Calibration of an Active Binocular Head

Sheng-Wen Shih, Yi-Ping Hung, *Member, IEEE*, and Wei-Song Lin

***Abstract*— In this paper, we show how an active binocular head, the IIS head, can be easily calibrated with very high accuracy. Our calibration method can also be applied to many other binocular heads. In addition to the proposal and demonstration of a four-stage calibration process, there are three major contributions in this paper. First, we propose a motorized-focus lens (MFL) camera model which assumes constant nominal extrinsic parameters. The advantage of having constant extrinsic parameters is to having a simple head/eye relation. Second, a calibration method for the MFL camera model is proposed in this paper, which separates estimation of the image center and effective focal length from estimation of the camera orientation and position. This separation has been proved to be crucial; otherwise, estimates of camera parameters would be very noise-sensitive. Thirdly, we show that, once the parameters of the MFL camera model is calibrated, a nonlinear recursive least-square estimator can be used to refine all the 35 *kinematic* parameters. Real experiments have shown that the proposed method can achieve accuracy of one pixel prediction error and 0.2 pixel epipolar error, even when all the joints, including the left and right focus motors, are moved simultaneously. This accuracy is good enough for many three-dimensional (3-D) vision applications, such as 3-D navigation, 3-D object tracking, and even 3-D reconstruction.**

***Index Terms*—Active vision, binocular head, head calibration, head/eye calibration, motorized lens calibration.**

I. Introduction

ACTIVE vision has become an important research topic in the field of computer vision. The major advantages of an active vision system over a traditional passive vision system is that it can utilize more degrees of freedom to adapt to its environment and, thus, has more potential in many applications. Being able to acquire information actively, many computer vision problems which are ill-posed to a passive vision system become linear and stable to an active observer [1]. Different kinds of binocular heads have been built for investigation of active vision problems, e.g., [2]–[8]. To perform experiments on active vision, we have also built a binocular head (referred to as the IIS head). This IIS head has four revolute joints and two prismatic joints, as shown in Fig. 1. The two joints on top of the IIS head are for camera vergence or gazing (referred to as joint $5L$ and joint $5R$). The next two joints below them are for tilting and panning the stereo cameras (referred to as joint 4 and joint 3). All of the above four joints are revolute and are mounted on an X–Y table which is composed of two prismatic joints (referred to as joints 2 and 1).

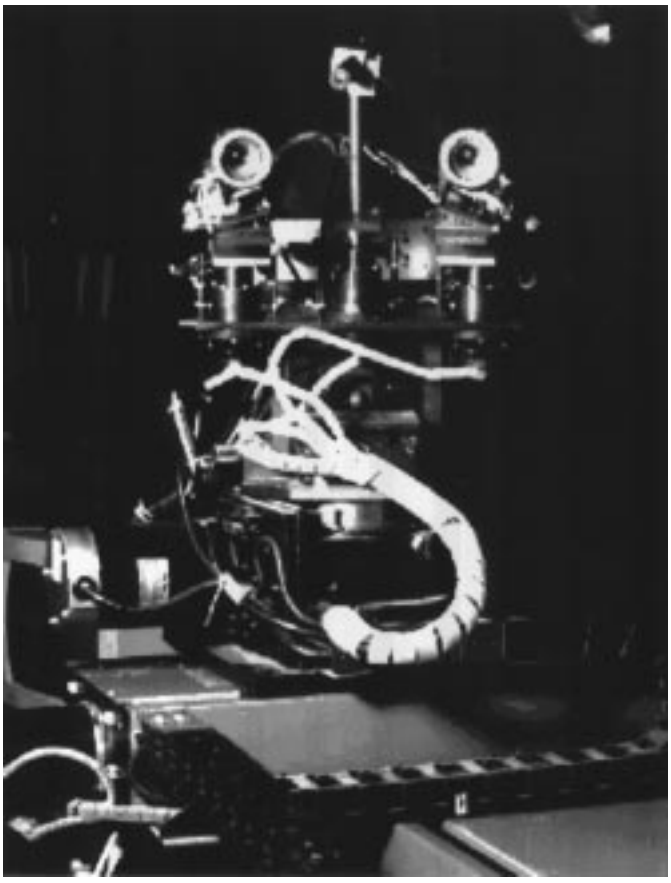

Fig. 1. Picture of the IIS head.

A. Importance of Calibration

Accurate camera calibration can greatly simplify solutions to many important vision problems such as the stereo vision problem, the three-dimensional (3-D) visual tracking problem, the mobile-robot visual guidance problem, the 3-D reconstruction problem, the 3-D visual information registration problem, etc. For example, it is well known that a well-calibrated stereo vision system would not only dramatically reduce the complexity of the stereo correspondence problem but also significantly reduce the 3-D estimation error. Therefore, extensive work has been devoted to developing camera calibration techniques (refer to [9]–[11] for a detailed survey of these calibration methods). Since accurately calibrating a passive stereo vision system is a fairly easy and efficient task, most

Manuscript received June 9, 1996; revised August 27, 1997 and December 26, 1997. This work was supported in part by the National Science Council, Taiwan, under Grants NSC-85-2213-E-001-016 and NSC-86-2745-E-001-007.

S.-W. Shih is with the Department of Computer Science and Information Engineering, National Chi Nan University, Nantou, Taiwan, R.O.C. (e-mail: swshih@ncnu.edu.tw).

Y.-P. Hung is with the Institute of Information Science, Academia Sinica, Nankang, Taipei, Taiwan, R.O.C. (e-mail: hung@iis.sinica.edu.tw).

W.-S. Lin is with the Institute of Electrical Engineering, National Taiwan University, Taipei, Taiwan, R.O.C. (e-mail: weisong@cc.ee.ntu.edu.tw).

Publisher Item Identifier S 1083-4427(98)04358-6.

of the traditional stereo matching algorithms are based on well-calibrated stereo cameras (refer to [12] and [13] for a detailed review). On the other hand, due to the complexity and difficulty of calibrating an active camera, most existing active vision systems are not accurately calibrated. Hence, some researchers chose to develop visual modules which rely on either rough camera models (e.g., Krotkov [14]) or self-calibration techniques (e.g., Deriche *et al.* [15], Du and Brady [16], Ma [17], and Zhang *et al.* [18]). Nevertheless, since an accurate model of an active stereo vision system can simplify solutions to the above mentioned vision problems, many researchers are still working on improving the calibration accuracy of their active binocular heads, e.g., Li *et al.* [19]–[21], Mclauchlan and Murray [22], Shih *et al.* [23], [24], and Young *et al.* [25]. Also, some researchers are using the techniques of the calibration-trio proposed by Tsai [9] and Tsai and Lenz [26]–[28].

B. Difficulties in Calibrating an Active Camera

There are at least three major difficulties in calibration of an active binocular head.

1) *High Complexity:* Calibration of an active vision system consists of three calibration processes, namely, motorized lens calibration, kinematic calibration, and head/eye calibration. Each calibration process is itself a complex problem.
2) *Ultra-High Accuracy Requirement:* The angular resolution (i.e., the view-angle of one pixel) of a camera is roughly several fractions of 1 mrad. However, the orientation accuracy of the state-of-the-art camera calibration technique [9] is at least one order worse than the camera resolution. Moreover, to achieve the required calibration accuracy, the gear backlash and some other nongeometric factors in the kinematics model may not be negligible.
3) *Lack of Accurate Techniques for Camera Parameter Estimation:* The goal of active camera calibration is to establish the relation between the actuating motor positions and the camera parameters of an active/dynamic stereo vision system. However, due to the correlation between the estimates of certain camera parameters, e.g., the correlation between the estimates of image center and camera orientation, an estimate of a set of camera parameters which yields very small two-dimensional (2-D) residual error does not guarantee that the physical camera parameter estimates are themselves accurate [29].

C. Review of Related Work

The calibration problem of an active binocular head can be decomposed into three subproblems: motorized lens calibration, kinematic calibration, and head/eye calibration.

In motorized lens calibration, the goal is to determine the relation between the lens setting and the camera parameters (both intrinsic and extrinsic). Several approaches have been used to solve this problem. Abbott [2] assumed a fixed image center and used low-order polynomials to describe the relation. Li [19], Li and Lavest [21], and Tarabanis *et al.* [30] used partial look up tables and interpolations to represent the relationship. Willson [31] proposed use of low order polynomials for some parameters and fixing the extrinsic parameters except for the component of the translation vector in the optical axis direction. It is well known that the image center is very critical to accurate 3-D computer vision [27]. However, camera calibration techniques that estimate the image center together with the camera orientation suffer from the instability problem for the estimated parameters [29]. According to our theoretical analysis [29], the Willson method [31] is a very accurate method for motorized lens calibration because it initially uses the auto-collimated laser technique to locate the image center and thus separates estimation of the camera orientation from estimation of the image center. Moreover, the Willson method provides a fixed-orientation camera frame, the origin of which is constrained such that it can only translate along the optical axis, while in other models the camera frame is changed without any constraint whenever the lens setting is changed, which makes subsequent head/eye calibration more complicated.

In kinematic calibration, the goal is to determine the relation between the joint values of the head and the orientation and position of the camera mounts. Many methods have been developed for kinematic calibration of binocular heads, e.g., Li *et al.* [20], Young *et al.* [25], Shih *et al.* [23], [24], [32], and Lenz and Tsai [28]. Most of the work has focused on the derivation of closed-form solutions. However, the major problem is that, due to measurement noise, the kinematic parameter estimates obtained by using closed-form solutions are not accurate enough to fit the ultra-high accuracy requirement in applications of active binocular heads [33].

In head/eye calibration, the goal is to determine the camera coordinate system and the camera mount reference frame. The Tsai and Lenz method [26] is recognized as the most accurate one (refer to [20] and [34]). However, if the kinematic model or the estimated camera poses are not accurate enough, then the results will not be satisfactory.

The above-mentioned three calibration stages can also be accomplished in one process provided that sufficiently accurate initial values of the active camera parameters are available. McLauchlan and Murray [22] used a variable state-dimension filter to recursively estimate the head/eye relation and the vertical and horizontal effective focal length. The reported tracking error (similar to the epipolar error) is approximately 0.5 pixel. This method is attractive because it requires no special calibration object except the trajectories of some static feature corners observed in the scene. However, their kinematics model is so oversimplified that no translation component is considered. Also, even though the tracking error is small, they do not know the amount of the absolute error, i.e., the 2-D image prediction error (refer to Section IV). Zhuang *et al.* [35] also proposed simultaneous calibration of a robot and a hand-mounted camera to avoid propagation of estimation errors in the three calibration stages. Notice that by using nonlinear optimization techniques, the active camera parameters can be further refined, and the accuracy will be improved to some extent. However, when applying the Zhuang *et al.* method [35] to the active binocular head calibration problem, one should

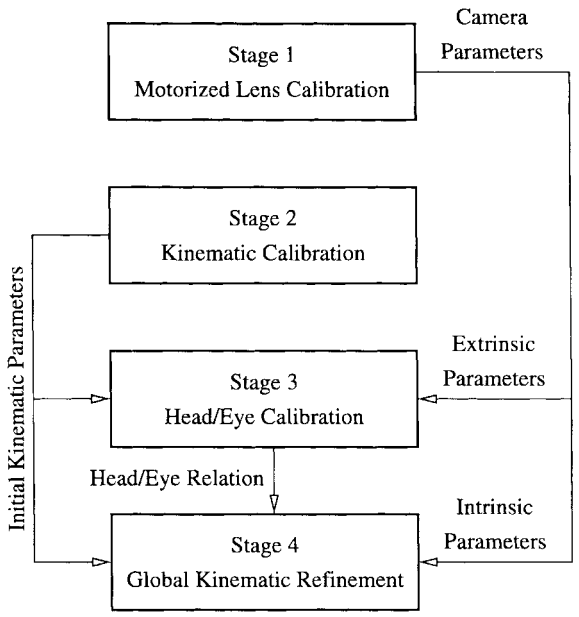

Fig. 2. Schematic diagram of the four-stage calibration process for an active binocular head.

avoid directly estimating all the active camera parameters simultaneously due to the correlations between the estimates of certain extrinsic and intrinsic camera parameters. Instead, based on our theoretical analysis [29], some of the intrinsic camera parameters are better estimated independently of the extrinsic camera parameters; otherwise, the estimates will be very sensitive to noise.

In this study, a four-stage calibration process was used to calibrate the IIS head which could achieve a root-mean-square (rms) prediction error of approximately one pixel and an epipolar error of about 0.2 pixel, even when all the joints are moved simultaneously. Fig. 2 shows the schematic diagram of the four calibration stages. In the first stage, the motorized lens is calibrated by using the *motorized-focus lens* (MFL) camera model, which identifies the functions of the intrinsic camera parameters with respect to the focus settings. In the second stage, the kinematic parameters are estimated by using the closed-form solution proposed in [24]. In the third stage, the head/eye relation is estimated with either a simple method using *coordinate measurement machines*, which will be described in Section IV, or the well-known Tsai and Lenz method [26]. In the fourth stage, all the *kinematic* parameters estimated in the previous two stages are refined by using a nonlinear recursive least-square estimator. This paper describes in detail the first stage and the fourth stage.

Notice that if one has already had accurate enough kinematic model that can be used as an initial value for nonlinear optimization, then he can skip the second stage of kinematic calibration. However, when the available information is not sufficient for constructing a kinematic model for the binocular head (which is usually the case, such as for the IIS head and the TRC UniSight/BiSight head we are using now), the second stage of kinematic calibration is recommended to obtain an accurate initial value for the subsequent nonlinear optimization.

This paper is organized as follows. Section II describes the MFL camera model and a calibration method. Section III describes the kinematics model of the IIS head. Section IV describes the global kinematic calibration process, including the nonlinear recursive least-square estimator. Section V gives the experimental results. Conclusions are given in Section VI.

II. Calibration of the Motorized Lens

A. Perspective Projection and Radial Lens Distortion

The camera model used in this work is the perspective projection model with radial lens distortion, which is commonly used in the field of computer vision. Let p_O be an object point in the 3-D space, and (x_O, y_O, z_O) be its coordinates, in millimeters, with respect to a fixed *object coordinate system* (OCS). Let the projected image coordinates, p_I in pixels, of the object point p_O be (u_I, v_I). The camera model used in this paper requires 12 camera parameters, i.e., u_0, v_0, s_u, s_v, f, κ, ϕ_x, ϕ_y, ϕ_z, t_x, t_y, and t_z, where u_0 and v_0 are the coordinates of the image center, s_u and s_v are the horizontal and vertical pixel spacings, f is the effective focal length, κ is the radial lens distortion coefficient, ϕ_x, ϕ_y, and ϕ_z are the X–Y–Z Euler angles of the camera orientation matrix, and t_x, t_y, and t_z are the location of the optical center, respectively. Equations that relate the 3-D and 2-D coordinates can be written as follows (refer to [9]):

$$(1-\kappa\rho^2)(u_I-u_0)s_u = f\frac{x_C}{z_C} \tag{1}$$

$$(1-\kappa\rho^2)(v_I-v_0)s_v = f\frac{y_C}{z_C} \tag{2}$$

where

$$\rho^2 = (u_I-u_0)^2 s_u^2 + (v_I-v_0)^2 s_v^2,$$

$$\begin{bmatrix} x_C \\ y_C \\ z_C \end{bmatrix} = \mathbf{R}\begin{bmatrix} x_O \\ y_O \\ z_O \end{bmatrix} + \mathbf{t}.$$

$\mathbf{R}$ is the 3×3 rotation matrix composed by using the X–Y–Z Euler angles, ϕ_x, ϕ_y, and ϕ_z, and $\mathbf{t} = [t_x \quad t_y \quad t_z]'$ is the 3×1 translation vector, where the prime denotes the *transpose* operation. Notice that three of the camera parameters, i.e., the effective focal length, f, and the vertical and horizontal pixel spacing, s_u and s_v, can only be solved up to a scale factor. In general, there are two ways to approach this problem. One is to compose them into two effective focal length parameters, namely, the horizontal and vertical effective focal lengths (refer to the Weng method [10]). Another method is the one adopted in the well known Tsai method [9] and is suitable for solid state cameras. This method is based on the fact that the horizontal and vertical pixel spacing of a solid state camera can be directly obtained from the manufacturer. However, the horizontal pixel spacing will be rescaled with an unknown factor after the image is sampled by a frame grabber (refer to [27]). Therefore, only the vertical pixel spacing, s_v, is known *a priori* in the calibration process. This is the method adopted in this paper. Nevertheless, if the vertical pixel spacing is unknown, we can simply set s_v to 1, which yields the same representation as in the Weng method [10].

Fig. 3. Motorized lens of the IIS head.

B. Motorized Lens

Fig. 3 shows a motorized-focus lens used in the IIS head. In general, off-the-shelf commercial zoom lenses are designed for manual operation and are not suitable for computer control. Using this kind of zoom lens in an active vision system would make the calibration work more complicated and time-consuming. Moreover, a lens without a zoom function is usually more accurate (due to smaller distortion), more compact and lighter-in-weight than is a zoom lens. Therefore, we have recently chosen to use a motorized focus lens for the IIS head instead of the motorized zoom lens we used a few years ago. Currently, only the focus setting is motorized, and the aperture setting is left manually adjustable. The lens is able to focus on objects at distances from 0.3 m to infinity, and its nominal focal length is 25 mm. Stepping motors are used to control the focus distance, and the operation range is divided into 2560 steps.

Let f, d, and F denote the effective focal length (i.e., the distance from the optical center to the image plane), the object distance and the focal length, respectively. It is well known that the above three factors can be related by the following lens equation:

$$\frac{1}{F} = \frac{1}{f} + \frac{1}{d}. \tag{3}$$

By using lens equation (3), it can be shown that the effective focal length, f, is in the region $[F, Fd_0/(d_0 - F)]$, where d_0 is the minimal object distance, which is approximately 0.3 m for our motorized-focus lens. Therefore, the resolution of the effective focal length of the IIS head can be computed to be approximately one micrometer when using the stepping motor with 2560 steps, which is much more accurate than the variance of the calibration error. Hence, the error caused by the finite resolution of the stepping motor is negligible. Furthermore, to eliminate the effect of backlash in the motorized lens, the stepping motor controller was designed to move the focus ring of the lens to the target position always from the same direction before it stops.

C. MFL Camera Model

When the focus setting is changed, the physical position, orientation and some intrinsic parameters of the camera will change accordingly. For most off-the-shelf lenses, the relation between the physical camera parameters and the focus setting is too complex to be modeled exactly. Therefore, look-up tables or polynomials are often used to approximate the relation. However, since the stepping resolution of the motorized lens is usually very high, storing the relations requires a huge comprehensive table. It is impractical to build such a huge table via calibration. Hence, low-order polynomials or interpolation techniques based on partial tables are widely used.

Recently, we have observed that accurate camera calibration techniques yielding very small 2-D residual error do not necessarily provide accurate estimates of physical camera parameters. Motivated by this observation, we have derived the analytic covariance matrices of the estimated camera parameters for four different approaches [29]. Based on this theoretical error analysis, we have found that, for small variations, some of the camera parameters are approximately linearly dependent. For example, if the estimated orientation matrix of the camera is

$$\hat{R} = \delta R(\delta\phi_x, \delta\phi_y, \delta\phi_z)\, R \tag{4}$$

where R is the true orientation matrix, and $\delta\phi_x$, $\delta\phi_y$, and $\delta\phi_z$ are the X–Y–Z Euler angles of the error orientation matrix, δR, then the estimates of u_0 and $\delta\phi_y$, v_0 and $\delta\phi_x$, and f and t_z will be linearly dependent for small variation (see [29]). This phenomenon can cause a serious problem in motorized lens calibration, and there are two different approaches to solving this problem. The first approach is to develop new techniques for acquiring more accurate estimates of the extrinsic and intrinsic camera parameters. The second approach is more interesting and will be described in more detail in this paper. This approach makes full use of the extra-degrees of freedom in the calibration (i.e., the linear dependency of the estimates of some camera parameters for small variations). For example, if we fix the camera orientation at a slightly incorrect value during camera calibration, then some other camera parameters will be deviate from their true values to compensate for the effect of small orientation error when the residual error is minimized. Motivated by this theoretical analysis result, we have proposed a camera model for motorized-focus lens, which turns out to be similar to the Willson model [31]. Willson has used these extra-degrees of freedom without theoretical analysis and has achieved a very accurate zoom lens model. In Willson's camera model for a motorized zoom lens, namely the *adjustable camera model*, several camera parameters including the camera orientation and the X–Y component of the translation vector are assumed to be independent of the lens settings. Moreover, the relations between the remaining camera parameters and the lens settings (focus, zoom, and aperture settings) are approximated by multivariable polynomials.

From a series of preliminary experiments on our MFL camera conducted in our laboratory [36], we found that

- the effective focal length varies linearly with respect to the focus setting;
- the camera parameters are independent of the aperture setting;

- the variation of the optical center is small with respect to the change of the focus setting.

Therefore, the MFL camera model is defined as follows.

1) The MFL camera model is an extension of the general camera model described in Section II-A. In the MFL camera model, some camera parameters are functions of the focus settings.
2) The vertical and horizontal pixel spacings, i.e., s_v and s_u, are independent of the focus setting.
3) The position and orientation of the CCS are independent of the focus setting, i.e., the extrinsic camera parameters are all fixed, which yields a very simple head/eye relation.
4) The relation between the image center, (u_0, v_0), and the focus setting is stored in look-up tables.
5) The relation between the coefficient of radial lens distortion, κ, and the focus setting is stored in look-up tables.
6) The relation between the effective focal length, f, and the focus setting is a linear function.

For convenience, we will refer to the fixed camera parameters, s_u, s_v, ϕ_x, ϕ_y, ϕ_z, t_x, t_y, and t_z, as the *nominally-invariant camera parameters*, and to the focus-setting-dependent camera parameters, i.e., u_0, v_0, f, and κ, as the *adjustable camera parameters*, respectively. Since we have eliminated the extra-degrees of freedom in estimating camera parameters by fixing the extrinsic camera parameters, calibration accuracy can be dramatically improved. The calibration procedure for the MFL camera model is described in the following subsection.

D. Calibration of the MFL Camera Model

To calibrate the motorized lens, we have developed a three-step process. In the first step, with a fixed camera position, all the camera parameters are determined simultaneously by using a general camera calibration technique with respect to one focus setting. To ensure calibration accuracy, one should collect a large number of 2-D–3-D calibration pairs in this step and use a nonlinear optimization calibration technique (e.g., the Weng method [10]) to estimate the camera parameters based on an initial solution obtained by a linear camera calibration method.

In the second step, the nominally-invariant camera parameters, i.e., s_u, s_v, ϕ_x, ϕ_y, ϕ_z, t_x, t_y, and t_z, are fixed to the values obtained in the first step, and the relations between the adjustable camera parameters and the focus setting are determined. For each focus setting, the following process is performed to obtain the values of the adjustable camera parameters.

1) *Initial Solution:* By using the estimate of the nominally-invariant camera parameters obtained in the first step and setting κ to zero, (1) and (2) become linear in u_0, v_0, and f. Hence, an initial solution for the adjustable camera parameters can be easily obtained by using linear least-square techniques.
2) *Nonlinear Refinement:* Using the estimated parameters obtained in the previous step as an initial solution, the adjustable camera parameters can be refined by using a nonlinear optimization technique to minimize the following objective function:

$$\epsilon(u_0, v_0, f, \kappa) = \sum_{i=1}^{N} \left\{ \left[(1-\kappa\rho_i^2)(u_{Ii}-u_0)s_u - f\frac{x_{Ci}}{z_{Ci}} \right]^2 + \left[(1-\kappa\rho_i^2)(v_{Ii}-v_0)s_v - f\frac{y_{Ci}}{z_{Ci}} \right]^2 \right\} \quad (5)$$

where N is the number of 2-D–3-D calibration pairs.

After repeating the above process several times for different focus settings, we have a table of the variable camera parameters with respect to several different focus settings.

In the third step, the effective focal length with respect to the focus setting is fitted with a polynomial while the image center and the radial lens distortion coefficient are still maintained in the look-up tables. Notice that only the effective focal length is fitted to a polynomial because its variation with respect to a changing focus setting is more regular and, thus, can be well approximated by a low-order polynomial. On the other hand, the variation of the other parameters is complicated and is better represented as look-up tables than fitted by polynomials.

III. Kinematics Model of the IIS Head

A. Kinematics Model

The kinematic models of the IIS head can be written as follows (refer to [24]):

$${}^0T_{6L} = V_0\, Q_1\, V_1\, Q_2\, V_2\, Q_3\, V_3\, Q_4\, V_{4L}\, Q_{5L}\, V_{5L} \quad (6)$$

and

$${}^0T_{6R} = V_0\, Q_1\, V_1\, Q_2\, V_2\, Q_3\, V_3\, Q_4\, V_{4R}\, Q_{5R}\, V_{5R} \quad (7)$$

where

$$Q_i = \begin{cases} Trans_Z(q_i), & \text{when joint } i \text{ is prismatic} \\ Rot_Z(q_i), & \text{when joint } i \text{ is revolute.} \end{cases}$$

V_i is referred to as the *shape matrix* containing the position and orientation of joint axis i, and q_i is the joint value of joint i, respectively. Fig. 4 gives an illustration of the relationship between the joint frames and the transformation matrices for the IIS head.

Notice that the unknown kinematic parameters are all contained in the shape matrix, and that there are two, four, and six unknown parameters to be estimated for a prismatic joint, a revolute joint, and a fixed-transformation matrix (e.g., the head-to-eye transformation matrices, V_{5R} and V_{5L}), respectively.

B. Eliminating the Effects of Gear Backlash

In general, the effects of gear backlash are classified into nongeometric terms in kinematic calibration (refer to Roth *et al.* [37]). Backlash will not only degrade the motion control performance for an active vision system but may also cause an image prediction error of several pixels. One may choose to use an expensive backlash-free gear to avoid all the problems caused by gear backlash. Another solution is to reduce the

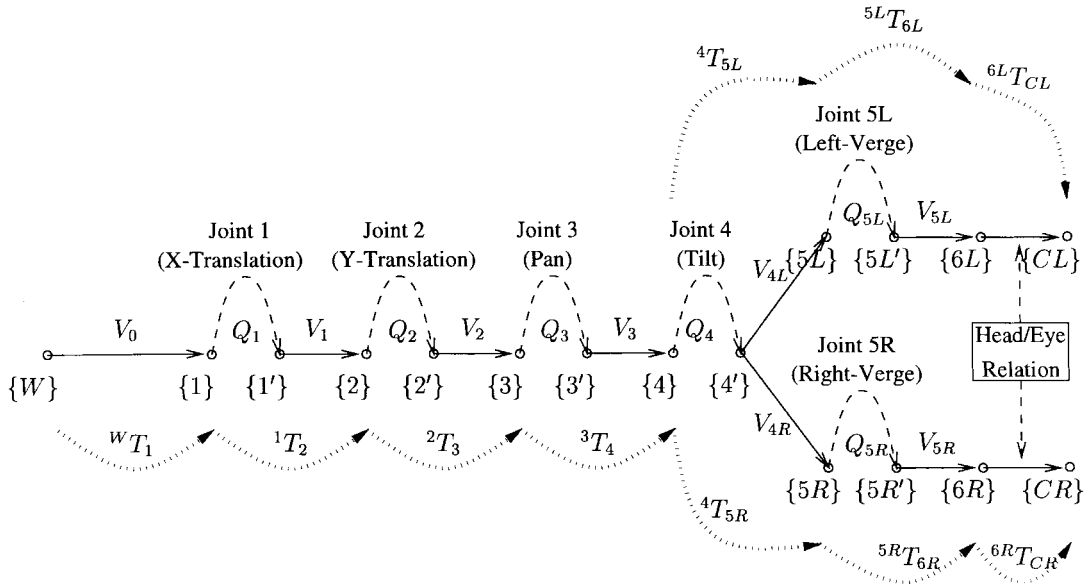

Fig. 4. Illustration of the joint frames and the transformation matrices for the IIS head.

effect of backlash by using either of the following two methods. The first method is to control the motor such that it always approaches the desired position in one direction before it stops. This method is simple and inexpensive, but it is more suitable for a stepping motor system since the overshoot of a servo-controlled system may cause the direction of motion to be unpredictable before the motor stop. The second method is to assume a constant amount of backlash for each joint and to calibrate the backlash constant by using the method described in [38].

IV. Global Kinematic Refinement

A. Calibration Setup

In order to obtain accurate calibration results, the setup used in the calibration should be carefully designed. We will first identify the requirements for the calibration setup when calibrating an active binocular head, and then describe the calibration setup we used for the kinematic refinement. Let the calibration space be the largest convex subset of the joint space of the IIS head which contains all the calibration configurations. If the calibration space is small compared to the joint space, then after the calibration, the binocular head will have a large probability of working outside the small calibration range, which may lead to larger kinematic inaccuracy in general. Hence, to obtain accurate calibration results, the amount of motion for each joint should be made as large as possible during the calibration data acquisition process. On the other hand, when the amount of motion for some revolute joints is large, it is very difficult to keep the calibration object in the field of view. One way to keep the calibration object within the field of view of a rotating camera is to use prismatic motions to compensate for the offset caused by revolute motions. Unfortunately, since there is no vertical translation joint in the IIS head setup, we have to use an auxiliary vertical translation stage to move the calibration object, as shown in Fig. 5. A black calibration

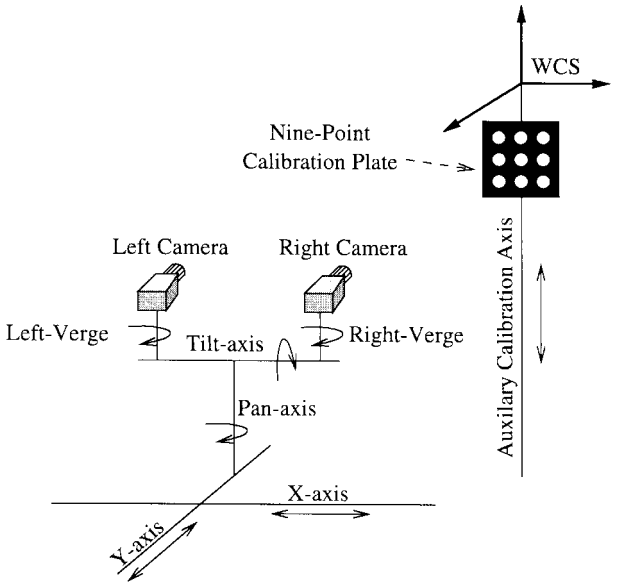

Fig. 5. Schematic diagram of the calibration setup.

plate having nine white disks located at nine grid positions are mounted on the auxiliary translation stage. The origin of the *world coordinate system* (WCS) is defined to be at the center of the nine calibration points when the auxiliary translation stage is at its home position, and the X–Y plane of the WCS is defined to be parallel to the calibration plate. The translation direction of the auxiliary translation stage with respect to the WCS is accurately calibrated by using a coordinate measurement machine (CMM). Therefore, the 3-D coordinates of the centroids of the nine disks on the calibration plate with respect to the WCS are always known even when the translation stage is moving up an down. The auxiliary prismatic joint and the calibration plate are then used for head calibration. They also can be used to provide initial estimates of the head/eye relation and the head-base-to-WCS relation.

Fig. 6. Portable CMM we used for providing the initial estimates of the kinematic parameters.

In the next subsection, we will describe a calibration method for providing an initial estimate of the set of kinematic parameters using a CMM. The main advantage of using a CMM is that it is relatively convenient to acquire highly accurate 3-D measurements, whereas when a vision system is used as the 3-D measurement device, the lighting conditions, background and shape of the calibration object will all affect calibration accuracy. Notice that, since we have a nonlinear refinement process in our four-stage calibration method, a low cost portable CMM is adequate to provide sufficiently accurate initial estimates of the kinematic parameters. The portable CMM we used to provide initial kinematic parameters is shown in Fig. 6. The original purpose of the portable CMM we used is mainly for reconstructing 3-D object models, and it is not as accurate as those very expensive CMM's used for robot calibration. In fact, the accuracy of the CMM we used is even worse than that of the stereo measurement system which we have implemented and described in [24], except that the former has a much larger working space. However, when using stereo cameras as the measurement device, we have to operate the setup very carefully to have the accurate calibration results shown in [24]. Also, according to our theoretical error analysis [33], the estimation errors of the kinematic parameters are inversely proportional to the range of the joint values used to obtain the calibration measurements. Because the joint value is limited to a relatively small range when using the stereo vision system as the 3-D measurement tool, which is due to the restricted observability of the calibration pattern, the kinematic parameters obtained by using the CMM is much more accurate than those obtained by using the stereo vision system. However, if the CMM is not available in the laboratory, one can use the results obtained with stereo vision measurements as the initial value, but may take longer time to converge (as shown in Section V).

B. Initial Estimation of the Head/Eye Relations with a CMM

If a CMM is available, then the process of the second stage and the third stage calibration will become very simple. In the second stage, the CMM is used to provide 3-D point measurements for kinematic calibration (similar to [24] except that the 3-D measurements are obtained by using a CMM instead of a stereo vision system). Here, the transformation matrix between the base frame of the binocular head and the reference frame of the CMM, ${}^{\mathrm{CMM}}T_1$, is estimated as part of the kinematic parameters. After the second stage calibration, the CMM is then used to measure the transformation from the CMM reference frame to the WCS, i.e., ${}^{W}T_{\mathrm{CMM}}$. Hence, the base-to-world transformation, ${}^{W}T_1$, can be computed from ${}^{\mathrm{CMM}}T_1$ and ${}^{W}T_{\mathrm{CMM}}$. Then, the following procedure can be used to estimate an initial value of the head/eye relations (refer to Fig. 7 for the relation between the transformation matrices).

1) Choose a configuration of the binocular head, and calibrate its left and right cameras.
2) Keep the two vergence joints unchanged (so that the stereo parameters obtained in the previous step are still valid for computing 3-D measurements using triangulation), and move the binocular head such that the nine-point calibration plate shown in Fig. 5 is in the field of view of the stereo cameras.
3) Compute the 3-D coordinates of the nine calibration points with respect to the left and right camera frame, respectively, using triangulation. Also, record the corresponding 3-D coordinates of the nine calibration points with respect to the WCS.
4) Compute the transformation matrices from the WCS to the left and right camera frames, ${}^{CL}T_W$ and ${}^{CR}T_W$, based on the Umeyama method [39] by using the three sets of 3-D coordinates of the nine calibration points with respect to the left camera frame, the right camera frame and the WCS.
5) Compute the left and right head/eye relationships as follows:
$$ {}^{6L}T_{CL} = ({}^{W}T_1 \ {}^{1}T_{6L})^{-1} \, {}^{W}T_{CL} \tag{8} $$
$$ {}^{6R}T_{CR} = ({}^{W}T_1 \ {}^{1}T_{6R})^{-1} \, {}^{W}T_{CR} \tag{9} $$
where ${}^{1}T_{6L}$ and ${}^{1}T_{6R}$ can be computed based on the forward kinematic model (6) and (7).
6) Combine the left and right head/eye relationships, ${}^{6L}T_{CL}$ and ${}^{6R}T_{CR}$, with the shape matrices of the left and right vergence joints, i.e., V_{5L} and V_{5R}, and define the *effective* left and right head/eye relationships to be $(V_{5L} \ {}^{6L}T_{CL})$ and $(V_{5R} \ {}^{6R}T_{CR})$, respectively.

Notice that when no CMM is available, a stereo vision system can first be used to provide the measurement data for calibrating the kinematic parameters of the binocular head as described in [24], and then the head/eye relation can be estimated by using the well known Tsai and Lenz method [26]. Once the kinematic parameters and the head/eye relation are available, the base-to-WCS relation can be computed by using a pair of camera pose measurements, ${}^{W}T_{CL}$ and ${}^{W}T_{CR}$, the forward kinematic model, ${}^{1}T_{6L}$ and ${}^{1}T_{6R}$, and the head/eye relation, ${}^{6L}T_{CL}$ and ${}^{6R}T_{CR}$.

C. Parameters to be Estimated in the Global Kinematic Refinement Process

The total number of unknown parameters to be estimated is 35, which includes the transformation matrix from the WCS to the base frame of the binocular head (six parameters, but there are four redundant parameters and only the two orientation parameters of the X prismatic joint axis are variable to be estimated), the orientation of the Y prismatic joint axis (two

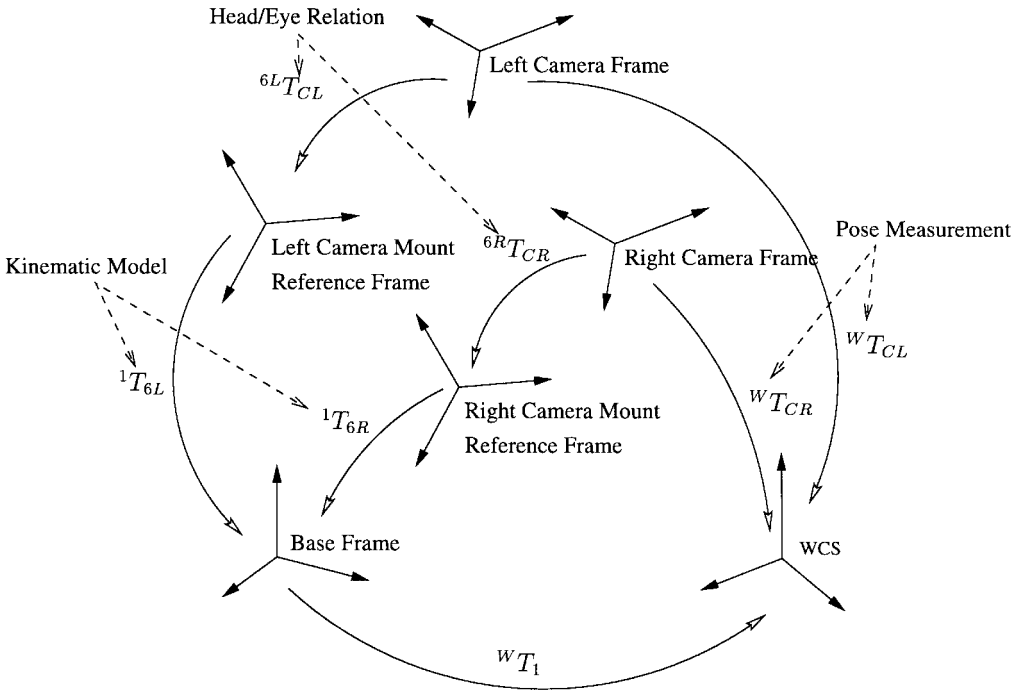

Fig. 7. The relationships between the base frame, camera mount reference frame, camera frame, and WSC.

parameters), the position and orientation of the revolute joint axes of Pan, Tilt, Left-Verge, and Right-Verge (16 parameters, four for each joint), the two transformation matrices of the effective head/eye relations (12 parameters, six for each camera), and three joint level parameters for the tilt joint. The tilt joint (a revolute joint) of the IIS head uses a leverage mechanism driven by a translation stage to achieve a higher payload. Therefore, three joint level parameters have to be estimated to map the position of the leverage translation stage into the revolute joint angle.

D. The 2-D Prediction Error and the Epipolar Error

Global kinematic refinement is performed by minimizing some errors which will be described in the following. Let $P_W(i)$, $i = 1, 2, \cdots, 9$, be the 3-D coordinates, with respect to the WCS, of the calibration points which are the centroids of the nine white disks on the calibration plate (see Fig. 5). From kinematic equations (6) and (7), we have

$$P_{CL}(i) = ({}^{0}T_{6L}\ {}^{6L}T_{CL})^{-1} P_W(i) \tag{10}$$

$$P_{CR}(i) = ({}^{0}T_{6R}\ {}^{6R}T_{CR})^{-1} P_W(i) \tag{11}$$

where $P_{CL}(i)$ and $P_{CR}(i)$ are the 3-D coordinates of the ith calibration point with respect to the left and right camera reference frames, respectively. For each of the nine calibration points, its 3-D coordinates with respect to the camera frame can be computed by using (10) and (11). Then, its image location can be predicted by using the camera model described in (1) and (2). The 2-D prediction error is defined as the difference between the observed image location and the predicted image location of a calibration point. For convenience, let $h_{pL}[P_W(i), q(t); \Theta]$ and $h_{pR}[P_W(i), q(t); \Theta]$ denote the predicted left and right 2-D image location vectors derived from (1), (2), (10), and (11), respectively, where $q(t)$ denotes the joint value vector of the binocular head at time t, and Θ denotes the 35 parameters to be estimated. In addition to the 2-D prediction error mentioned above, we also have to consider the epipolar error, which is very important for stereo vision systems. Let $h_{eL}[p_L(i), p_R(i), q(t); \Theta]$ and $h_{eR}[p_L(i), p_R(i), q(t); \Theta]$ be the left and right epipolar errors, where $p_L(i)$ and $p_R(i)$ are the stereo correspondence of the ith calibration point from the left and right image planes (details of the definition and computation of the epipolar error can be found in [38]). Notice that the ideal values of $h_{pL}[P_W(i), q(t); \Theta]$, $h_{pR}[P_W(i), q(t); \Theta]$, $h_{eL}[p_L(i), p_R(i), q(t); \Theta]$ and $h_{eR}[p_L(i), p_R(i), q(t); \Theta]$ are $p_L(i)$, $p_R(i)$, 0, and 0, respectively. Let $z(t)$ be the 36-component vector containing the actual measurements of the nine calibration points, i.e.,

$$z(t) = \begin{pmatrix} \vdots \\ p_L(i) \\ p_R(i) \\ 0 \\ 0 \\ \vdots \end{pmatrix}. \tag{12}$$

Define the measurement equation as follows:

$$z(t) = h[q(t); \Theta] + w(t) \tag{13}$$

where

$$h[q(t); \Theta] = \begin{pmatrix} \vdots \\ h_{pL}[P_W(i), q(t); \Theta] \\ h_{pR}[P_W(i), q(t); \Theta] \\ h_{eL}[p_L(i), p_R(i), q(t); \Theta] \\ h_{eR}[p_L(i), p_R(i), q(t); \Theta] \\ \vdots \end{pmatrix} \tag{14}$$

is the measurement model, and $w(t)$ is the Gaussian noise with zero mean and covariance matrix $R(t)$.

E. Nonlinear Recursive Least-Square Estimator

Based on the measurement model (14), we can estimate the kinematic parameters by minimizing some objective functions either with a batch nonlinear optimization process or with a recursive method. The batch process requires a predetermined number of measurements that can provide satisfactory calibration results. However, the suitable number of measurements is usually unknown in advance. Also, the initial estimates of the kinematic parameters may not be accurate enough to generate a sequence of robot head configurations such that for each configuration the calibration object is kept within the field of view of the stereo cameras. As a result, calibration data for batch optimization have to be collected under careful supervision, which is a time-consuming and error-prone task. On the other hand, a recursive optimization process will update the kinematic parameters in each iteration, hence, after several iterations, the kinematic parameters will be accurate enough for completely automatic generation of calibration configurations. Therefore, we choose to use a recursive method, i.e., the recursive least-square estimator. Certainly, when the error of the recursive estimator is reduced to a satisfactory level, we can apply a batch nonlinear optimization method and utilize all the data collected in the recursive optimization process to further refine the calibration results. However, our experience is that the result of recursive optimization process is already accurate enough for most computer vision applications. To use a recursive least-square estimator, the objective function is defined as follows:

$$J(\Theta) = [\Theta - \Theta(0)]' P^{-1}(0)[\Theta - \Theta(0)] + \sum_{\tau=1}^{t} v(\tau;\Theta)' R^{-1}(\tau) v(\tau;\Theta) \tag{15}$$

where

$$v(t;\Theta) = \{z(t) - h[q(t);\Theta]\}. \tag{16}$$

$\Theta(0)$ is the initial estimate of the unknown parameters, $P(0)$ is its covariance matrix and $R(\tau)$ is the covariance matrix of noise vector $w(\tau)$ as defined following (14).

The update equations for a nonlinear recursive least-square estimator are as follows (refer to Mendel [40]):

$$P^{-1}(t+1) = P^{-1}(t) + H'(t+1)R^{-1}(t+1)H(t+1), \tag{17}$$

$$\Theta(t+1) = \Theta(t) + P(t+1)H'(t+1)R^{-1}(t+1) \cdot v[t+1;\Theta(t)] \tag{18}$$

where $P(t)$ is the covariance matrix of the parameter vector, $\Theta(t)$, and $H(t)$ is the Jacobian matrix, i.e., $\partial h/\partial\Theta$.

V. EXPERIMENTS

A. Calibration Results of the MFL Camera Model

In the first experiment, the proposed intrinsic parameter calibration method was tested. A traditional camera calibration method was used to estimate all the camera parameters simultaneously corresponding to the focus setting that provided clear-focus images of the calibration objects. In total, 1000 2-D–3-D pairs of calibration points were used in this calibration to ensure parameter accuracy. The estimated nominally-invariant extrinsic parameters were then used to estimate the adjustable intrinsic parameters in the MFL camera model. Figs. 8 and 9 show the calibration results of the adjustable camera parameters. Notice that, to test the repeatability of the proposed method, the adjustable camera parameters with respect to 26 focus motor positions (0, 100, ..., 2500) were estimated four times subject to the same extrinsic parameters. Figs. 10 and 11 show the trajectories of the estimated left and right image center subject to 26 focus settings for the first real experiment. In Figs. 8 and 9, the four curves obtained from four real experiments are close to each other, which implies that the proposed method is quite robust. Also, our experimental results showed that by using the proposed calibration method, the residual rms error of (5) is only 0.24 pixels (computed with 26 focus settings and 14 300 calibration points), which shows that the proposed method is very accurate.

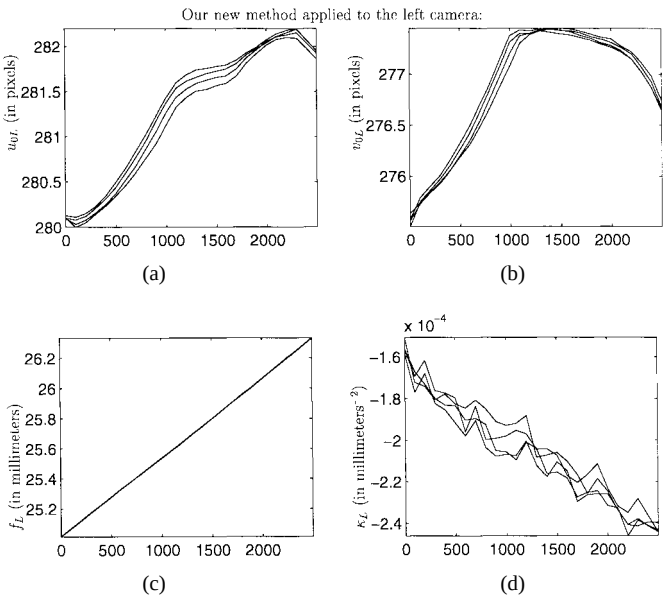

Fig. 8. Variations of the intrinsic camera parameters of the left camera estimated by using the proposed method, with respect to different position of the focus motor. The horizontal axes represent the position of the focus motor. The four curves represent the results of four different real experiments. (a) u_{0L} (in pixels), (b) v_{0L} (in pixels), (c) the effective focal length, f_L (in millimeters), and (d) the radial lens distortion coefficient, κ_L (in millimeters $^{-2}$).

Notice that, in Figs. 8 and 9 the estimates of the effective focal length obtained in four different real experiments are very close to each other whereas the estimates of the radial lens distortion coefficient look noisy. However, the four estimates of the lens distortion coefficient will have approximately the same amount of distortion when using $(1 - \kappa\rho^2)$ in (1) and (2). Therefore, the calibration results obtained using any of the four curves are all acceptable. As to the effective focal lengths, linear functions are fitted, and the resulting relations

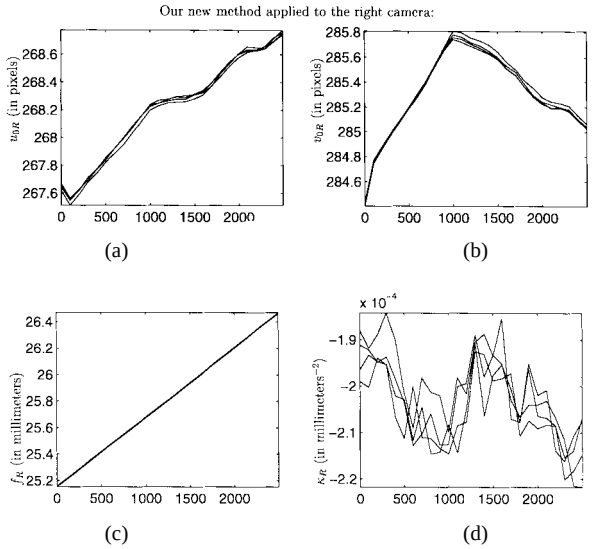

Fig. 9. Variations of the intrinsic camera parameters of the right camera estimated by using the proposed method, with respect to different positions of the focus motor. The horizontal axes represent the position of the focus motor. The four curves represent the results of four different real experiments. (a) u_{0R} (in pixels), (b) v_{0R} (in pixels), (c) the effective focal legnth, f_R (in millimeters), and (d) the radial lens distortion coefficient, κ_R (in millimeters $^{-2}$).

for the left and right cameras in our experiment are

$$f_L(M_L) = 25.01005 + 5.191702 \times 10^{-4} M_L \quad (19)$$

and

$$f_R(M_R) = 25.153\,86 + 5.241\,816 \times 10^{-4} M_R \quad (20)$$

where M_L and M_R denote the left and right focus motor positions, respectively.

Equations (19) and (20) together with the twenty-six-entry lookup tables of the image centers, $[u_{0L}(M_L), v_{0L}(M_L)]$ and $[u_{0R}(M_R), v_{0R}(M_R)]$, and the lens distortion coefficients, $\kappa_L(M_L)$ and $\kappa_R(M_R)$, were used to generate the intrinsic parameters of the IIS head in the subsequent applications.

For comparison, we have also utilized the traditional method, which is used in the first step of the proposed MFL camera calibration method, to directly estimate all the camera parameters corresponding to the 26 focus motor positions. The major difference is that, with the traditional method, all the camera parameters are estimated simultaneously all the time; whereas in our new method, the extrinsic parameters are fixed, and only some of the intrinsic parameters need to be estimated. To test the repeatability of the traditional method, we also performed four real experiments, and some typical results are shown in Figs. 12 and 13. Fig. 13 shows the trajectory of the estimated image center subject to change of the focus setting for the first real experiment. Notice that the trajectory is quite irregular and varies by 20–30 pixels, which is large compared to the more regular trajectories with 1–3 pixel variation in Figs. 10 and 11. Also, from these figures, the intrinsic parameters vary dramatically subject to change of the focus motor position. Therefore, the interpolation results are less reliable, and a comprehensive lookup table is required for this approach, which makes it impractical.

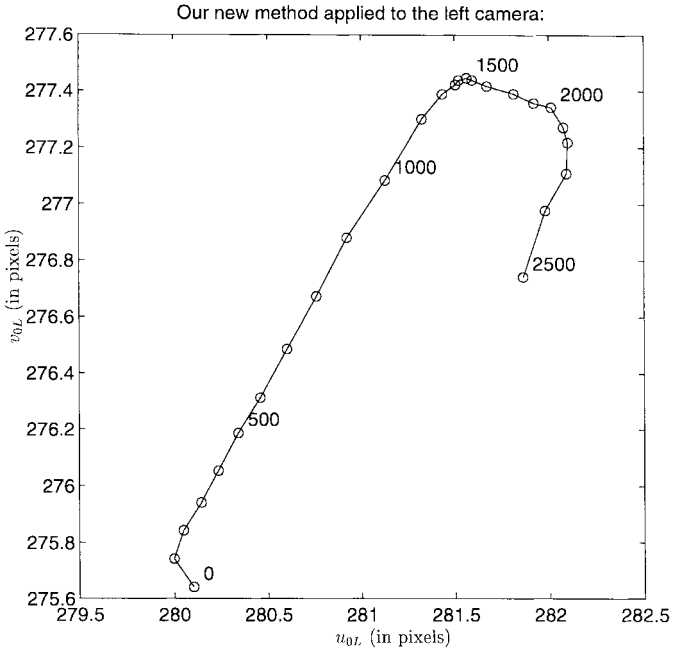

Fig. 10. Trajectory of the left image center subject to the focus motor position estimated by using the proposed method. Values of the focus motor positions are marked every five samples around the corresponding estimated image centers. Notice that the variation of the left image center is about two pixels.

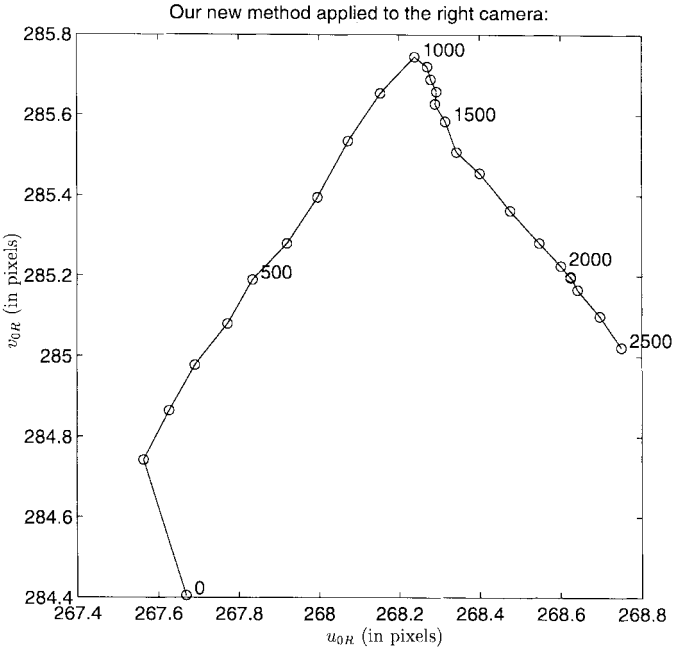

Fig. 11. Trajectory of the right image center subject to the focus motor position estimated by using the proposed method. The values of the focus motor positions are marked every five samples around the corresponding estimated image centers. Notice that the variation of the right image center is less than two pixels.

B. Initialization of the Recursive Estimator

The initial values of unknown parameters, $\Theta(0)$, are obtained by applying the second stage and the third stage calibration methods, based on a low-cost portable CMM (refer to Section IV-B), and the initial covariance matrix,

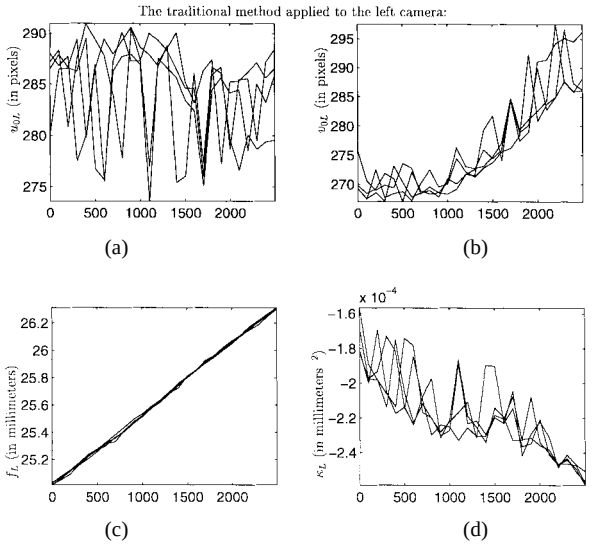

Fig. 12. Variations of the intrinsic camera parameters of the left camera estimated by using a traditional direct nonlinear method, with respect to different positions of the focus motor. The horizontal axes represent the position of the focus motor. The four curves represent the results of four different real experiments. (a) u_{0L} (in pixels), (b) v_{0L} (in pixels), (c) the effective focal length, f_L (in millimeters), and (d) the radial lens distortion coefficient, κ_L (in millimeters $^{-2}$).

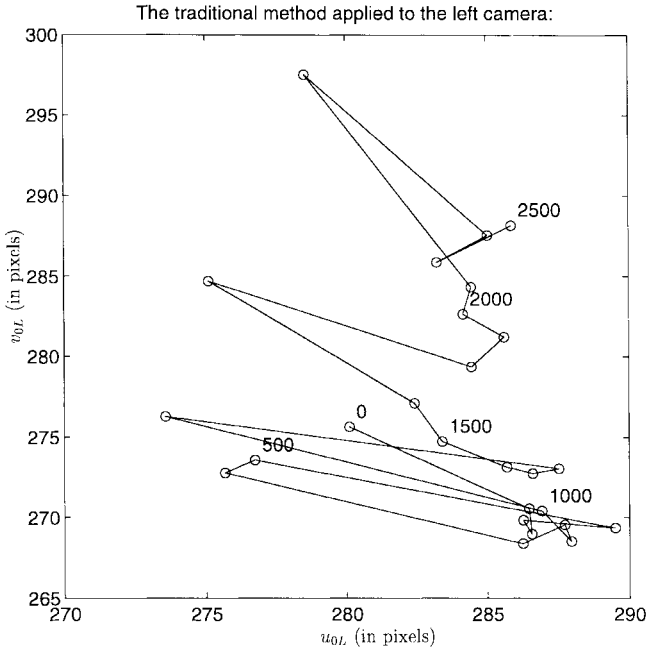

Fig. 13. Trajectory of the left image center subject to the focus motor position estimated by using a traditional direct method. The values of the focus motor positions are marked every five samples around the corresponding estimated image centers. Notice that the variation of the left image center is larger than 15 pixels. Similar results have been obtained for the right image center when using the traditional method.

$P(0)$, is set to diagonal, in which the initial variances of each translation and each orientation parameters are set to (0.2 mm)^2 and $(0.02^\circ)^2$, respectively. Notice that a larger initial variance might cause a faster convergence rate; however, it might also cause the instability problem since the number of unknowns is very large. Therefore, a smaller initial variance may yield a slower convergence rate, but the recursive filter will be more stable.

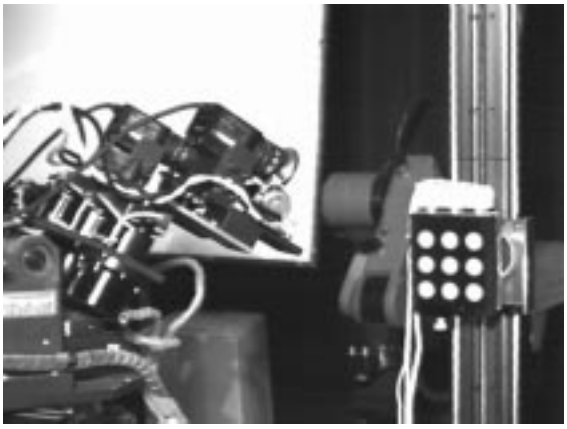

Fig. 14. Picture of the IIS head fixating at the nine-point calibration plate during calibration.

C. Acquiring Calibration Data

In order to calibrate the IIS head, we had to generate a sequence of joint values to control the binocular head such that the calibration plate was visible to the stereo cameras all the time. This task could be accomplished by using an inverse kinematics module. According to our inverse kinematics scheme described in the Appendix, the gaze angle and the joint values of the X–Y table and of the auxiliary Z-axis were chosen to be the given parameters. Therefore, in the real experiments, the gaze angle and the above three joint values were randomly generated whereas the remaining joint values were determined by solving the inverse kinematics equation. Notice that at the beginning of the calibration process, if the head parameters, $\Theta(t)$, were not accurate enough to generate joint values for tracking of the calibration plate, then the calibration process had to be interrupted, and manual operation could be imposed to drive the calibration object back to the view-field of the stereo cameras such that new measurements could be obtained to update the kinematic parameters. However, since our initial parameters were accurate enough, the calibration object was successfully tracked during the whole calibration process.

The process of data acquisition started with random generation of the gaze angle and the three joint values (the X–Y table and the auxiliary Z-axis). Then, the joint values of the remaining axes were determined automatically using the inverse kinematics module. This set of new joint values, $q(t)$, was then be sent to the motion controller of the IIS head. Fig. 14 shows a picture of the IIS head fixating at the nine-point calibration plate during calibration. After the motion command was completed, stereo images of the calibration plate were acquired, and the centroids of the nine calibration disks were computed and sorted to generate correct stereo correspondences. Fig. 15 shows a typical picture of the detected nine stereo correspondences. The image location accuracy of the centroid detection for the nine calibration disks was about 0.1 pixel, and the total time for image acquisition and processing was approximately 0.5 s, using a Sparc 20 workstation equipped with a Data Cell S2200 frame grabber. The nine detected stereo correspondences and the joint values, $q(t)$, were fed into the nonlinear recursive least-

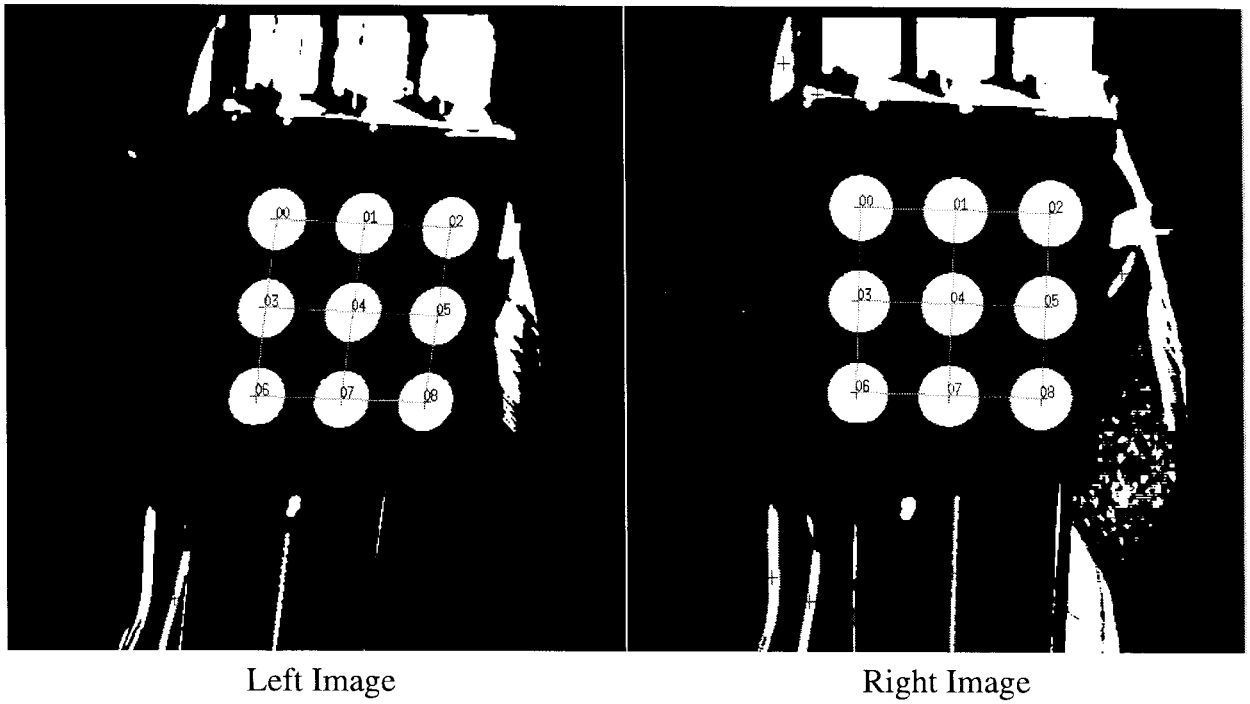

Fig. 15. Typical picture of the detected stereo correspondences.

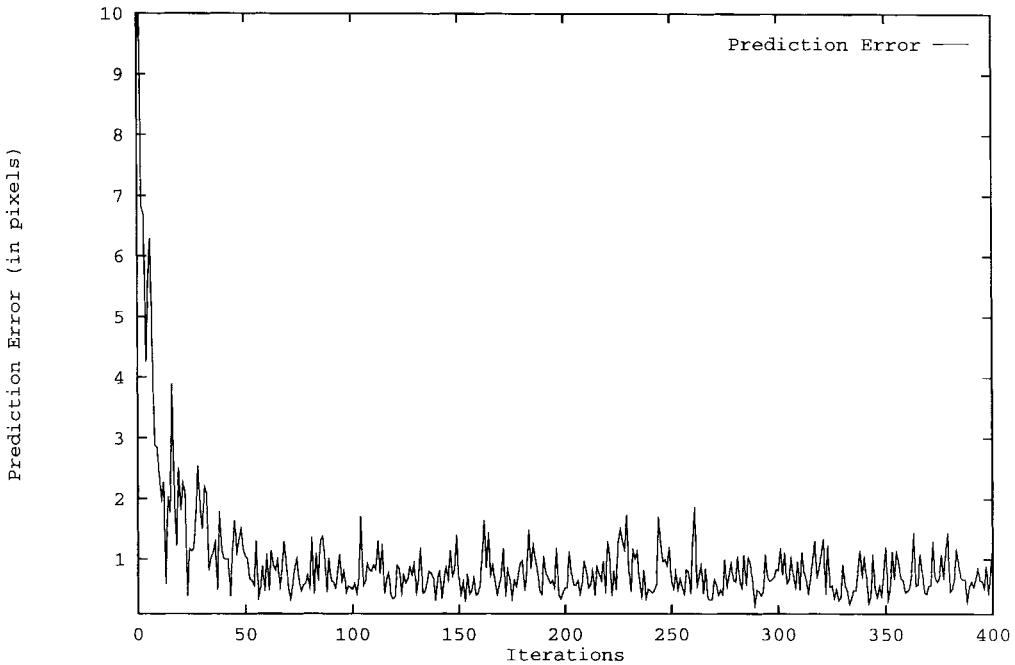

Fig. 16. Decrease of the 2-D prediction error with our iterative refinement procedure. A portable CMM was used in the second and third stages for providing the initial kinematic parameters.

square estimator described in Section IV-E. The distribution ranges of the generated joint values are listed in Table I.

D. Real Experimental Results

Figs. 16–18 show some results of the real experiments. Fig. 16 shows the rms 2-D prediction error based on both the left and right 2-D predictions (18 2-D predictions from the left and right images corresponding to the nine calibration points). Initially, the prediction error was about ten pixels, and it decreased rapidly to about one pixel after iteratively correcting the 35 kinematic parameters of the binocular head. Notice that since the effective focal length of the IIS head was about 25 mm, and the pixel dimension was about 10 μm, one pixel of rms prediction error was approximately equal to 0.4 mrad of rms orientation error, which represents an extremely high accuracy calibration result for an active binocular head.

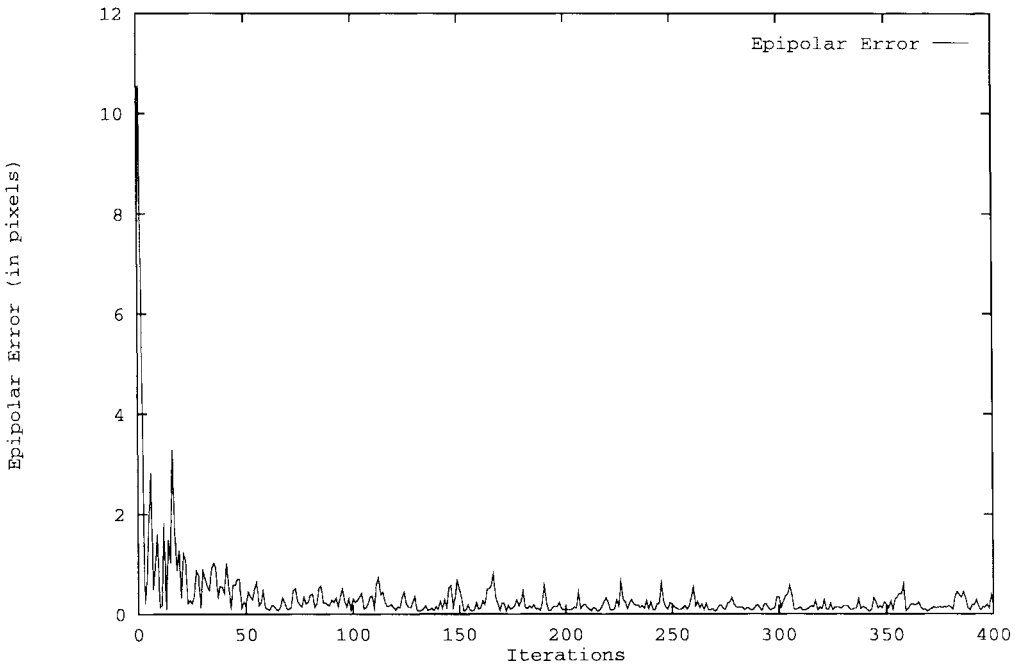

Fig. 17. Decrease of the epipolar error with our iterative refinement procedure. A portable CMM was used in the second and third stages for providing the initial kinematic parameters.

TABLE I
CALIBRATION RANGE OF EACH JOINT

Joint	Name	Working Range	Calibration Range
1	X	0–1000 mm	0–200 mm
2	Y	0–500 mm	0–500 mm
3	Pan	0°–200°	80°–133°
4	Tilt	0°–60°	1°–45°
5L	Left Vergance	0°–80°	0°–43°
5R	Right Vergance	0°–80°	0°–44°
6L	Left Effective Focal Length	25.01–26.34 mm	25.11–25.52 mm
6R	Right Effective Focal Length	25.15–26.50 mm	25.21–25.65 mm

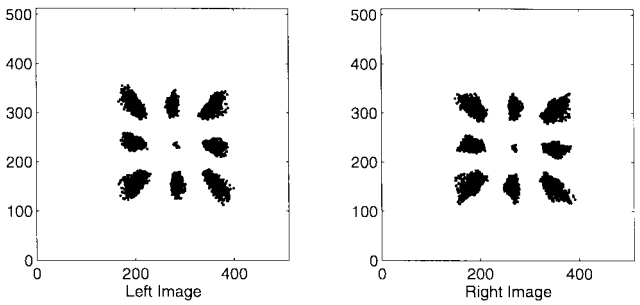

Fig. 18. Footprints of the nine calibration points in the left and right images.

Fig. 17 shows another important performance index of an active binocular head, i.e., the rms epipolar error. As shown in this figure, the initial value of the extrinsic and intrinsic parameters caused an epipolar error of about ten pixels, and after several hundred iterations, the epipolar error was reduced to 0.2 pixels on the average. Notice that the 2-D feature extraction error of the circular calibration points was approximately 0.1 pixel. This shows that the epipolar lines of the IIS head computed based on the calibrated parameters are very accurate; hence, the epipolar constraint can be fully used to reduce the searching range for stereo correspondences when dealing with stereo vision problems with this active binocular head.

Recall that the joint values required to fixate the stereo cameras at the calibration object were determined by using the inverse kinematics module described in the Appendix. Here, the mission of the inverse kinematics module was to compute a set of feasible joint values such that the given interest point would be fixated at the center of the left and right images. In this experiment, the point of interest was selected as the centroid of the nine disks on the calibration plate. Fig. 18 shows distributions of the image locations of the nine calibration points on the left and right image planes, respectively, corresponding to the 1000 iterations in the experiment. Because the center calibration point was fixated, its footprints in the image were confined within a much smaller region than were those of the other eight calibration points.

The last experiment is to repeat the global kinematic refinement experiment except that the initial kinematic parameters used is the estimates obtained by applying the stereo vision technique [24] in kinematic calibration (stage 2) and by applying the Tsai and Lenz method [26] in head/eye calibration (stage 3). Fig. 19 shows the rms 2-D prediction error of the nine calibration points. Initially, the prediction error was about seven pixels, and it decreased to about one pixel after several hundred iterations of correcting the 35 kinematic parameters. However, because the calibration range (i.e., the range that each joint can be moved when measuring the calibration points) of using a stereo vision system as the 3-D measurement tool is much smaller than that of using a CMM, the initial estimates of the kinematic parameters are less accurate than

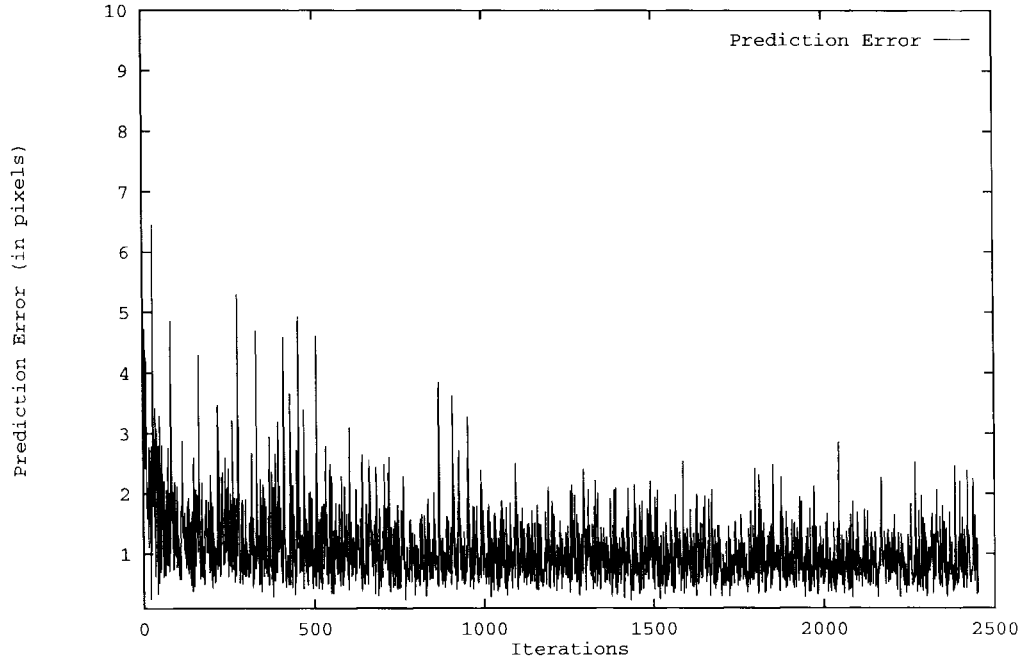

Fig. 19. Decrease of the 2-D prediction error with our iterative refinement procedure. The stereo vision technique and the Tsai and Lenz method were used in the second and the third stages, respectively, for providing the initial kinematic parameters.

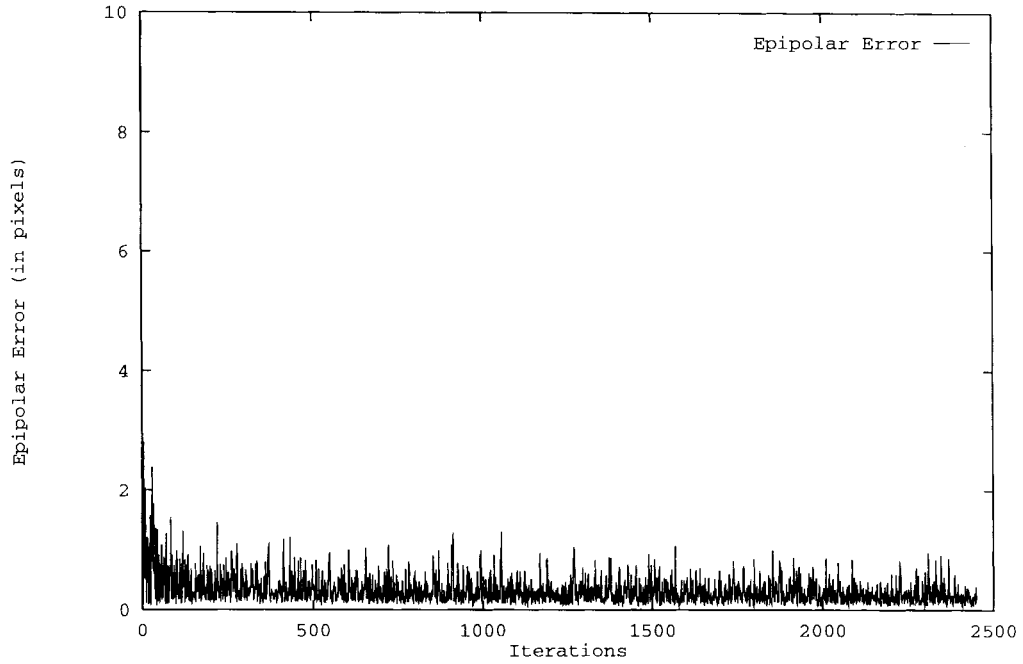

Fig. 20. Decrease of the epipolar error with our iterative refinement procedure. The stereo vision technique and the Tsai and Lenz method were used in the second and the third stages, respectively, for providing the initial kinematic parameters.

those obtained by using the CMM (refer to [33]), and the residual error of the nonlinear optimization process is higher than the one obtained by using the initial value that is obtained with the CMM. However, while the peak 2-D prediction error is slightly larger than that obtained by using a CMM, as shown in Fig. 19, the rms epipolar error roughly converges to 0.3 pixel with a converged peak error of 1 pixel, as shown in Fig. 20. This 2-D prediction accuracy and epipolar accuracy is good enough for most computer vision applications such as 3-D tracking and 3-D reconstruction. Notice that Figs. 19 and 20 show the experimental results of 2500 iterations while Figs. 16 and 17 show only 400 iterations because of faster convergence.

VI. Conclusion

Active stereo vision is attracting more and more research interest. Many binocular heads have been built to investigate the active vision problems. However, accurate calibration of

a binocular head has been such a difficult task that little work relying on well-calibrated heads has been carried out in the past. The main difficulties of the calibration problem for a binocular head are the high complexity, ultrahigh accuracy requirement and the lack of accurate camera parameter estimation techniques (accurate camera calibration techniques may not provide accurate physical camera parameters). Until now, no accurate calibration results for any binocular head have been reported when the camera orientation and position and the motorized lens are moved simultaneously. In this paper, we have demonstrated how an active stereo vision system can be easily calibrated with very high accuracy. Consequently, many 3-D applications can directly make full use of an active binocular head. In addition to the proposal and demonstration of a four-stage calibration process, this paper has made three major contributions.

1) We have proposed an MFL camera model for cameras using motorized lens to adjust the focus setting. The MFL camera model was initially motivated by our theoretical analysis results (described in [29]), and it turned out to be similar to the Willson model [31]. We have made full use of the extra degrees of freedom in the estimation process (i.e., the linear dependency of some estimates of the camera parameters under small variations) by fixing the extrinsic parameters. This model is simple and especially suitable for cameras having a low distortion lens because the linear dependencies under small variations become invalid when the amount of lens distortion is large (refer to [29]).
2) We have proposed a new method for calibrating an MFL camera model to map the value of the focus setting to the intrinsic camera parameters. The calibration method for the adjustable camera model is efficient and accurate for two reasons. The first reason is that we use a linear solution to provide accurate initial values for the subsequent nonlinear optimization method. The second reason is that the estimations of the image center and effective focal length are made independent of the estimation of the camera orientation and the Z-component of the translation vector because the image center with the camera orientation and the effective focal length with the Z-component of the translation vector are linearly dependent under small variations.
3) We have shown that the kinematic parameters (including the head/eye relations) can be refined by using the nonlinear recursive least-square estimator. The nonlinear measurement function contains both the prediction of the 2-D image location and the epipolar error. Totally, 35 parameters are estimated simultaneously, including three joint level parameters. An inverse kinematics module has been implemented to fixate both the left and right cameras to the center point of the nine calibration points. Therefore, the whole refinement process can be executed automatically without manual operation.

In this paper, we have used real experiments to demonstrate that our calibration method can achieve approximately one pixel prediction error and 0.2 pixel epipolar error when all the joints including the left and right focus motors are moved simultaneously. This accuracy is good enough for many 3-D vision applications such as 3-D navigation, 3-D object tracking, or even 3-D reconstruction. Based on the highly accurate calibration, we have successfully done some experiments on panoramic stereo imaging and are now working on reconstruction of a 3-D environment model using our active binocular system.

APPENDIX
INVERSE KINEMATICS

For an active binocular head, the goal of inverse kinematics is to compute the joint values of the robot head such that the optical axes of both stereo cameras will pass through a given 3-D target point. The IIS head has totally eight degrees of freedom controlled by the following eight motors: the left and right focus motors, the left and right vergence motors, the tilt motor, the pan motor, and the X- and Y-motors. Therefore, to solve a unique set of joint values for control of our binocular head, eight constraints should be provided. The eight constraints chosen are as follows.

1) The 3-D coordinates of a target point with respect to the base frame of the binocular head: To direct the binocular head to a new observation configuration, the most efficient way is to choose a 3-D target point in the environment by means of its 3-D coordinates. Given a pair of stereo images, one can choose a 3-D target point by determining its corresponding image points in the stereo image pair and then compute the 3-D coordinates of the target point with respect to the base frame of the binocular head. This condition provides three constraints.
2) The joint values of the first two joints of the IIS head, i.e., the position of the X–Y table: This condition provides two constraints. The IIS head was originally designed to have an X–Y table to emulate an mobile robot platform. Thus, the inverse kinematics module has to be independent of the X–Y table such that when the IIS head is mounted on a real mobile robot platform, the inverse kinematics module will still work.
3) The gaze angle of the binocular head: The gaze origin is defined as the point locating on the tilt joint axis and being closest to the pan joint axis (see Fig. 21). The gaze angle is defined as the angle between the tilt joint axis and the line passing through the target point and the gaze origin. This condition provides one constraint.
4) The 3-D target point should be clearly focused in the stereo images: Owing to the MFL camera model we used, the nominal values of the extrinsic camera parameters are independent of the focus settings. Therefore, the joint values of the focus motors can be determined after all the joint values corresponding to the extrinsic parameters of the stereo cameras are determined. After the position and orientation of the stereo cameras are determined, the 3-D coordinates of the target point with respect to both the left and the right camera frames can be computed by using the kinematic model of

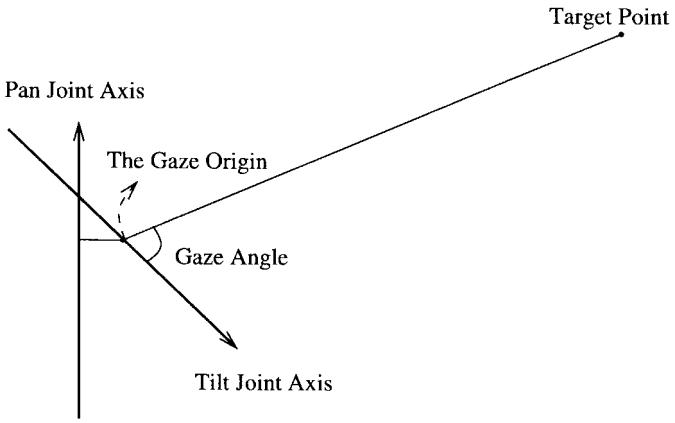

Fig. 21. Gaze origin and gaze angle.

the binocular head. Since the Z-component of the 3-D coordinates of the target point is essentially the object distance, the focus setting for focusing on the 3-D target point can be computed by using the lens equation (3). This condition provides two constraints.

In practice, the optical axes of stereo cameras may not intersect with each other at all. Therefore, in our implementation, we only constrain the optical axis of the left camera to exactly pass through the target point while the right optical axis is only required to be as close to the target point as possible. Each time we solve the inverse kinematics problem to control the binocular head, we first compute a simplified kinematic model based on the estimated kinematic parameters. From the simplified kinematic model, an analytic solution to the inverse kinematics problem can be easily computed. The exact solution to the inverse kinematics problem is then obtained by using an iterative nonlinear method. Current implementation of our inverse kinematics solution (including the computations of the simplified kinematic model, the initial solution and the iterative solution) take only a few milliseconds on a Sparc 20.

ACKNOWLEDGMENT

The authors are grateful to the anonymous reviewers for their valuable suggestions for paper revision.

REFERENCES

[1] J. Aloimonos and A. Badyopadhyay, "Active vision," in *1st Int. Conf. Computer Vision,* 1987, pp. 35–54.
[2] A. L. Abbott, "Dynamic integration of depth cues for surface reconstruction from stereo images," Ph.D. dissertation, Univ. Illinois, Urbana-Champaign, 1990.
[3] H. I. Christensen, "A low-cost robot camera head," *Int. J. Pattern Recognit. Artif. Intell.,* vol. 7, no. 1, pp. 69–87, 1993.
[4] N. J. Ferrier and J. J. Clark, "The Harvard binocular head," *Int. J. Pattern Recognit. Artif. Intell.,* vol. 7, no. 1, pp. 9–31, 1993.
[5] E. Milios and M. Jenkin, "Design and performance of Trish, a binocular robot head with torsional eye movements," *Int. J. Pattern Recognit. Artif. Intell.,* vol. 7, no. 1, pp. 51–68, 1993.
[6] K. Pahlavan and J. O. Eklundh, "Heads, eyes and head-eye systems," *Int. J. Pattern Recognit. Artif. Intell.,* vol. 7, no. 1, pp. 33–49, 1993.
[7] J. R. G. Pretlove and G. A. Parker, "The surrey attentive robot vision system," *Int. J. Pattern Recognit. Artif. Intell.,* vol. 7, no. 1, pp. 89–107, 1993.
[8] P. M. Sharkey, D. W. Murray, S. Vandevelde, I. D. Reid, and P. F. McLauchlan, "A modular head/eye platform for real-time reactive vision," *Mechatronics,* vol. 3, no. 4, pp. 517–535, 1993.
[9] R. Y. Tsai, "A versatile camera calibration technique for high-accuracy 3-D machine vision metrology using off-the-shelf TV cameras and lenses," *IEEE J. Robot. Automat.,* vol. RA-3, no. 4, pp. 323–344, 1987.
[10] J. Weng, P. Cohen, and M. Herniou, "Camera calibration with distortion models and accuracy evaluation," *IEEE Trans. Pattern Anal. Machine Intell.,* vol. 14, pp. 965–980, Oct6. 1992.
[11] G. Q. Wei and S. D. Ma, "Implicit and explicit camera calibration: Theory and experiments," *IEEE Trans. Pattern Anal. Machine Intell.,* vol. 16, no. 5, pp. 469–480, 1994.
[12] U. R. Dhond and J. K. Aggarwal, "Structure from stereo—A review," *IEEE Trans. Syst., Man, Cybern.,* vol. 19, no. 6, pp. 1489–1510, 1989.
[13] S. B. Marapane and M. M. Trivedi, "Multi-primitive hierarchical (mph) stereo analysis," *IEEE Trans. Pattern Anal. Machine Intell.,* vol. 16, pp. 227–240, Mar. 1994.
[14] E. P. Krotkov, *Active Computer Vision by Cooperative Focus and Stereo.* New York: Springer-Verlag, 1989.
[15] R. Deriche, Z. Zhang, Q. T. Luong, and O. Faugeras, "Robust recovery of the epipolar geometry for and uncalibrated stereo rig," in *Lecture Notes Comput. Sci., Comput. Vision—ECCV'94,* J.-O. Eklundh, Ed., 1994, vol. 800, pp. 565–576.
[16] F. Du and M. Brady, "Self-calibration of the intrinsic parameters of cameras for active vision systems," *Comput. Vis. Pattern Recognit.,* pp. 477–482, 1993.
[17] S. D. Ma, "A self-calibration technique for active vision systems," *IEEE Trans. Robot. Automat.,* vol. 12, pp. 114–120, Feb. 1996.
[18] Z. Zhang, Q. T. Luong, and O. Faugeras, "Motion of an uncalibrated stereo rig: Self-calibration and metric reconstruction," *IEEE Trans. Robot. Automat.,* vol. 12, pp. 103–113, Feb. 1996.
[19] M. Li, "Camera calibration of the kth head/eye system," Tech. Rep. CVAP147, Computation. Vision Active Perception Lab., Dept. Numerical Anal. Comput. Sci., Royal Inst. Technol., Stockholm, Sweden, 1994.
[20] M. Li, D. Betsis, and J. M. Lavest, "Kinematic calibration of the kth head-eye system," Tech. Rep. CVAP171, Computation. Vision Active Perception Lab., Dept. Numerical Anal. Comput. Sci., Royal Inst. Technol., Stockholm, Sweden, 1994.
[21] M. Li and J. M. Lavest, "Some aspects of zoom-lens camera calibration," Tech. Rep. CVAP172, Computation. Vision Active Perception Lab., Dept. Numerical Anal. Comput. Sci., Royal Inst. Technol., Stockholm, Sweden, 1995.
[22] P. F. McLauchlan and D. W. Murray, "Active camera calibration for a head-eye platform using the variable state-dimension filter," *IEEE Trans. Pattern Anal. Machine Intell.,* vol. 18, pp. 15–22, Jan. 1996.
[23] S. W. Shih, J. S. Jin, K. H. Wei, and Y. P. Hung, "Kinematic calibration of a binocular head using stereo vision with the complete and parametrically continuous model," *SPIE Intell. Robots Comput. Vision XI,* vol. 1825, pp. 643–657, 1992.
[24] S. W. Shih, Y. P. Hung, and W. S. Lin, "New closed-form solution for kinematic parameter identification of a binocular head using point measurements," *IEEE Trans. Syst., Man, Cybern. B,* vol. 28, pp. 258–267, Apr. 1998.
[25] G. S. Young, T. H. Hong, M. Herman, and J. C. S. Yang, "Kinematic calibration of an active camera system," in *IEEE Int. Conf. Computer Vision Pattern Recognition,* 1992, pp. 748–751.
[26] R. Y. Tsai and R. K. Lenz, "A new technique for fully autonomous and efficient 3-D robotics hand/eye calibration," *IEEE Trans. Robot. Automat.,* vol. 5, pp. 345–358, June 1989.
[27] R. K. Lenz and R. Y. Tsai, "Techniques for calibration of the scale factor and image center for high accuracy 3-D machine vision metrology," *IEEE Trans. Pattern Anal. Machine Intell.,* vol. 10, pp. 713–720, May 1988.
[28] ———, "Calibrating a Cartesian robot with eye-on-hand configuration independent of eye-to-hand relationship," *IEEE Trans. Robot. Automat.,* vol. 11, pp. 916–928, 1995.
[29] S. W. Shih, Y. P. Hung, and W. S. Lin, "Accuracy analysis on the estimation of camera parameters for active vision systems," Tech. Rep. TR-IIS-96-006, Inst. Inform. Sci., Acdemia Sinica, Nankang, Taipei, Taiwan, R.O.C., 1996 (ftp://140.109.23.242/pub/TechRpt/TR-IIS-96-006.ps.gz).
[30] K. A. Tarabanis, R. Y. Tsai, and D. S. Goodman, "Modeling of a computer-controlled zoom lens," in *IEEE Int. Conf. Robotics Automation,* 1992, pp. 1545–1551.
[31] R. G. Willson, "Modeling and calibration of automated zoom lenses," Ph.D. dissertation, Dept. Elect. Comput. Eng., Carnegie Mellon University, Pittsburgh, PA, 1994.
[32] S. W. Shih, Y. P. Hung, and W. S. Lin, "Comments on a 'linear solution to kinematic parameter identification of robot manipulator' and some modifications," *IEEE Trans. Robot. Automat.,* vol. 11, no. 5, pp.

777–780, 1995.
[33] ——, "Error analysis on closed-form solutions for kinematic calibration," Tech. Rep. TR-IIS-96-014, Inst. Inform. Sci., Acdemia Sinica, Nankang, Taipei, Taiwan, R.O.C., 1996, (ftp://140.109.23.242/pub/TechRpt/TR-IIS-96-014.ps.gz).
[34] C. C. Wang, "Extrinsic calibration of a vision sensor mounted on a robot," *IEEE Trans. Robot. Automat.*, vol. 8, pp. 161–175, Apr. 1992.
[35] H. Zhuang, K. Wang, and Z. S. Roth, "Simultaneous calibration of a robot and a hand-mounted camera," *IEEE Trans. Robot. Automat.*, vol. 11, pp. 649–660, Oct. 1995.
[36] P. Y. Tsai, "Calibration of a camera with motorized lens and its application to depth from focusing," M.S. thesis, Dept. Comput. Sci. Inform. Eng., National Taiwan Univ., Taipei, Taiwan, R.O.C., 1995.
[37] Z. S. Roth, B. W. Mooring, and B. Ravani, "An overview of robot calibration," *IEEE J. Robot. Automat.*, vol. RA-3, pp. 377–385, Oct. 1987.
[38] S. W. Shih, Y. P. Hung, and W. S. Lin, "Calibration of an active binocular head," Tech. Rep. TR-IIS-96-007, Inst. Inform. Sci., Acdemia Sinica, Nankang, Taipei, Taiwan, R.O.C., 1996, (ftp://140.109.23.242/pub/TechRpt/TR-IIS-96-007.ps.gz).
[39] S. Umeyama, "Least-squares estimation of transformation parameters between two point patterns," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, pp. 376–380, Apr. 1991.
[40] J. M. Mendel, *Lessons in Digital Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1987.

Sheng-Wen Shih received the M.S. and Ph.D. degrees in electrical engineering from National Taiwan University, Taipei, Taiwan, R.O.C., in 1990 and 1996, respectively.

From 1990 to 1997, he was with the Laboratory of Intelligent System, Institute of Information Science, Academia Sinica, Taipei, Taiwan. He is currently an Assistant Professor in the Department of Computer Science and Information Engineering, National Chi Nan University, Nantou, Taiwan. His current research interests are in active vision, virtual reality, and human–computer interface.

Yi-Ping Hung (S'84–M'89) received the B.S. in electrical engineering from National Taiwan University, Taipei, Taiwan, R.O.C., in 1982. He received the M.S. from the Division of Engineering, the M.S. from the Division of Applied Mathematics, and the Ph.D. from the Division of Engineering at Brown University, Providence, RI, in 1987, 1988, and 1989, respectively.

He spent the summers of 1987 and 1988 at Philips Laboratories, Briarcliff Manor, NY, working on robot vision and automation. From October 1989 to February 1990, he worked as a Postdoctoral Research Associate in the Laboratory for Engineering Man/Machine Systems (LEMS), Brown University. He joined the Institute of Information Science, Academia Sinica, Taipei, in February 1990. He served as the Deputy Director of the Institute of Information Science from 1996 to 1997, and was promoted to a Research Fellow in 1997. He has also been an Adjunct Associate Professor with the Department of Computer Science and Information Engineering, National Taiwan University, since 1990. He has published more than 60 technical papers in the fields of computer vision, pattern recognition, image processing, and robotics. In addition to the above topics, his current research interests include virtual reality, visual surveillance, human–computer interface, and visual communication (especially in video technologies related to MPEG-4 and MPEG-7).

Dr. Hung received the Young Researcher Publication Award of Academia Sinica in 1997

Wei-Song Lin was born in Taiwan, R.O.C., in 1951. He received the B.S. degree in engineering science and the M.S. degree in electrical engineering from National Cheng Kung University, Taipei, Taiwan, in 1973 and 1975, respectively. He received the Ph.D. degree from the Institute of Electrical Engineering, National Taiwan University, Taipei, in 1982.

Beginning in 1977, he worked in the Telecommunication Research and Training Institutes as a Research Engineer developing computer technologies. He was an Associate Professor with the Institute of Electrical Engineering from 1983 to 1987. From 1982 to 1984, he was also the head of the Electronic Instrument Division, Ching Ling Industrial Research Center, where he mainly worked on design and implementation of instruments. Since 1987, he has been a Professor at National Taiwan University. His current research interests are in the field of computer control, sensing, and intelligent processing.

Dr. Lin received the Excellent Research Award from the National Science Council four times and the Distinguished Paper Award from the Association of Chinese Electrical Engineers. He is a member of the International Association of Science and Technology for Development (IASTEP).