relation to the recognition of targets by robots. And also it played an important role in reverse engineering.

Recently, systems based on vision cameras allow automated and non-contact 3D measurements, which are based on two alternative techniques: stereovision and triangulation from structured light stripe vision. Structured light stripe vision inspection avoids the so-called correspondence problem of passive stereo vision and has gained the widest acceptance [1–4] in industry inspections due to fast measuring speed, very simple optical arrangement, non-contact, moderate accuracy, low cost, and robust nature in the presence of ambient light source in situ. Since the early 1970s [5], research has been active on shape reconstruction and object recognition by projecting structured light stripes onto objects.

A basic structured light stripe vision sensor consists of one camera and one laser projector, which together form an active stereo pair. The projector is displaced relative to the camera in space. For example, one of arrangements is shown in Fig. 1, the light stripe is modulated by the depth of the measured surface. Thus, the deformed light stripe contains rich 3D characteristic information of the object surface. The camera captures the image of the object with the deformed light stripe. The task of structured light vision inspection is to acquire the 3D characteristic information of the measured surface from the 2D deformed light stripe image, i.e. 3D reconstruction. A structured light stripe vision sensor generates dense world points by sampling image points on each light stripe in the image. The absence

* Corresponding author. Tel.: +86 10 82316930; fax: +86 10 82314804.
E-mail address: fuqiang_zhou@yahoo.com (F. Zhou).

0262-8856/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.imavis.2004.07.006

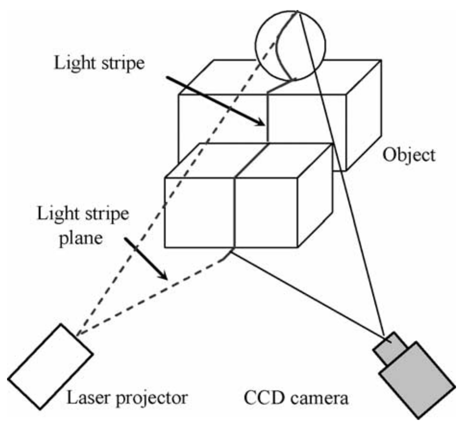

Fig. 1. Overview of the structured light vision sensor.

of the difficult matching problem makes active stereo an attractive method for many vision inspection tasks with the trade-off that the sensor must be fully calibrated prior to use [6].

The procedures to estimate the sensor model parameters are referred to sensor calibration. Sensor calibration is necessary step in vision inspection in order to extract metric information from 2D images. The traditional approach to calibrating a structured light stripe sensor incorporates two separate stages: camera calibration and projector calibration.

In the camera calibration stage, the world-to-image perspective transformation matrix and the distorted coefficient are estimated using at least six non-coplanar world points and their corresponding image projection points [7]. The transformation matrix can be decomposed into camera intrinsic parameters (e.g. effective focal length, principal point and skew coefficient) and extrinsic parameters (including the 3D position and orientation of the camera frame relative to a certain world coordinate system). Camera calibration is performed by observing a calibration object whose geometry in 3D space is known with very good precision. Calibration can be done very efficiently [8, 9]. The calibration target usually consists of two or three planes orthogonal to each other. Sometimes, a plane undergoing a precisely known translation is also used [10]. These approaches require an expensive calibration apparatus, and an elaborate set-up. Zhang proposed a flexible new technique for camera calibration by view a plane from the different unknown orientations, this approach is very easy to use and obtained very good calibration accuracy [11].

In the projector calibration stage, the coefficients of the equation of light stripe plane relative to the same 3D coordination frame must be determined. The projector calibration is related to two principal procedures. One is constructing non-collinear control points. The 3D world coordinates of those control points and their corresponding 2D image coordinates can be estimated in this procedure. The other is establishing the mathematical model of the sensor and estimating the parameters of the model.

The approaches for estimating the structured light vision sensor's model parameters are determined by the established mathematical model itself. Currently, there has been two major ways to establish the mathematical models of structured light stripe vision sensor. One is based on the simple triangular methodology, and another is based on the perspective, translation and rotation transforms in homogenous coordinate system [4]. The perspective projection model is broadly used in the 3D vision inspection. Until now, the extraction of calibration world points for a structured light vision sensor has remained a difficult problem. This is because usually the known world points on the calibration object do not normally fall on the light stripe plane illuminated from the projector, which poses difficulty for the sensor calibration.

Different approaches for calibrating structured light vision sensor have been proposed in many literatures.

In Refs. [12,13], several thin threads that are non-coplanar are strained in the space, and the illuminated light stripe plane from the projector of the structured light vision sensor will intersect the threads at several bright light dots, and these bright light dots are used as the control points. Then a 3D coordinate measuring system such as movable theodolite measurement system is used to measure the 3D coordinates of the bright light dots. However, the strained thread method has the following disadvantage: (1) another 3D calibration target is needed to calibrate camera; (2) the procedure to generate control points for projector calibration is too complex, and only a few points are obtained; (3) as the bright light dot is in fact a kind of distribution of light intensity, it is difficult to match the world bright light dots on the thread strictly with the image bright light dots in the camera image plane, thus the accuracy of the obtained control points is usually poor.

In the Ref. [14], a more sophisticated calibration target is used, which has a zigzag-like face. This method uses the intersection, which are generated by the emitted light stripe plane intersecting with the ridge of the zigzag, as the control points. In this method, the light stripe plane is perpendicular to the datum plane lying on the zigzag is required. Moreover, it has similar problems as the strained thread method.

Chen and Kak [15] proposed the method of the known world line and image point correspondences to directly get the image-to-world transformation matrix. In this method, at least six world lines are required. Later Reid [16] extended the concept to the known world plane and image point correspondence. In Reid's method, a number of known world planes are required.

Xu et al. [17] and Huynh [18] proposed, respectively, a method that is based on the invariance of the cross-ratio to generate the world points on the light stripe plane. The control points can be got by a calibration target whose geometry in 3D space is known with very good precision. The 3D calibration target usually consists of two or three planes orthogonal to each other and is difficult to be manufactured accurately. It is also very difficult to capture the good calibration image by simultaneously viewing the different planes of the 3D calibration object. These approaches require an expensive calibration apparatus.

Besides the methods already mentioned, there are other methods in the literature, such as, standard ruler calibration target mounted on a step-motor-controlled stage with one linear translation axis [19–21], a photoelectrical aiming device combined with a 3D movable platform [22], and so on. But because all these methods are demanding for the precision of the moveable stage, inconvenient for operation, and time-consuming in calibration, none of them is suitable for or widely accepted on-line calibration.

To sum up, the methods of calibrating a structured light vision sensor presented in the literature mainly have the following drawbacks:

(1) it is difficult to generate large number of highly accurate control points.
(2) some methods require an expensive calibration apparatus, and elaborate setup.
(3) the method is well unsuitable for on line sensor calibration in situ.

These factors obstruct the improvement of the calibration accuracy of structured light vision sensor and their engineering 3D metrology applications.

In this paper, inspired by the work of Huynh and the work of Zhang, we propose a novel approach that employs the invariance of a cross-ratio to generate an arbitrary number of control points on the light stripe plane by viewing a plane from unknown orientations. The captured images of the planar target can be used for camera calibration and projector calibration simultaneously. The proposed techniques only requires the camera to observe a planar target shown at a few (at least two) different orientations. The target can be printed on a laser printer and attached to a 'reasonable' planar surface (e.g. a hard book cover). The planar target can be moved by hand. The motion need not be known. In many respects the approach taken here is similar to bundle adjustment procedures used in photogrammetry in that a nonlinear method is used to adjust the system parameters to best accommodate all measured data. The difference is that in our calibration approach, one of the components (the laser striper) is not camera, although it can be considered in some respects to be a special type of linear camera.

The paper is organized as follows. Section 2 introduces the mathematical model of the structured light stripe vision sensor. The model includes the perspective projection model of the camera and the projector model. Section 3 describes the procedure of calibrating the sensor. This procedure includes camera calibration, constructing the control points lying onto the light stripe plane and projector calibration. Section 4 provides the experimental results. Real data is used to validate the proposed technique.

2. Mathematical model of the structured light stripe vision sensor

2.1. Camera model

Fig. 2 shows the perspective projection model of camera in the structured light vision sensor. Note that o_n is usually the center of the camera image plane $\boldsymbol{\pi}_c$, o_c is the projection center of the camera and z_c axis is the optical axis of the camera lens. The relative coordinate frames are defined as follows: $o_n x_n y_n$ is 2D normalized image coordinate frame, $o_u x_u y_u$ is 2D image plane coordinate frame, $o_w x_w y_w z_w$ is 3D world coordinate frame and $o_c x_c y_c z_c$ is 3D camera coordinate frame. In addition, we define some known geometry relations in the camera model, such as $o_c x_c // o_u x_u // o_n x_n$, $o_c y_c // o_u y_u // o_n y_n$ and $o_c z_c \perp \boldsymbol{\pi}_c$.

Given one point P_w in 3D space, its homogeneous coordinate in 3D world coordinate frame is denoted by $\tilde{\mathbf{p}}_w = (x_w, y_w, z_w, 1)^T$ and its homogeneous coordinate in 3D camera coordinate frame is denoted by $\tilde{\mathbf{p}}_c = (x_c, y_c, z_c, 1)^T$. P_n is the ideal perspective projection of P_w in the image plane of camera, its 2D homogeneous coordinate in $o_n x_n y_n$ is denoted by $\tilde{\mathbf{p}}_n = (x_n, y_n, 1)^T$ and its 2D homogeneous coordinate in $o_u x_u y_u$ is denoted by $\tilde{\mathbf{p}}_u = (x_u, y_u, 1)^T$. P_d is the real projection of P_w in the image plane of camera, its 2D real coordinate in $o_u X_u Y_u$ is denoted by $\tilde{\mathbf{p}}_d = (x_d, y_d, 1)^T$.

Based on the pinhole camera imaging theory and perspective projection theory, the camera model is described as follows:

(1) Rigid body transformation from the 3D world coordinate to the 3D camera coordinate:

$$\tilde{\mathbf{p}}_c = \mathbf{H}_w^c \tilde{\mathbf{p}}_w \tag{1}$$

where

$$\mathbf{H}_w^c = \begin{pmatrix} \mathbf{R}_w^c & \mathbf{T}_w^c \\ 0^T & 1 \end{pmatrix}$$

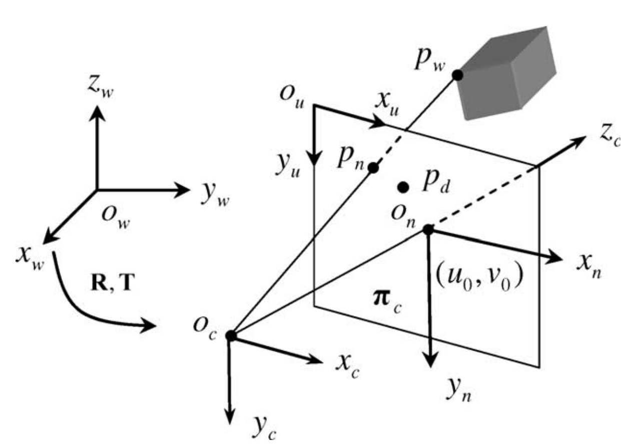

Fig. 2. Perspective projection model of camera in the structured light vision sensor.

is 4×4 transformation matrix, which relates the world coordinate frame to the camera coordinate frame. $\mathbf{T}_\mathrm{w}^\mathrm{c}$ is 3×1 translation vector and $\mathbf{R}_\mathrm{w}^\mathrm{c}$ is 3×3 orthogonal rotation matrix.

(2) Transformation from the 3D camera coordinate to the 2D normalized image coordinate

$$\lambda_1\tilde{\mathbf{p}}_\mathrm{n} = (\mathbf{I}|0)\tilde{\mathbf{p}}_\mathrm{c}, \quad \lambda_1 \neq 0 \tag{2}$$

where $\mathbf{I}$ is 3×3 identity matrix, λ_1 is an arbitrary scale factor.

(3) Transformation from the 2D normalized image coordinate to the 2D undistorted image coordinate

$$\lambda_2\tilde{\mathbf{p}}_\mathrm{u} = \mathbf{A}\tilde{\mathbf{p}}_\mathrm{n}, \quad \lambda_2 \neq 0 \tag{3}$$

where

$$\mathbf{A} = \begin{pmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{pmatrix}$$

is called the intrinsic parameters matrix, f_x and f_y are the effective focal length in pixels of the camera in the x and y direction, respectively, (u_0,v_0) is the principal point coordinates of the camera. λ_2 is an arbitrary scale factor.

According to the Eqs. (1)–(3), the ideal camera model can be represented as follows:

$$\lambda\tilde{\mathbf{p}}_\mathrm{u} = \mathbf{A}(\mathbf{R}_\mathrm{w}^\mathrm{c}|\mathbf{T}_\mathrm{w}^\mathrm{c})\tilde{\mathbf{p}}_\mathrm{w}, \quad \lambda = \lambda_1\lambda_2 \neq 0 \tag{4}$$

We choose the following model to handle lens distortion effects.

$$\begin{cases} x_\mathrm{d} = x_\mathrm{u} + x_\mathrm{u}r^2(k_1 + k_2r^2) + 2p_1x_\mathrm{u}y_\mathrm{u} + p_2(r^2 + 2x_\mathrm{u}^2) \\ y_\mathrm{d} = y_\mathrm{u} + y_\mathrm{u}r^2(k_1 + k_2r^2) + p_1(r^2 + 2y_\mathrm{u}^2) + 2p_2x_\mathrm{u}y_\mathrm{u} \end{cases} \tag{5}$$

where $r^2 = x_\mathrm{u}^2 + y_\mathrm{u}^2$, k_1 and k_2 are the coefficients of the radial distortion, p_1 and p_2 are the coefficients of tangential distortion.

Eqs. (4) and (5) completely describe the real perspective projection model of camera in the structured light vision sensor. All the parameters in the camera model can be estimated by camera calibration.

2.2. Projector model

Fig. 3 shows the Projector model of the structured light stripe vision sensor, which relatives the light stripe plane coordinate frame to the camera coordinate frame. It is useful to identify several coordinate frames as shown in Fig. 3. $o_\mathrm{w}x_\mathrm{w}y_\mathrm{w}z_\mathrm{w}$ is 3D world coordinate frame. $o_\mathrm{c}x_\mathrm{c}y_\mathrm{c}z_\mathrm{c}$ is 3D camera coordinate frame. $o_\mathrm{n}x_\mathrm{n}y_\mathrm{n}$ is 2D normalized image coordinate frame. These three coordinate frames have the same definition as the frames in Section 2.1.

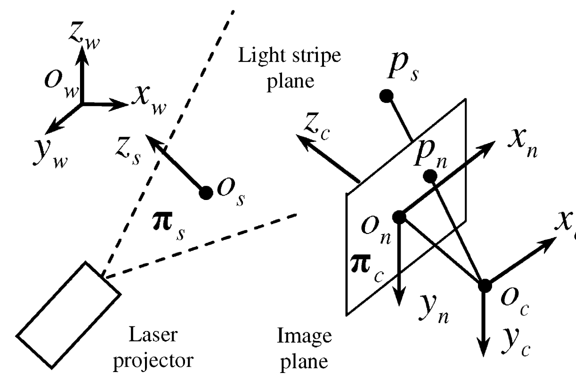

Fig. 3. Projector model of the structured light stripe vision sensor.

Given an arbitrary point p_s lying on the light stripe plane $\boldsymbol{\pi}_\mathrm{s}$. p_n is ideal projection point in the image plane $\boldsymbol{\pi}_\mathrm{c}$. p_s is the intersection point of $o_\mathrm{c}p_\mathrm{n}$ and $\boldsymbol{\pi}_\mathrm{s}$. p_n is the intersection point of $o_\mathrm{c}p_\mathrm{n}$ and $\boldsymbol{\pi}_\mathrm{c}$. Therefore, if the plane equation of $\boldsymbol{\pi}_\mathrm{s}$ and the line equation of $o_\mathrm{c}p_\mathrm{n}$ are known in $o_\mathrm{c}x_\mathrm{c}y_\mathrm{c}z_\mathrm{c}$, the 3D coordinate of the measured point p_s can be determined completely in $o_\mathrm{c}x_\mathrm{c}y_\mathrm{c}z_\mathrm{c}$.

Let $\mathbf{p}_\mathrm{c} = (x_\mathrm{c}, y_\mathrm{c}, z_\mathrm{c})^T$ be the coordinates of a point on the surface of object being profiled, in $o_\mathrm{c}x_\mathrm{c}y_\mathrm{c}z_\mathrm{c}$ relative to the object. The equation of the light stripe plane $\boldsymbol{\pi}_\mathrm{s}$ is described by

$$ax_\mathrm{c} + by_\mathrm{c} + cz_\mathrm{c} + d = 0 \tag{6}$$

Eqs. (4)–(6) constitute a complete real mathematical model of the structured light stripe vision sensor. Based on this model, the calibration of the structured light stripe vision sensor includes two separate steps: (1) camera calibration, i.e. estimating the intrinsic parameters and extrinsic parameters of the camera model, for which at least six non-coplanar world points and their corresponding image projection points are required; and (2) projector calibration, i.e. the calibration of the light stripe plane emitted from the projector. At least three non-collinear control points on the light stripe plane are required for this step. Often more points are necessary to improve the calibration accuracy.

3. A flexible new calibration method

3.1. Camera calibration

The camera intrinsic parameters including distortion coefficients can be calibrated with multiple views of the planar target by applying the calibration algorithm described in Ref [11]. All the views of the planar target are acquired by one camera in different positions, thus each view has separate extrinsic parameters, but common intrinsic parameters.

Due to the non-linear nature of Eq. (5), simultaneous estimation of the parameters involves using an iterative algorithm to minimize the residual between the model and N observations. Typically, this procedure is performed with least squares fitting, where the sum of squared residuals is minimized. The objective function is then expressed as

$$F = \sum_{i=1}^{N}((x_{mi} - x_{di})^2 + (y_{mi} - y_{di})^2) \tag{7}$$

where (x_{di}, y_{di}) is the real image coordinate and (x_{mi}, y_{mi}) is the estimated image coordinate by 3D world point according to the camera projective model.

Two coefficients for both radial and tangential distortion are often enough [9]. Then, a total of eight intrinsic parameters are estimated. The more detailed algorithm is described in Ref. [11].

3.2. Transformation from world coordinate frame to 3D camera coordinate frame

Without loss of generality, we assume the calibration target plane is on $z_w = 0$ of the local world coordinate frame Let's denote the ith column of the rotation matrix by $\mathbf{r}_i$. From Eq. (4), we have

$$\lambda\tilde{\mathbf{p}}_u = \mathbf{A}(\mathbf{r}_1 \quad \mathbf{r}_2 \quad \mathbf{T})\tilde{\mathbf{p}}_w \tag{8}$$

by abuse of notation, we still use $\tilde{\mathbf{p}}_w$ to denote the homogeneous coordinate of a point on the calibration object's plane, i.e. $\tilde{\mathbf{p}}_w = [x_w, y_w, 1]^T$. Therefore, a calibration plane point $\tilde{\mathbf{p}}_w$ and its image $\tilde{\mathbf{p}}_u$ is related by homography $\mathbf{H}$:

$$\lambda\tilde{\mathbf{p}}_u = \mathbf{H}\tilde{\mathbf{p}}_w$$

with

$$\mathbf{H} = \mathbf{A}(\mathbf{r}_1 \quad \mathbf{r}_2 \quad \mathbf{T}) = (\mathbf{h}_1 \quad \mathbf{h}_2 \quad \mathbf{h}_3) \tag{9}$$

where $\mathbf{h}_i$ denotes the ith column of the matrix $\mathbf{H}$.

Given four or more points $\tilde{\mathbf{p}}_{wi}$, where $i = 1,\ldots,n$ for $n \geq 4$, in general position (i.e. no three points are collinear) on calibration plane and their perspective projection $\tilde{\mathbf{p}}_{ui}$ under a perspective center o_c. If camera intrinsic parameters are known, the transformation from world coordinate frame to 3D camera coordinate frame can be obtained as follows [11]:

$$\begin{cases} \mathbf{r}_1 = \lambda\mathbf{A}^{-1}\mathbf{h}_1 \\ \mathbf{r}_2 = \lambda\mathbf{A}^{-1}\mathbf{h}_2 \\ \mathbf{r}_3 = \mathbf{r}_1 \times \mathbf{r}_2 \\ \mathbf{T} = \lambda\mathbf{A}^{-1}\mathbf{h}_3 \end{cases} \tag{10}$$

where $\lambda = 1/||\mathbf{A}^{-1}\mathbf{h}_1|| = 1/||\mathbf{A}^{-1}\mathbf{h}_2||$. Then the rotation matrix relates the world coordinate frame to the camera coordinate frame is expressed as:

$$\hat{\mathbf{R}} = (\mathbf{r}_1 \quad \mathbf{r}_2 \quad \mathbf{r}_3) \tag{11}$$

The matrix $\hat{\mathbf{R}}$ does not satisfy the orthonormality constraint of a standard rotation matrix, but we can normalize and orthonormalize it using the singular value decomposition (SVD):

$$\hat{\mathbf{R}} = \mathbf{UWV}^T \tag{12}$$

The orthonormal version of $\hat{\mathbf{R}}$ is given by

$$\mathbf{R} = \mathbf{UW}'\mathbf{V}^T \tag{13}$$

where$\mathrm{W}' = \mathrm{diag}(1, 1, |\mathbf{UV}^T|)$.

In practice, the linear algorithm describes above is quite noisy because nine parameters for a system of equations with six degrees of freedom. However, the linear estimation is close to the true solution to sever as a good initialization for non-linear optimization technique. All the more accurate parameters of the transformation from the world frame to the camera frame can be estimated by Levenberg–Marquardt non-linear optimization of the cost function described in Eqs. (7) and (8).

3.3. Computing 3D camera coordinates of control points

In Fig. 4, $o_c x_c y_c z_c$ is the 3D camera coordinate frame. $o_i x_i y_i z_i$ is the ith 3D local world coordinate frame defined on the planar calibration target for $z_i = 0$, $i = 1,\ldots,m$. $o_i z_i$ is the normal of the calibration plane. The plane $\boldsymbol{\pi}_c$ is image plane, the plane $\boldsymbol{\pi}_s$ is the light stripe plane and the plane $\boldsymbol{\pi}_i$ is the calibration target plane. The plane $\boldsymbol{\pi}_s$ intersects the plane $\boldsymbol{\pi}_i$ at the straight line $\mathbf{L}_s$. L_j, which consists of three or more known world points, is one column in the planar target. l_j is the perspective projection of L_j.

As shown in Section 2.2, to calibrate the projector, at least three non-collinear control points on the light stripe plane are required in the 3D camera coordinate frame. Often

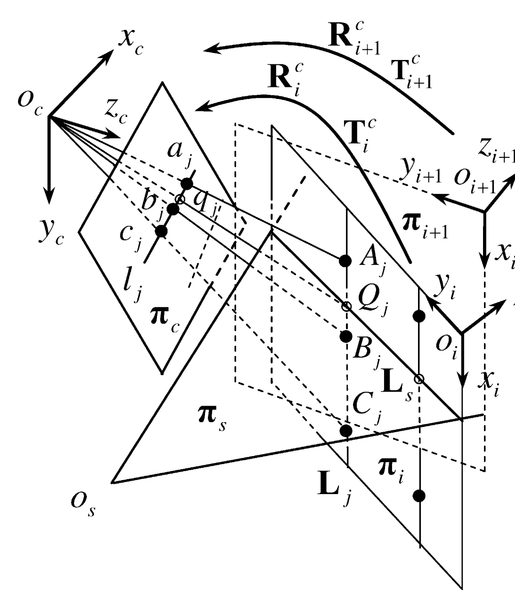

Fig. 4. Computing 3D camera coordinates based on the invariance of the cross-ratios. Using three collinear world points (dark circles $\{A_j, B_j, C_j \mid j = 1,\ldots,n\}$), open circles Q_j's world points lying on $\boldsymbol{\pi}_i$ can be computed. o_c and o_s are the perspective centers of the camera and projector.

more control points are necessary to improve the calibration accuracy. Therefore, we must obtain the 3D camera coordinates of the control points lying on $\mathbf{L}_s$ prior to projector calibration.

Under perspective projection, the length or two lengths' ratio will change, but the ratio of two ratios, which is called the cross-ratios, remains unchanged. That is to say, the four collinear points (A_j,Q_j,B_j,C_j) and their projected points in the images plane (a_j,q_j,b_j,c_j) have the same cross-ratios. The cross-ratio is defined as:

$$r(A_j,Q_j,B_j,C_j) = \frac{A_jB_j}{Q_jB_j} : \frac{A_jC_j}{Q_jC_j} = \frac{a_jb_j}{q_jb_j} : \frac{a_jc_j}{q_jc_j} \tag{14}$$

where $j=1,\ldots,n$.

The image coordinates of the points a_j, q_j, b_j, c_j can be estimate with sub-pixel accuracy by image processing. Then the normalized image coordinates of the points a_j, q_j, b_j, c_j can be obtained from their corresponding image coordinates according to camera model. As shown in Fig. 4, the known collinear world points A_j, B_j, C_j do not fall onto $\boldsymbol{\pi}_s$, the world coordinate of the control point Q_j that falls on $\boldsymbol{\pi}_s$ can be computed according to Eq. (14). Then the camera coordinates of the control points Q_j can be computed from their corresponding world coordinates based on Eq. (1).

We only obtain the collinear control points if we use one view of the planar calibration target from an arbitrary orientation. Therefore, we must construct the calibration control points by using multiple views of the same planar calibration target from the different orientations. All views of the calibration plane are acquired by one camera in the different positions.

We use m views of the captured planar target image with n control points for each view. Let $\tilde{\mathbf{p}}_j^i = (x_j^i \quad y_j^i \quad z_j^i \quad 1)^T$ and $\tilde{\mathbf{p}}_{cj}^i = (x_{cj}^i \quad y_{cj}^i \quad z_{cj}^i \quad 1)^T$ be the homogeneous coordinate of the jth control point in the ith local world coordinate frame and its camera coordinate, respectively, where $i=1\sim m$, $j=1\sim n$.

The 3D camera coordinates of the non-collinear control points can be computed from their local world coordinates frame as follows.

$$\tilde{\mathbf{p}}_{cj}^i = \mathbf{H}_i^c \tilde{\mathbf{p}}_j^i \tag{15}$$

where

$$\mathbf{H}_i^c = \begin{pmatrix} \mathbf{R}_i^c & \mathbf{T}_i^c \\ \mathbf{0}^T & 1 \end{pmatrix}$$

is 4×4 transformation matrix, which relates the ith local world coordinate frame to the 3D camera coordinate frame. $\mathbf{T}_i^c$ is 3×1 translation vector and $\mathbf{R}_i^c$ is 3×3 orthogonal rotation matrix.

$\mathbf{H}_i^c$ is estimated as the scheme described in Section 3.2 with additional at least six the known world points for each view, which fall onto the calibration target and don't fall onto the light stripe plane, and their corresponding normalized image coordinates.

$\mathbf{H}_i^c$ and $\tilde{\mathbf{p}}_j^i$ are determined, $\tilde{\mathbf{p}}_{cj}^i$ for all the control points lying on the different views, respectively, can be computed according to (14).

Theoretically speaking, as the scheme described above, many non-collinear control points lying on the light stripe plane can be constructed by viewing the same planar calibration target from multiple unknown different orientations.

The camera coordinates of all the non-collinear control points lying on the light stripe plane and their corresponding normalized image coordinates can be denoted as follows.

$$\tilde{\mathbf{p}}_{cj}^i = \cup \mathbf{H}_i^c \tilde{\mathbf{p}}_j^i \quad \mathbf{Q} = \cup \mathbf{Q}_j^i \tag{16}$$

where $i=1\sim m, j=1\sim n$. $\mathbf{Q}_j^i$ is the normalized coordinate of images of the jth control points respective to the ith view.

3.4. Projector calibration

An arbitrary number of control points (x_{ci}, y_{ci}, z_{ci}) ($i=1,\ldots,k$) on the structured light stripe plane can easily be obtained as scheme described above. Use Eq. (6) to fit the k control points with the nonlinear least squares. The objective function is the square sum of the distance from the control points to the fitted plane:

$$f(\mathbf{x}) = \sum_{i=1}^{k} d_i^2, \quad i = 1\ldots k \tag{17}$$

where $\mathbf{x}=(a,b,c,d)$, $d_i = |ax_{ci} + by_{ci} + cz_{ci} + d|/(a^2+b^2+c^2)^{1/2}$

After the camera and projector calibration are completed, we can reconstruct the metric 3D camera coordinates from the 2D real image points based on Eqs. (4)–(6).

3.5. Summary

In summary, for the light stripe plane illuminated from the projector, the procedure for the proposed calibration method is as follows:

(1) Extracting all the image coordinates of the light stripe
(2) Correcting the distortion for those image coordinates according to Eq. (5)
(3) Fitting the feature line with the image coordinates of undistorted image coordinates obtained in (2).
(4) Fitting the straight lines with the undistorted image coordinates of the known world collinear points (at least three) with lying on the planar calibration target but not falling on the light stripe plane.
(5) Computing the intersection points of the feature line and the fitted lines in (4), respectively.
(6) Computing the normalized image coordinates of the control points which are computed in (5) according to Eq. (3).

Fig. 5. Designed structured light vision sensor.

(7) Computing the cross-ratio as given by Eq. (14) for each image line l_j and computing the local world coordinates.
(8) Computing the ith transformation $\mathbf{H}_i^c$ with another feature points lying on the ith calibration plane but not on the light stripe and their corresponding image coordinates.
(9) Computing the 3D camera coordinates of the control points.
(10) Estimating the equation of the light stripe plane in 3D camera coordinate frame as described in Section 3.4.

Because our method uses unknown orientation planar calibration target, the proposed calibration method can be easily extended to handle an arbitrary large number of control points. The calibration is very easy and flexible.

4. Experiments

We designed a structured light vision sensor, which is shown in Fig. 5. The sensor consists of an off-the-shelf WATEC CCD camera (902H) with 16 mm lens and a stripe laser projector (wave length, 650 nm, line width ≤ 1 mm). The image resolution is 768×576. The distance between the camera and the laser projector is about 500 mm. The Fig. 6 shows the designed planar calibration target. The calibration target contains a pattern of 3×3 squares, so there are 36 corner points. The size of the pattern is 35×35 mm^2. The distance between the near two squares is 35 mm in the horizontal and the vertical directions. It was printed with a high-quality printer and put on glass.

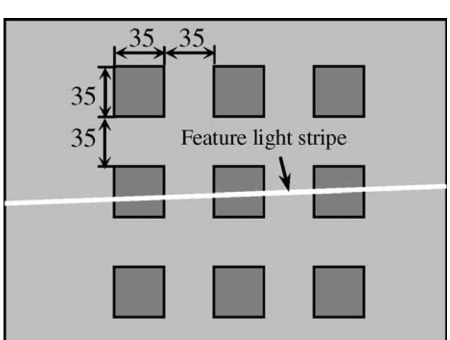

Fig. 6. The sketch of the planar calibration target and the feature light stripe lying on it.

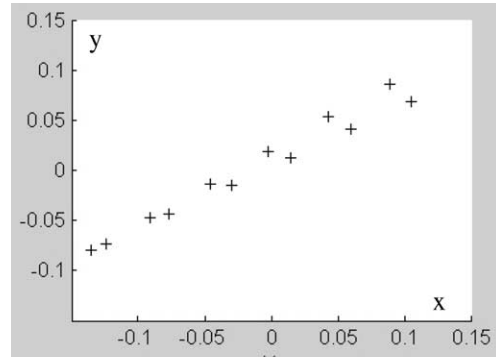

Fig. 7. The normalized image coordinates of the control points lying on the light stripe plane.

Five images of the plane under different orientations were taken. The 36 corner points were detected as the intersection of straight lines fitted to each square. The camera intrinsic parameters including distortion coefficients can be calibrated with those corner points by applying the calibration algorithm described in the reference [11]. The camera intrinsic parameters are

$$
\begin{aligned}
&f_x = 1521.204 \text{ pixels}, f_y = 1515.462 \text{ pixels}\\
&u_0 = 400.987 \text{ pixels}, v_0 = 284.554 \text{ pixels}\\
&k_1 = -4.352\times10^{-1}, k_2 = 1.955\\
&p_1 = -1.789\times10^{-3}, p_2 = -1.295\times10^{-3}
\end{aligned}
$$

Let the laser stripe plane fall onto the calibration squares as shown in Fig. 6, two images of the plane under different positions were taken. After correcting the distortion of the two images, we extract the feature light stripes and the corner points from the two undistorted images. The calibration control points used for calibrating the structured light vision sensor can be constructed with the scheme described above. The normalized image coordinates are shown in Fig. 7 and the 3D camera coordinates are shown in Fig. 8, respectively. We use 12 control points to calibrate the light stripe plane, and get the structured light stripe plane

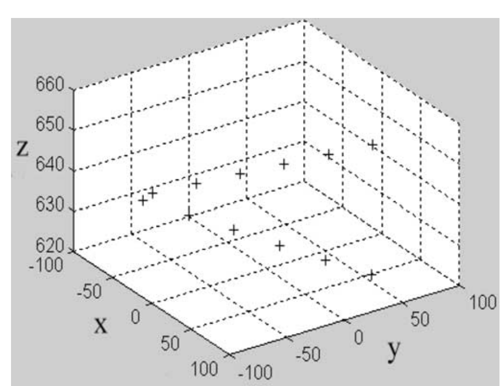

Fig. 8. The 3D camera coordinates of the calibration control points lying onto the light stripe plane.

Table 1
Sensor calibration accuracy evaluation with camera coordinates of control points

Number		Camera coordinates of control points constructing by invariance of cross-ratio			Camera coordinates of control points computing based on sensor model			Camera coordinates errors		
		x (mm)	y (mm)	z (mm)	x (mm)	y (mm)	z (mm)	Δx (mm)	Δy (mm)	Δz (mm)
View 1	1	−85.434	−50.255	630.201	−85.357	−50.427	630.411	−0.077	0.172	−0.210
	2	−57.316	−29.498	628.309	−57.242	−29.625	628.508	−0.074	0.127	−0.199
	3	−29.197	−8.742	626.418	−29.097	−8.800	626.602	−0.100	0.058	−0.184
	4	−1.078	12.013	624.525	−1.140	11.885	624.709	0.062	0.128	−0.184
	5	26.989	32.843	622.668	27.030	32.728	622.802	−0.041	0.115	−0.134
	6	55.158	53.526	620.742	55.234	53.596	620.892	−0.076	−0.070	−0.150
View 2	1	−78.418	−46.628	632.424	−78.378	−46.777	632.921	−0.040	0.149	−0.497
	2	−48.964	−28.285	637.039	−48.961	−28.335	637.474	−0.003	0.050	−0.435
	3	−19.606	−9.811	641.717	−19.523	−9.880	642.031	−0.083	0.069	−0.314
	4	9.800	8.599	646.363	9.811	8.510	646.571	−0.011	0.089	−0.208
	5	39.194	27.024	651.018	39.214	26.943	651.122	−0.020	0.081	−0.104
	6	68.612	45.416	655.656	68.644	45.393	655.677	−0.032	0.023	−0.021
							RMS errors	0.059	0.103	0.255

parameters as follows

$$a = -0.5324, \quad b = 0.7547, \quad c = 0.3832,$$

$$d = 248.998$$

If the local world coordinates and the normalized image coordinates of the control points are obtained with good accuracy, the camera coordinates of control points which are computed based on invariance of the cross-ratio as scheme described above should be accurate. We can use those camera coordinates as the approximate ground values of control points. In addition, we can estimate the camera coordinates by sensor model and we use those estimated coordinates as the measured values of control points. We evaluate the sensor calibration accuracy by comparing the approximate ground values to the measured values. The results are listed Table 1. The RMS errors are as follows $\Delta x = 0.059$ mm, $\Delta y = 0.103$ mm, $\Delta z = 0.255$ mm.

The distance between any two points lying onto the same feature light stripe can be computed as follows

$$\mathrm{d}w = \sqrt{(x_{\mathrm{w},i} - x_{\mathrm{w},i+1})^2 + (y_{\mathrm{w},i} - y_{\mathrm{w},i+1})^2 + (z_{\mathrm{w},i} - z_{\mathrm{w},i+1})^2} \tag{18}$$

$$\mathrm{d}rc = \sqrt{(x_{\mathrm{c},i} - x_{\mathrm{c},i+1})^2 + (y_{\mathrm{c},i} - y_{\mathrm{c},i+1})^2 + (z_{\mathrm{c},i} - z_{\mathrm{c},i+1})^2} \tag{19}$$

where $(x_{\mathrm{w},i}, y_{\mathrm{w},i}, z_{\mathrm{w},i})$ and $(x_{\mathrm{c},i}, y_{\mathrm{c},i}, z_{\mathrm{c},i})$ $(i=0,\ldots,5)$ are the local world coordinate of the ith point and the corresponding 3D camera coordinate which is obtained by measuring, respectively.

Generally speaking, dw should be computed accurately based on invariance of the cross-ratios if the high-accuracy

Table 2
Sensor accuracy evaluation with distance between two control points

(Point 1 point 2)	View 1			View 2		
	dw (mm)	drc (mm)	Δd (mm)	dw (mm)	drc (mm)	Δd (mm)
(0 1)	35.001	35.026	−0.025	35.004	35.017	−0.013
(0 2)	70.001	70.090	−0.088	70.005	70.059	−0.054
(0 3)	105.002	104.918	0.084	105.008	104.977	0.031
(0 4)	140.003	140.013	−0.010	140.010	139.978	0.032
(0 5)	175.003	175.149	−0.146	175.013	175.010	0.003
(1 2)	35.001	35.064	−0.063	35.002	35.043	−0.041
(1 3)	70.001	69.892	0.109	70.004	69.960	0.044
(1 4)	105.002	104.987	0.015	105.007	104.961	0.046
(1 5)	140.002	140.123	−0.121	140.009	139.994	0.016
(2 3)	35.001	34.829	0.172	35.003	34.918	0.085
(2 4)	70.002	69.924	0.078	70.005	69.918	0.087
(2 5)	105.002	105.060	−0.058	105.008	104.951	0.057
(3 4)	35.001	35.095	−0.094	35.002	35.000	0.002
(3 5)	70.001	70.231	−0.230	70.005	70.033	−0.028
(4 5)	35.000	35.136	−0.136	35.003	35.033	−0.030
Δd RMS error (30)				0.085 mm		

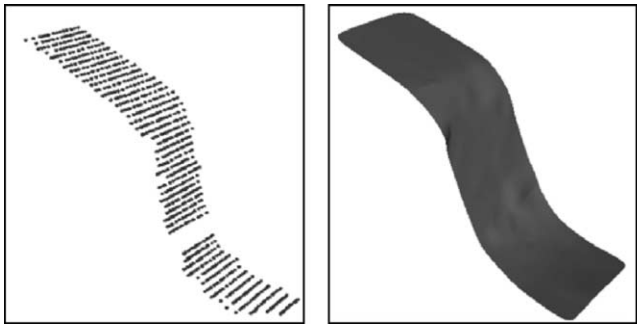

Fig. 9. Reconstruction of the lathe cutter image using the proposed calibration method.

calibration target is used. Therefore, we can also evaluate the sensor calibration accuracy by comparing the deviates between dw and drc. The obtained data are listed in Table 2. The RMS error of the 30 distances is 0.085 mm.

Using the above calibrated parameters, a lathe cutter is measured by the designed sensor. The lathe cutter is fixed on the one-dimensional guide and may be moved in one direction. The results of the reconstruction are illustrated in Fig. 9.

Note that the better sensor calibration accuracy can be obtained if we use the manufactured planar target with good accuracy. We only use the printed planar target with a high-quality printer, but we have obtained very good results.

5. Conclusions

A novel calibration approach for a structured light vision sensor by viewing a plane from unknown orientations was presented. The technique only requires the sensor to observe a planar target shown at a few (at least two) different orientations. Either the sensor or the planar target can be freely moved. The motion need not be known. The proposed approach can generate sufficient non-collinear control points for structured light stripe vision. The experiments conducted on a real structured light vision sensor that consists of one camera and one single light stripe plane laser projector reveal that the proposed approach is of high accuracy and is practical in the vision inspection applications. The proposed approach greatly reduces the cost of the calibration equipment and simplifies the calibrating procedure. It advances structured light vision inspection one step from laboratory environments to real world use.

In addition, note that the method proposed in this paper is not only applicable for the structured light vision sensor with one light stripe plane, but is also suitable for the multiple light stripes vision sensor and structured light grid vision sensor.

Acknowledgements

This work was funded by the Outstanding Younger Foundation of Education Department and National Outstanding Younger Foundation in China.

References

[1] E. Mouaddib, J. Batle, J. Salvi, Recent progress in structured light in order to solve the correspondence problem in stereo vision, Proceedings of the IEEE International Conference on Robotics and Automation 1997; 130.

[2] Y. Zhang, R. Kovacevic, Real-time sensing of sag geometry during GTA welding, J. Manufact. Sci. Eng. 119 (2) (1997) 151.

[3] B.F. Jones, P. Plassmann, An instrument to measure the dimensions of skin wounds, IEEE Trans. Biomed. Eng. 42 (5) (1995) 464.

[4] R. Jarvis, A perspective on range finding techniques for computer vision, IEEE Trans. Pattern Anal. Mach. Intell. 5 (2) (1983) 122.

[5] Y. Shirai, M. Suwa, Recognition of polyhedrons with a range finder, Proceedings on 2nd International Joint Conference on Artificial Intelligence, 1971 pp. 81.

[6] R. Jarvis, in: A.K. Jain, P.J. Flynn (Eds.), Range Sensing for Computer Vision Three-Dimensional Object Recognition Systems vol. 1 of Advances in Image Communications, Elsevier Science Publishers, Amsterdam, 1993, p. 17.

[7] O. Faugeras, Three-Dimensional Computer Vision, MIT Press, 1993.

[8] T. Melen., Geometrical Modeling and Calibration of Video Cameras for Underwater Navigation, Doctoral Dissertation, Norwegian University of Science and technology, Trondheim, Norway, (1994).

[9] J. Heikkila, O. Silven, Calibration procedure for Short Focal Length Off-the-shelf CCD Cameras, Proceedings on 13th International Conference Pattern Recognition, Vienna 1 (1996) 166.

[10] R.Y. Tsai, A Versatile Camera Calibration Technique for high accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV cameras and lenses, IEEE J. Robotics Automation 3 (4) (1987) 323.

[11] Z. Zhang, A flexible new technique for camera calibration, IEEE Trans. Pattern Anal. Mac. Intell. 22 (11) (2000) 1330.

[12] R. Dewar, Self-generated targets for spatial calibration of structured light optical sectioning sensors with respect to an external coordinate system, Proc. Robots and Vision'88 Conf. (1988) 5.

[13] K.W. James, Non-contact machine vision metrology within a CAD coordinate system, Proc. Autofact'88 Conf. 12 (1988) 9.

[14] F. Duan, F. Liu, S. Ye, A new accurate method for the calibration of line structured light sensor, Chinese J. Sci. Instrum. 21 (1) (2000) 108.

[15] C.H. Chen, A.C. Kak, Modeling and Calibration of a structured light scanner for 3D robot vision, Proc. IEEE Conf. Robotics Automation 2 (1987) 807.

[16] I.D. Reid, Projective calibration of a laser-stripe range finder, Image Vis. Comput. 14 (9) (1996) 659.

[17] G. Xu, L. Liu, J. Zeng, D. Shi, A new method of calibration in 3D vision system based on structure-light, Chinese J. Comput. 18 (6) (1995) 450.

[18] D.Q. Huynh, Calibration a structured light stripe system: a novel approach, Int. J. Comput. Vis. 33 (1) (1999) 73.

[19] M. Chang, W.C. Tai, 360-deg profile noncontact measurement using a neural network, Opt. Eng. 34 (12) (1995) 3572.

[20] W. Deng, N. Lu, Back-propagation neural network method for 3-D measurement, Chinese J. Huazhong Univ. Sci. Technol. 27 (1) (1999) 79.

[21] M. Alan, Nonlinear calibration of a laser stripe profiler, Opt. Eng. 41 (1) (2002) 205.

[22] G. Zhang, Z. Wei, Z. Sun, X. Li, A method of structured light based 3D vision inspection using BP neural network, Chinese J. Sci. Instrum. 23 (1) (2002) 31.