Calibration of an Active Binocular Head

Sheng-Wen Shih[†], Yi-Ping Hung[‡], and Wei-Song Lin[†]

[†]Institute of Electrical Engineering, National Taiwan University, Taipei, Taiwan.

[‡]Institute of Information Science, Academia Sinica, Nankang, Taipei, Taiwan.

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 MFL (Motorized-Focus Lens) camera model which assumes constant nominal extrinsic parameters. The advantage of having constant extrinsic parameters is to have a simple head/eye relation. Second, a calibration method for the MFL camera model is proposed in this paper, which separates the estimation of the image center and effective focal length from the estimation of the camera orientation and position. This separation has been proved to be crucial; otherwise, the 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 39 *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 3D vision applications, such as 3D navigation, 3D object tracking, and even 3D reconstruction.

Keywords

Active Vision, Binocular Head, Motorized Lens Calibration, Head/Eye Calibration, Head 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 the 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 were built for investigating the active vision problems, e.g, [2], [3], [4], [5], [6], [7] and [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 Figure 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 joint 2 and joint 1).

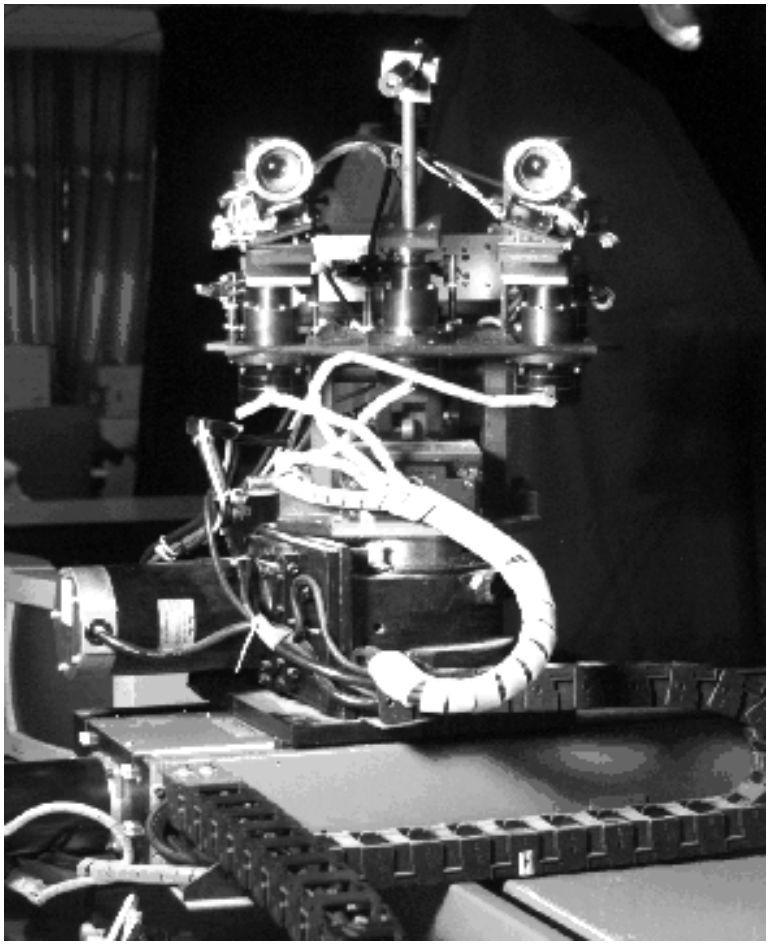

Fig. 1. A picture of the IIS head.

A. Importance of Calibration

Accurate camera calibration can greatly simplify the solutions of many important vision problems, such as the stereo vision problem, the 3D visual tracking problem, the mobile-robot visual guidance problem, the 3D reconstruction problem, the 3D 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 3D estimation error. Therefore, extensive work has been devoted to developing camera calibration techniques (refer to [9], [10] and

[11] for detailed survey of these calibration methods). Since accurately calibrating a passive stereo vision system is a fairly easy and efficient work, most 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 in 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, because knowing accurate model of an active stereo vision system can simplify the solutions of 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], [20], [21], Mclauchlan and Murray [22], Shih *et. al.* [23], [24] and Young *et. al.* [25]. Also, there are some other people using the techniques of the calibration-trio proposed by Tsai [9] and Tsai and Lenz [26], [27] and [28].

B. Difficulties of Calibrating an Active Camera

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

1. High complexity: The calibration of an active vision system consists of three calibration processes, namely, the motorized lens calibration, the kinematic calibration and the head/eye calibration. Each calibration process is itself a complex problem already.
2. Ultra-high accuracy requirement: The angular resolution (i.e, the view-angle of one pixel) of a camera is roughly several fractions of one milli-radian. 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 non-geometric terms in the kinematic 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 certain camera parameters, e.g, the correlation between the image center and the camera orientation, an estimate of a set of camera parameters

which yields very small 3D 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 sub-problems: the motorized lens calibration, the kinematic calibration and the head/eye calibration.

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

In the 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], Lenz and Tsai [28], Shih *et. al.* [23], [24], [32], and Young *et. al.* [25]. Most of the work is focused on the derivation of closed-form solutions. However, the major problem is that, due to the measurement noise, the kinematic parameter estimates obtained by using

closed-form solutions are not accurate enough to fit the ultra-high accuracy requirement in the applications of active binocular heads [33].

In the 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 to be the most accurate one (refer to [34] and [20]). 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 an accurate enough initial value 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 but the trajectories of some static feature corners observed in the scene. However, their kinematic 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 2D image prediction error (refer to section IV). Zhuang *et. al.* [35] also proposed to simultaneously calibrate a robot and a hand-mounted camera to avoid the propagation of estimation error 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 avoid directly estimating all the active camera parameters simultaneously, due to the correlation between certain extrinsic and intrinsic camera parameters. Instead, based on our theoretical analysis [29], some of the intrinsic camera parameters had better be estimated independently to the extrinsic camera parameters; otherwise, the estimates would be very sensitive to noise.

In this paper, a four-stage calibration process is used to calibrate the IIS head, which can achieve a root mean square prediction error of approximately 1 pixel and an epipolar error of about 0.2 pixel, even when all the joints are moved simultaneously. Figure 2 shows the schematic diagram of the four calibration stages. The first stage is to calibrate the motorized lens by using the MFL (*Motorized-Focus Lens*) camera model, which identifies

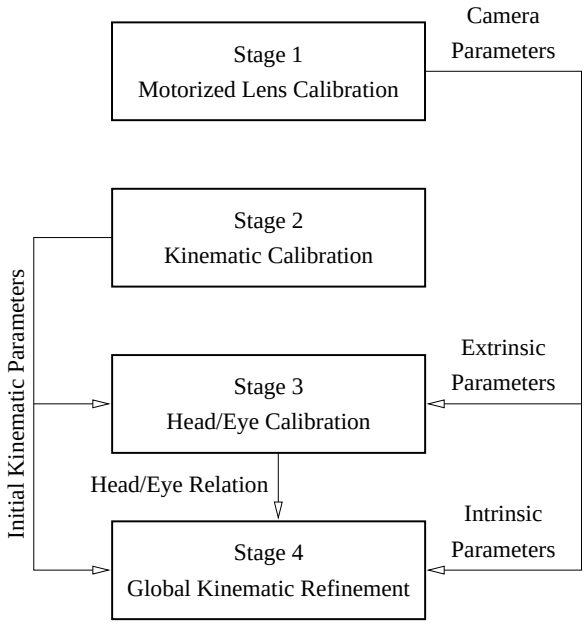

Fig. 2. The schematic diagram of the four-stage calibration process for an active binocular head.

the functions of the intrinsic camera parameters with respect to the focus settings. The second stage is to estimate the kinematic parameters by using the closed-form solution proposed in [24]. The third stage is to estimate the head/eye relation by using either a simple method, using *Coordinate Measurement Machines*, which will be described in section IV, or the well-known Tsai and Lenz method [26]. The forth stage is to refine all the *kinematic* parameters estimated in the previous two stages by using a nonlinear recursive least-square estimator. This paper describes in details the first stage and the fourth stage.

This paper is origanized as follows. Section II describes the MFL camera model and a calibration method. Section III describes the kinematic model and inverse kinemtaics of a binocular head. Section IV describes the global kinematic calibration process, including the nonlinear recursive least-square estimator. Section V describes 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 3D space, and (x_O, y_O, z_O) be its coordinates, in millimeters, with respect to a fixed OCS (*Object Coordinate System*). 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 twelve 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 3D and 2D 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 by 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 by 1 translation vector, where the prime denotes the *transpose* operation. Notice that three of the camera parameters, i.e, the effective focal length, f, 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 length (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].

B. Motorized Lens

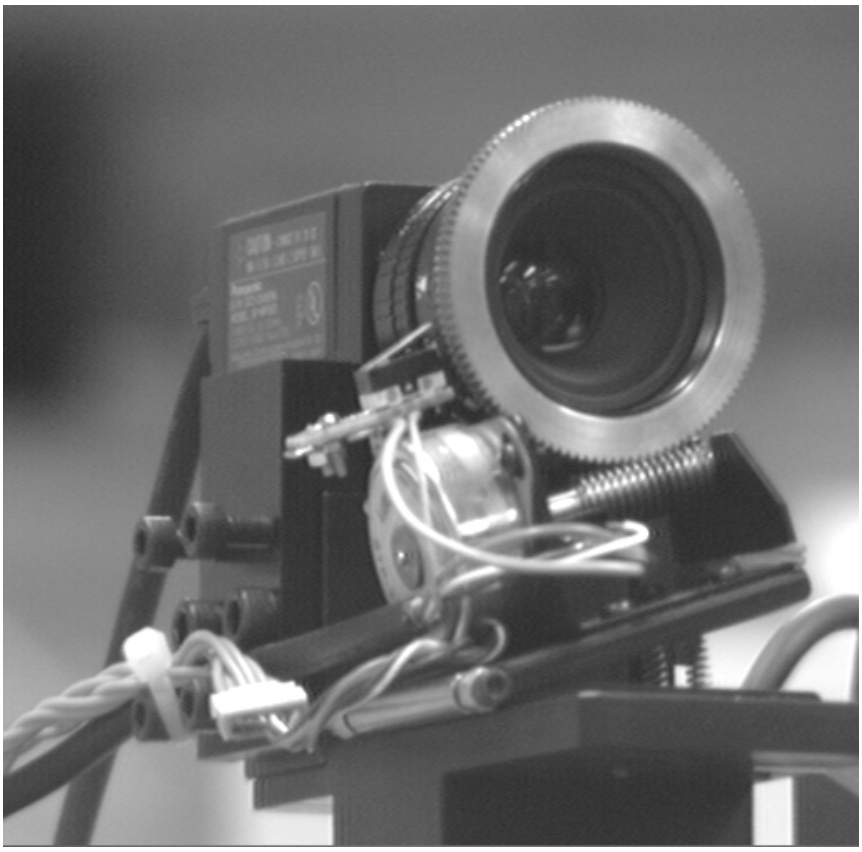

Fig. 3. The motorized lens of the IIS head.

Figure 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 zoom

function is usually more accurate (due to smaller distortion), more compact and lighter-in-weight than a zoom lens. Therefore, we recently chose 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 the distance from 0.3 meter to infinity, and its nominal focal length is 25 millimeters. Stepping motors were 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 $\left[F, \frac{Fd_0}{d_0 - F}\right]$, where d_0 is the minimal object distance which is approximately 0.3 meter 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 the 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 stopped.

C. The 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 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 were widely used.

Recently, we observed that accurate camera calibration techniques yielding very small 2D residual error did 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 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 as follows:

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

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 u_0 and $\delta\phi_y$, v_0 and $\delta\phi_x$, and f and t_x are linearly dependent for small variation (see [29]). This phenomenon can cause serious problem in motorized lens calibration, and there are two different approaches for solving this problem. The first approach is to develop some new techniques for acquiring more accurate estimate of the extrinsic and intrinsic camera parameters. The second approach is more interesting and will be described in more details in this paper. This approach makes full use of the extra-degrees of freedom in the calibration (i.e, the linear dependency of some camera parameters for small variations). For example, if we fix the camera orientation at a slightly incorrect value during the camera calibration, some other camera parameters would be deviated from their true value to compensate for the effect of small orientation error when minimizing the residual error. Motivated by this theoretical analysis results, we proposed a camera model for motorized-focus lens, which turned out to be similar to the Willson model [31]. Willson has used these extra-degrees of freedom without theoretical analysis and achieved very high accurate zoom lens model. In Willson's camera model for 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 to the lens settings. Moreover, the relation between the remaining camera parameters and the lens settings (focus, zoom and aperture settings) are approximated by multi-variable polynomials.

According to a series of preliminary experiments about our MFL camera held in our laboratory [36], we found that

- the effective focal length varies linearly with respect to the focus setting,
- the camera parameters are independent to 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 to the focus setting.
3. The position and orientation of the CCS are independent to 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 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 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, the calibration accuracy can be dramatically improved. The calibration procedure of the MFL camera model is described in the following subsection.

D. Calibration of the MFL Camera Model

To calibrate the motorized lens, we 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 the calibration accuracy, one should collect a large number of 2D-3D 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 relation 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, equations (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 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:

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

   where N is the number of 2D-3D 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 changing focus setting is more regular and thus can be well approximated by a low-order polynomial. Whereas, the variation of the other parameters are complicated that are better represented as look-up tables than fitted by polynomials.

III. Kinematic Model and Inverse Kinematics of a Binocular Head

A. Kinematic Model of the IIS head

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

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

and

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

where

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

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. Figure 4 gives an illustration of the relationship between the joint frames and the transformation matrices for the IIS head.

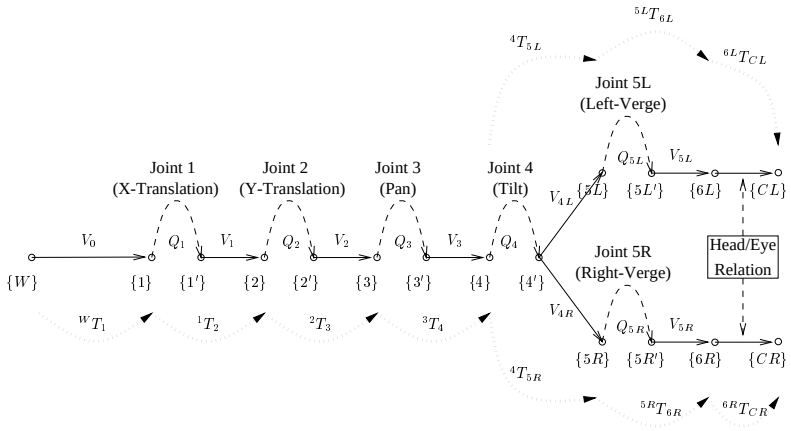

Fig. 4. An illustration of the joint frames and the transformation matrices for the IIS head.

Notice that the unknown kinematic parameters are all contained in the shape matrix and 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 world-to-base transformation matrix, V_0), respectively.

B. Eliminating the Effects of Gear Backlash

In general, the effects of gear backlash are classified into non-geometric 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 expansive backlash-free gear to avoid all the troubles caused by gear backlash. However, one may also eliminate the effects of backlash by either of the following two ways. The first way 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 would cause the direction of motion unpredictable before the motor stopped. The second way is to calibrate the amount of backlash described in Appendix-A.

C. 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 3D 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 controlling our binocular head, eight constraints should be provided. The eight constraints are chosen to be:

1. The 3D coordinates of a target point with respect to the base frame of the binocular head: For directing the binocular head to a new observation configuration, identifying the interesting point in the environment by its 3D coordinates is the most efficient way. Even for an unknown environment, one can also specify the interesting point from the first snapshot of the scene and translating it to the 3D coordinates 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 the X-Y table to emulate an mobile robot platform. Thus the inverse kinematics module has to be independent to the X-Y table such that when the IIS head is mounted on a real mobile robot platform, the inverse kinematics module still works.
3. The gaze angle of the binocular head: Define the gaze origin to be the point on the tilt joint axis which is closest to the pan joint axis (see Figure 5). 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 3D 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 to 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 3D 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 the binocular head. Since the Z-component of the 3D coordinates of the target point is essentially the object distance, the focus setting for focusing at the 3D target point can be computed by using the lens equation (3). This condition provides two constraints.

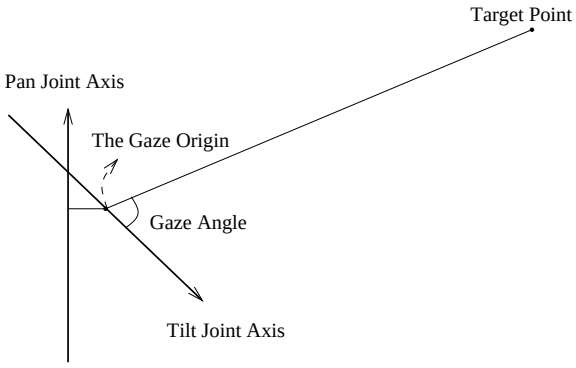

Fig. 5. The gaze origin and the gaze angle

In practice, the optical axes of the 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 when we solving 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 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) cost only a few milliseconds in Sparc 20.

IV. Global Kinematic Refinement

A. Calibration Setup

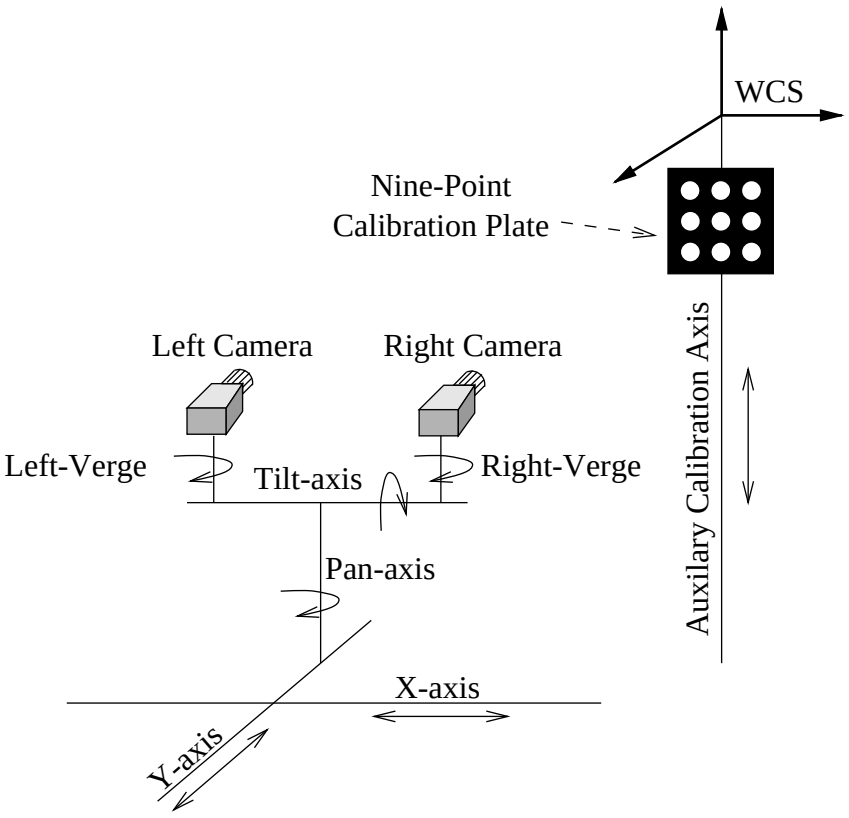

Fig. 6. A schematic diagram of the calibration setup.

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 comparing to the joint space, then after the calibration, the binocular head would have large probability of working outside the small calibration

range, which may lead to larger kinematic inaccuracy in general. Hence, for obtaining 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 as shown in Figure 6. A black calibration plate having nine white disks located at nine grid positions were mounted to the auxiliary translation stage. The X-Y plane of the WCS (*World Coordinate System*) is defined to be on the calibration plate when the auxiliary prismatic joint is at its home position. The translation direction of the auxiliary prismatic joint with respect to the WCS is accurately calibrated by using a CMM (Coordinate Measurement Machine). Therefore, the 3D coordinates of the centroids of the nine disks on the calibration plate with respect to the WCS are always known. The auxiliary prismatic joint and the calibration plate are then used for the head calibration. They also can be used for providing an initial estimate of the head/eye relation and the head-base-to-WCS relation.

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 the availability of highly accurate 3D measurements, whereas when using vision system as the 3D measurement device, the lighting condition, background and shape of the calibration object will all affect the calibration accuracy. Notice that, since we have a nonlinear refinement process in our four-stage calibration method, a low cost CMM is adequate for providing accurate enough initial estimate of the kinematic parameters. Nevertheless, if no CMM is available, then the stereo cameras of the binocular head can be used as the measurement device, which also can be very accurate if the setup are carefully operated [24].

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 3D point mea-

surements for the kinematic calibration (similar to [24] except that the 3D measurements are obtained by 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, ${}^{CMM}T_1$, 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, ${}^{W}T_{CMM}$. Hence, the base-to-world transformation, ${}^{W}T_1$, can be computed from ${}^{CMM}T_1$ and ${}^{W}T_{CMM}$. Then, the following procedure can be used to estimate an initial value of the head/eye relations (refer to Figure 7 for the relation between the transformation matrices):

1. Choose a configuration of the binocular head and perform a stereo camera calibration.
2. Keep the two vergence joint unchanged (so that the stereo parameters obtained in the previous step are still valid, for computing 3D measurement using triangulation) and move the binocular head such that the nine-point calibration plate shown in Figure 6 is in the filed of view of the stereo cameras.
3. Compute the 3D coordinates of the nine calibration points with respect to the left and right camera frame, respectively, using triangulation. Also, record the corresponding 3D coordinates of the nine calibration point with respect to the WCS.
4. Compute the transformation matrices from the WCS to the left and right camera frame, ${}^{CL}T_W$ and ${}^{CR}T_W$, based on the Umeyama method [38] by using the three sets of 3D coordinates of the nine calibration points with respect to the left camera frame, right camera frame and the WCS.
5. Compute the left and right head/eye relationships as follows:

$$ {}^{6L}T_{CL} = \left({}^{W}T_1 \, {}^{1}T_{6L}\right)^{-1} {}^{W}T_{CL}, \tag{8} $$

$$ {}^{6R}T_{CR} = \left({}^{W}T_1 \, {}^{1}T_{6R}\right)^{-1} {}^{W}T_{CR}, \tag{9} $$

   where ${}^{1}T_{6L}$ and ${}^{1}T_{6R}$ can be computed based on the forward kinematic model, equations (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 $\left(V_{5L} \, {}^{6L}T_{CL}\right)$ and $\left(V_{5R} \, {}^{6R}T_{CR}\right)$, respectively.

Notice that when no CMM is available, the head/eye relation can also be estimated by using the well known Tsai and Lenz method [26], and then the base-to-WCS relation can be computed by using a pair of 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 ${}^{6R}T_{CR}$.

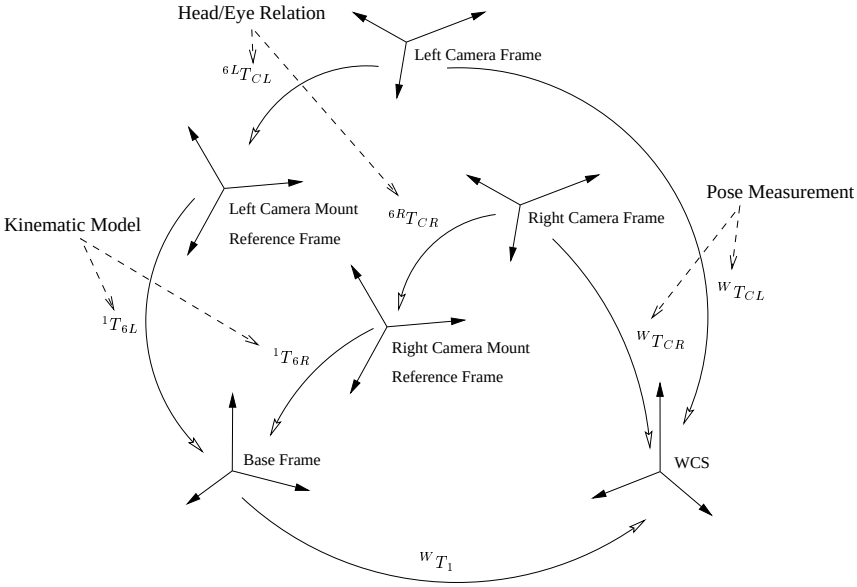

Fig. 7. The relationships between the base frame, camera mount reference frame, camera frame, and WCS.

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

The total number of unknown parameters to be estimated is 39, which include the transformation matrix from the WCS to the base frame of the binocular head (six parameters, including the orientation parameters of the X prismatic joint), the orientation of the Y prismatic joint axis (two parameters), the position and orientation of those revolute joint axes of Pan, Tilt, Left-Verge and Right-Verge (sixteen parameters, four for each joint), the two transformation matrices of the effective head/eye relations (twelve 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 for achieving higher payload. Therefore, three joint level parameters have to be estimated for mapping the position of the leverage translation stage into the revolute joint angle.

D. The 2D Prediction Error and the Epipolar Error

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

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

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

where $P_{CL}(i)$ and $P_{CR}(i)$ are the 3D coordinates of the ith calibration point with respect to the left and right camera reference frames, respectively. For each of the nice calibration points, its 3D coordinates with respect to the camera frame can be computed by using equations (10) and (11). Then, its image location can be predicted by using the camera model described in equations (1) and (2). The 2D prediction error is defined as the difference between the observed image location and the predicted image location of a calibration point. For convenience, let $h_{pL}\left(P_W(i), q(t); \Theta\right)$ and $h_{pR}\left(P_W(i), q(t); \Theta\right)$ denote the predicted left and right 2D image location vectors derived from equations (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 39 parameters to be estimated. In addition to the 2D prediction error mentioned above, we also have to consider the epipolar error, which is very important for stereo vision systems. Let $h_{eL}\left(p_L(i), p_R(i), q(t); \Theta\right)$ and $h_{eR}\left(p_L(i), p_R(i), q(t); \Theta\right)$ be the left and right epipolar error, where $p_L(i)$ and $p_R(i)$ are the stereo correspondence of the ith calibration point from the left and right image plane (details of the definition and computation of the epipolar error is described in Appendix-B). Notice that the ideal values of $h_{pL}\left(P_W(i), q(t); \Theta\right)$, $h_{pR}\left(P_W(i), q(t); \Theta\right)$, $h_{eL}\left(p_L(i), p_R(i), q(t); \Theta\right)$ and $h_{eR}\left(p_L(i), p_R(i), q(t); \Theta\right)$ 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:

$$z(t) = h\left(q(t); \Theta\right) + w(t), \tag{13}$$

where

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

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. When using a batch nonlinear optimization process, the number of measurements should be made as large as possible in order to reduce estimation error. However, since the initial estimate of the kinematic parameters may not accurate enough for automatic data acquisition, the process of collecting large amount of the calibration data for batch processing should be performed manually, which is a time-consuming and error-prone task. Therefore, we choose to use a recursive least-square estimator to avoid the above-mentioned problem. To use a recursive least-square estimator, the objective function is defined as follows:

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

where

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

$\Theta(0)$ is the initial estimate of the unknown parameters and $P(0)$ is its covariance matrix.

The update equations for a nonlinear recursive least-square estimator are as follows (refer to Mendel [39]):

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

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

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

V. Experiments

A. Calibration Results of the Adjustable 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. Totally 1,000 2D-3D pairs of calibration points were used in this calibration to ensure parameter accuracy. The estimated constant extrinsic parameters were then used to estimate the adjustable intrinsic parameters in the MFL camera model. Figures 8–13 show the calibration results of the adjustable camera model. Notice that for testing 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. In Figures 8–13, the four curves obtained from four real experiments are closed to each other, which implies that the proposed method is very accurate.

Figure 8 shows the variations of the image centers of the *left* camera subject to 26 focus motor positions in four different real experiments. Figure 9 shows the trajectory of the estimated image center for the first real experiment. Similarly, Figures 10 and 11 show the estimated image center of the *right* camera subject to 26 focus settings. Figures 12 and 13 show the variations of the left and right effective focal lengths and the left and right radial

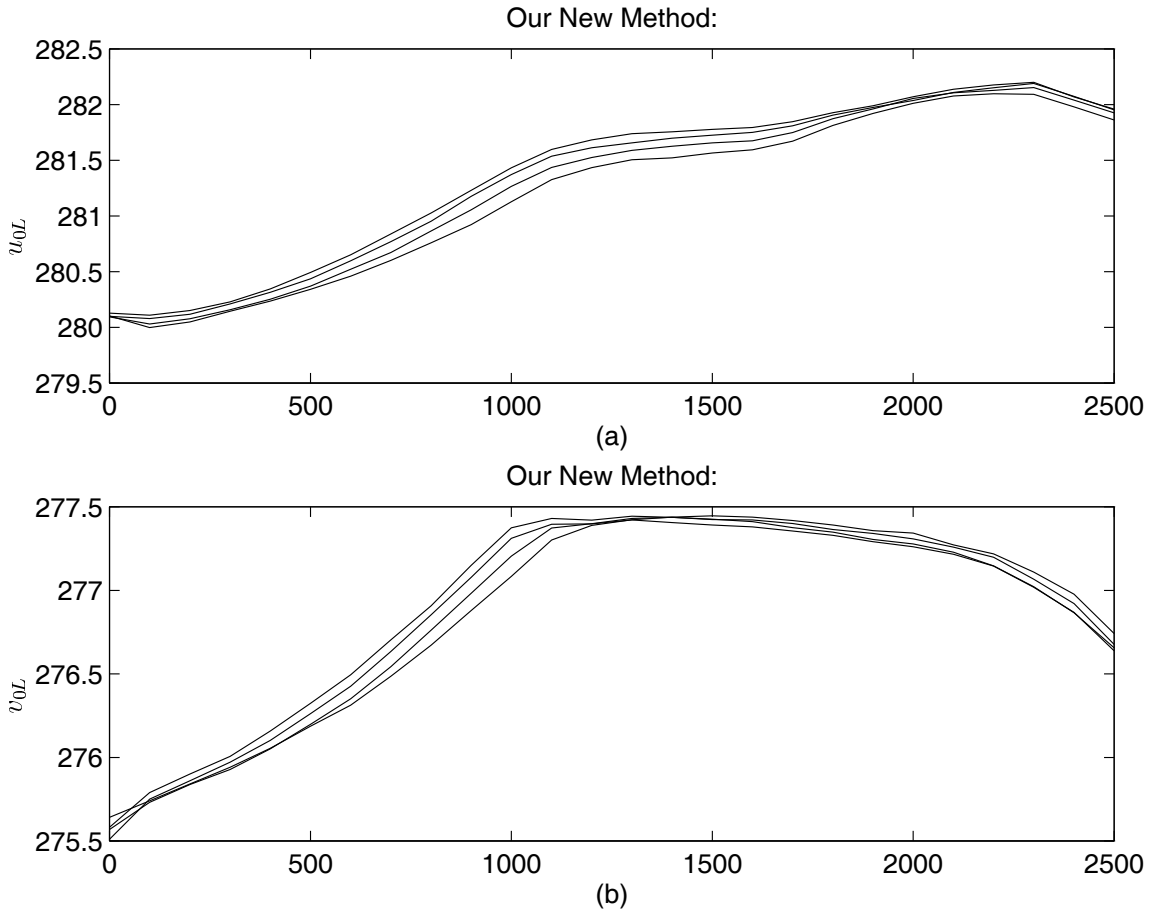

Fig. 8. Variations of the left image center subject to the focus motor position estimated by using the proposed method. The horizontal axes are the motor position and the vertical axes are the coordinates of the image center (in pixels). Four curves represent for the results of four different real experiments. (a) u_{0L} versus the left focus motor position. (b) v_{0L} versus the left focus motor position.

lens distortion coefficients, respectively. Notice that 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 lens distortion coefficient will cause approximately the same amount of distortion when using $(1-\kappa\rho^2)$ in equations (1) and (2). Therefore, the calibration results using any of the four curves is all acceptable. As to the effective focal lengths, linear functions are fitted, and the resulted relations for the left and right cameras in our experiment are

$$f_L\left(M_L\right) = 25.01005 + 5.191702 \times 10^{-4}\ M_L, \tag{19}$$

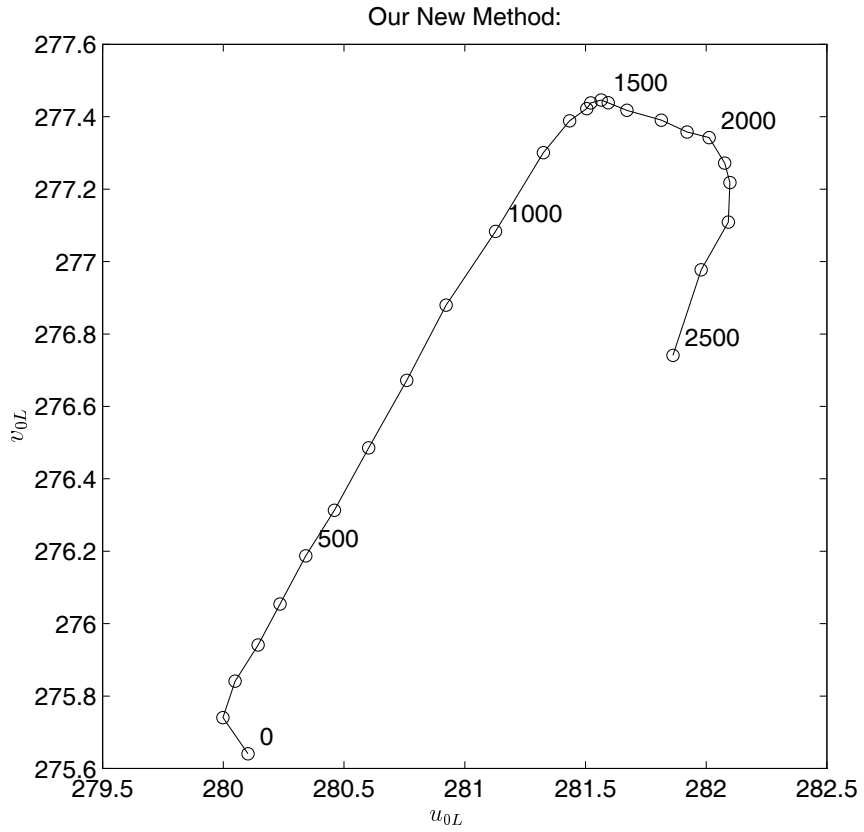

Fig. 9. Trajectory of the left image center subject to the focus motor position estimated by using the proposed method. Values of the focus motor positions are marked every five samples around the corresponding estimated image centers.

and

$$f_R(M_R) = 25.15386 + 5.241816 \times 10^{-4}\ M_R, \tag{20}$$

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 $\kappa_R(M_R)$, are 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 direct method, all the camera parameters are estimated simultaneously

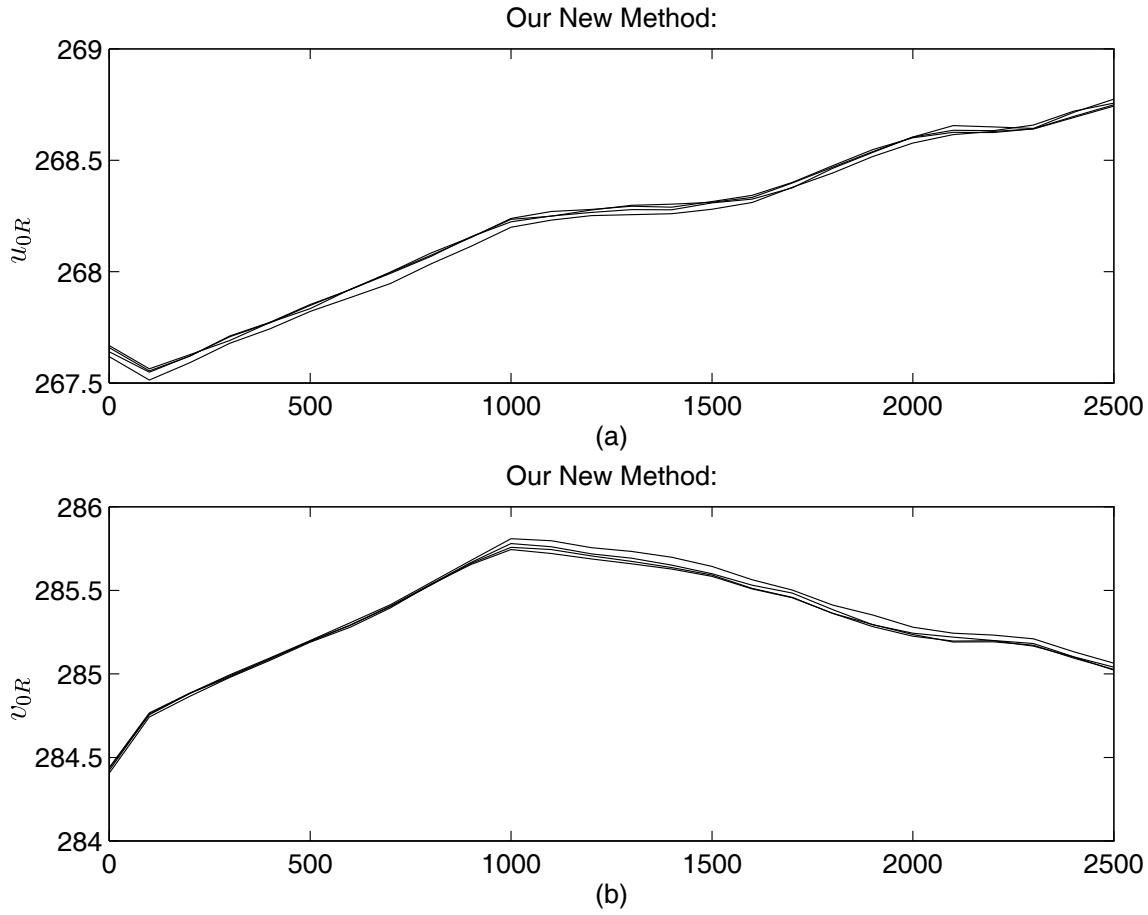

Fig. 10. Variations of the right image center subject to the focus motor position estimated by using the proposed method. The horizontal axes are the motor position and the vertical axes are the coordinates of the image center (in pixels). Four curves represent for the results of four different real experiments. (a) u_{0R} versus the right focus motor position. (b) v_{0R} versus the right focus motor position.

all the time, whereas, in our new method, the extrinsic parameters are fixed and only some of the intrinsic parameters are estimated. For testing the repeatability of the direct method, we also performed four real experiments, and some typical results are shown in Figures 14–16. Figure 15 shows the trajectory of the estimated image center subject to change of focus setting for the first real experiment. Notice that the trajectory is quite irregular and varies with 20–30 pixels, which is large compared to the more regular trajectories with 1–3 pixels variation in Figures 9 and 11. Also, from these figures, the intrinsic parameters vary dramatically subject to the change of focus motor position.

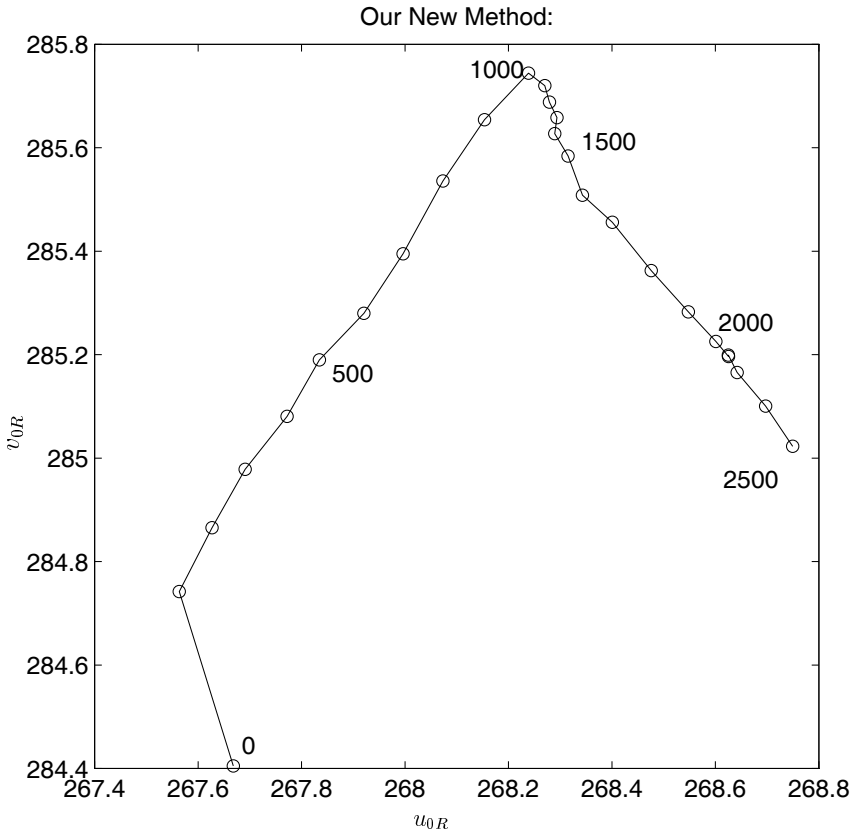

Fig. 11. Trajectory of the right image center subject to the focus motor position estimated by using the proposed method. Values of the focus motor positions are marked every five samples around the corresponding estimated image centers.

Therefore, the interpolation results are less reliable and a comprehensive lookup table is required for this approach which makes itself impractical.

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, $P(0)$, 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 larger initial variances might cause faster convergence rate; however, it might also cause instability problem, since the number of unknowns is very large. Therefore, a smaller initial variances may yield a slower convergence rate, but

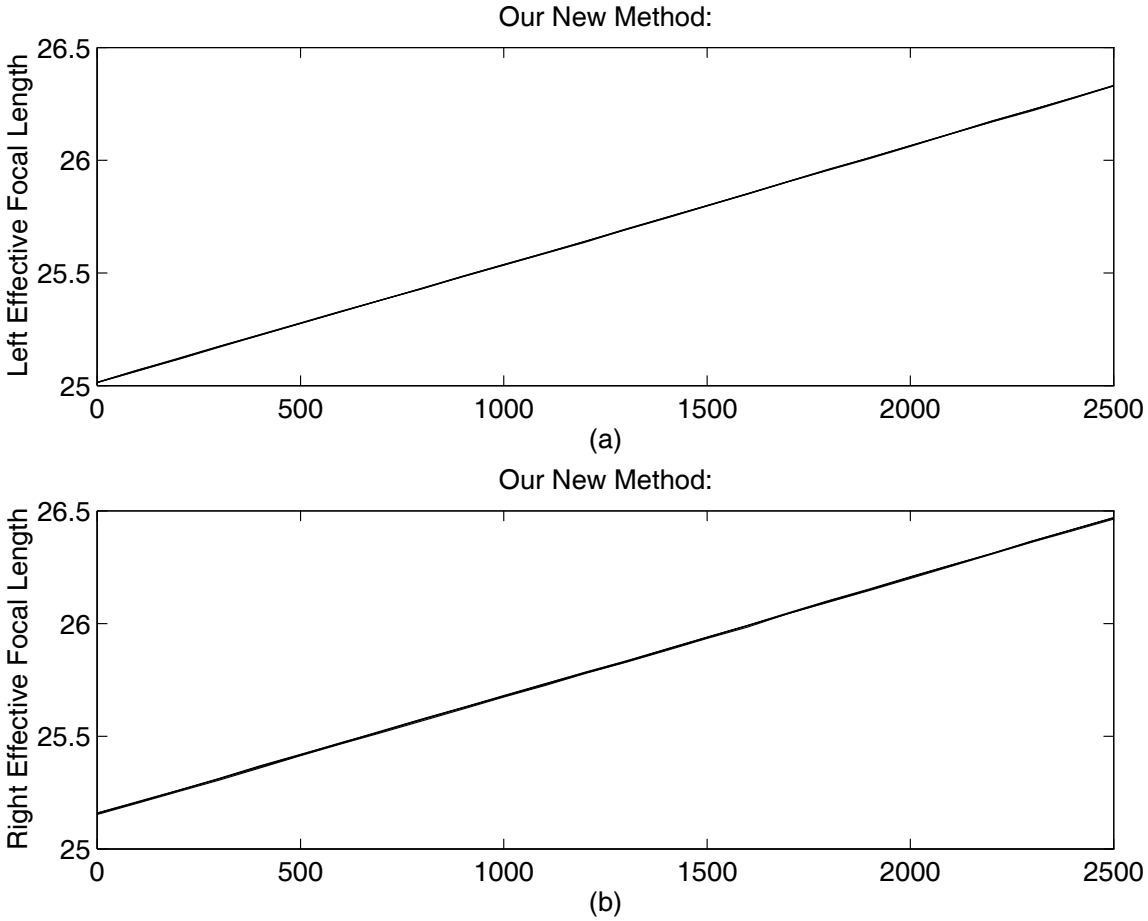

Fig. 12. Variations of the left and right effective focal lengths subject to the focus motor position estimated by using the proposed method. The horizontal axes are the motor position and the vertical axes are the amount of effective focal lengths (in millimeters). Four curves represent for the results of four different real experiments. (a) f_L versus the left focus motor position. (b) f_R versus the right focus motor position.

the recursive filter will be more stable.

C. Acquiring Calibration Data

In order to calibrate the IIS head, we have to generate a sequence of joint values to control the binocular head such that the calibration plate are visible to the stereo cameras all the time. This task can be accomplished with an inverse kinematics module. According to our inverse kinematics scheme described in section III-C, the gaze angle and the joint values of the X-Y table and of the auxiliary Z-axis are chosen to be given. In the real experiments

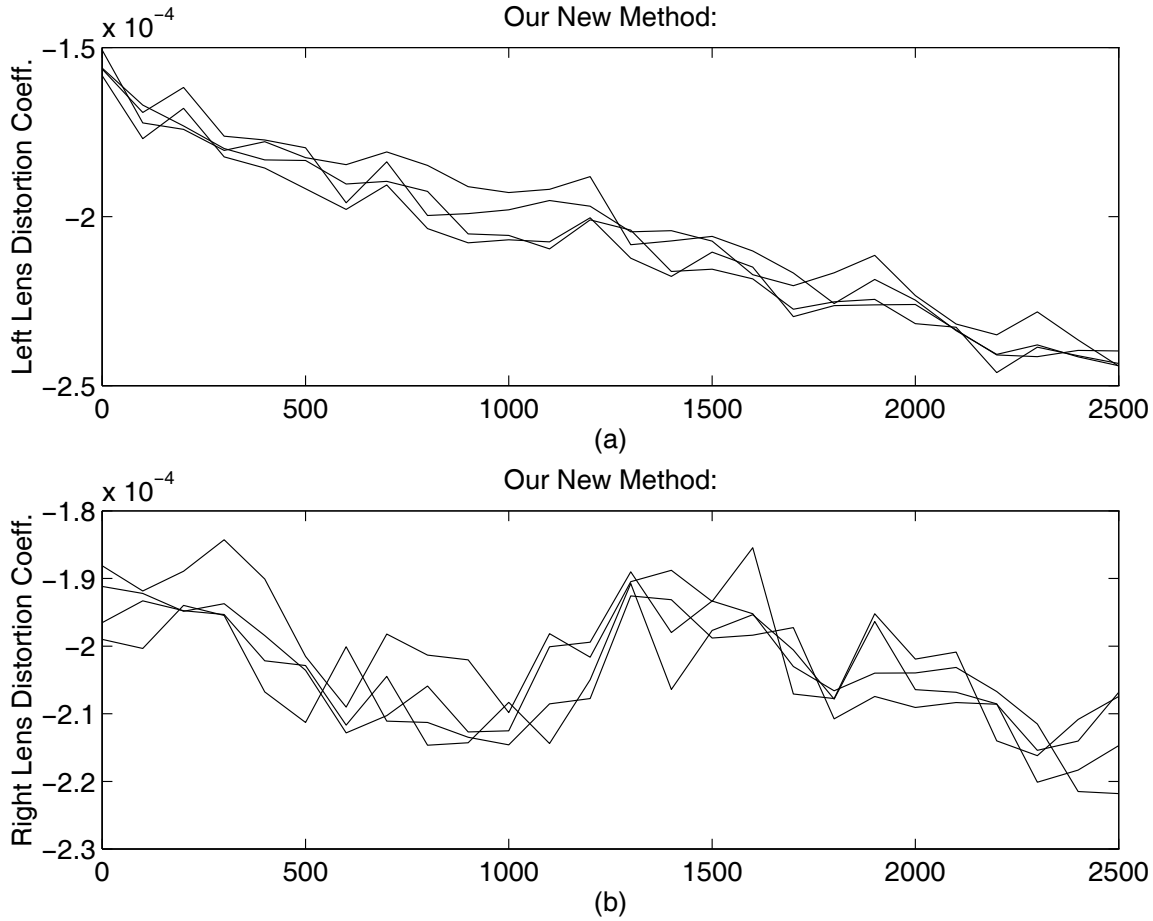

Fig. 13. Variations of the left and right radial lens distortion coefficients subject to the focus motor position estimated by using the proposed method. The horizontal axes are the motor position and the vertical axes are the amount of radial lens distortion coefficients (in millimeters^{-2}). Four curves represent for the results of four different real experiments. (a) κ_L versus the left focus motor position. (b) κ_R versus the right focus motor position.

the gaze angle and the three joint values were randomly generated, whereas the rest joint values were determined by solving the inverse kinematics problem. Notice that at the beginning of the calibration process, if the head parameters, $\Theta(t)$, is not accurate enough to generate joint values for tracking the calibration plate, then the calibration process should be interrupted and manual operations can be imposed to drive the calibration object back to the view-field of the stereo cameras such that new measurements can be obtained for updating the kinematic parameters. However, since our initial parameters are accurate enough, the calibration object was successfully tracked during the whole

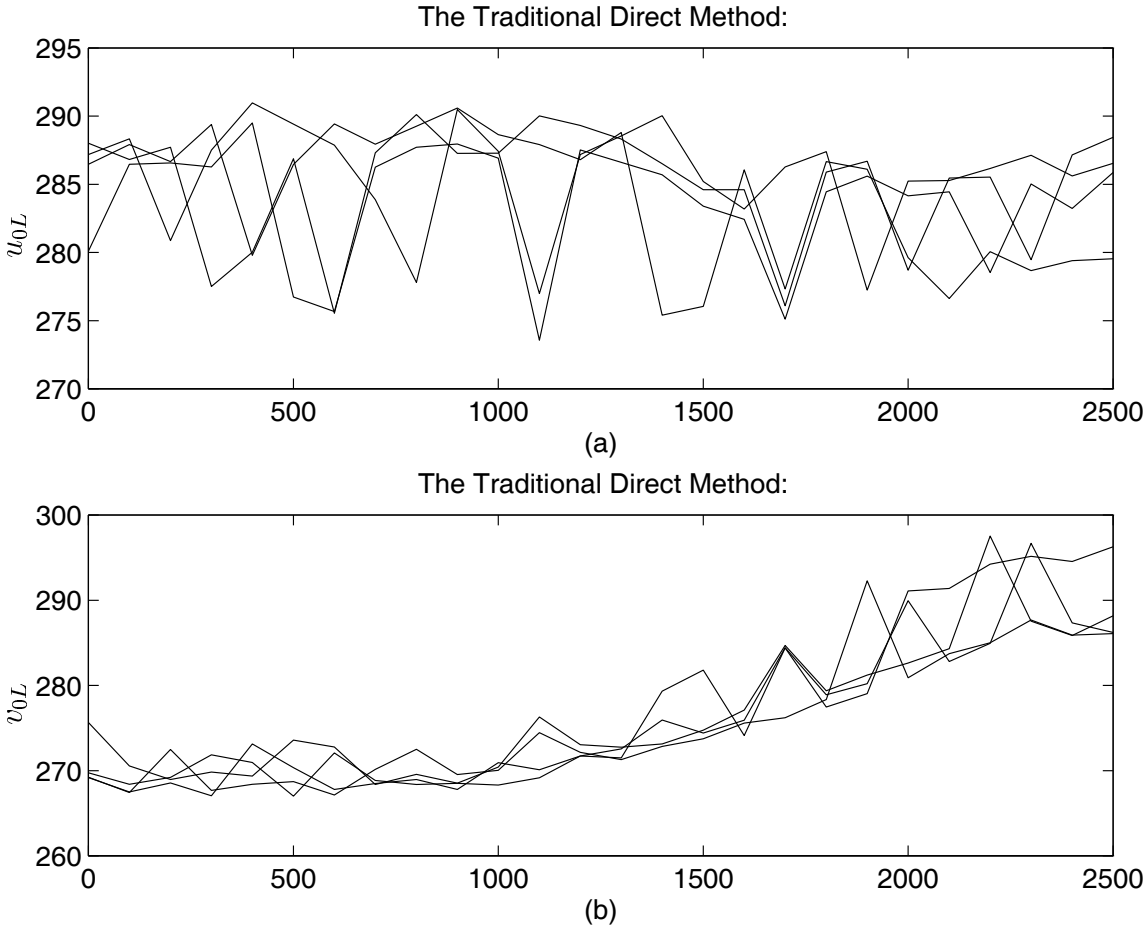

Fig. 14. Variations of the left image center subject to the focus motor position estimated by using a traditional direct method. The horizontal axes are the motor position and the vertical axes are the coordinates of the image center (in pixels). Four curves represent for the results of four different real experiments. (a) u_{0L} versus the left focus motor position. (b) v_{0L} versus the left focus motor position.

calibration process.

The process of the data acquisition starts with the 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 are determined automatically with the inverse kinematics module. This set of new joint values, $q(t)$, will then be sent to the motion controller of the IIS head. 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

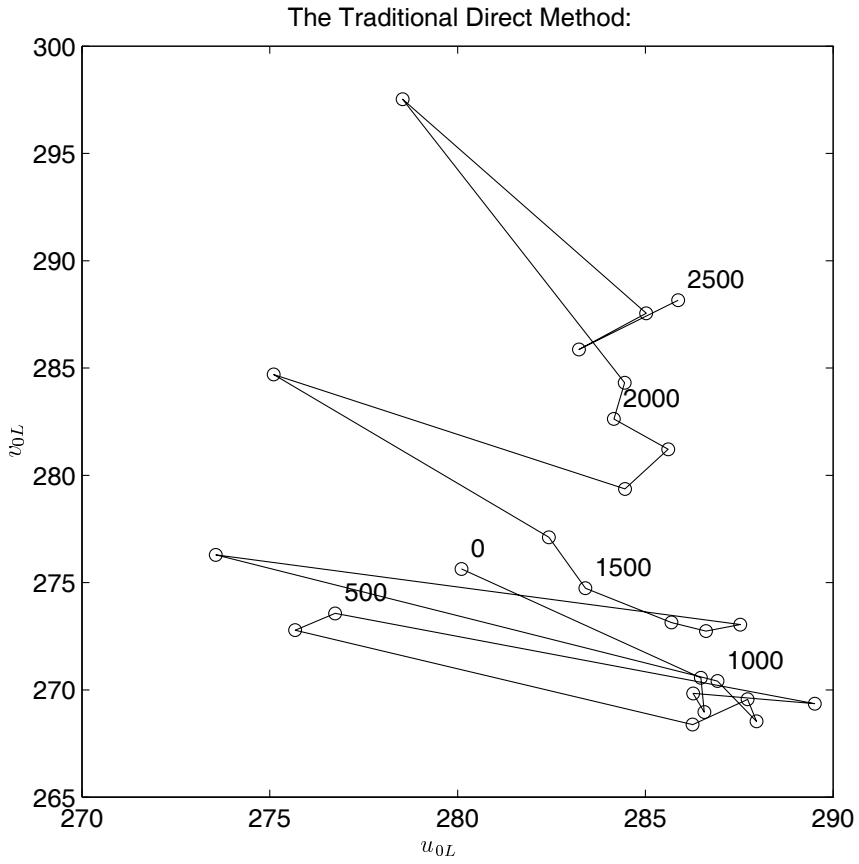

Fig. 15. Trajectory of the left image center subject to the focus motor position estimated by using a traditional direct method. Values of the focus motor positions are marked every five samples around the corresponding estimated image centers.

generate correct stereo correspondences. Figure 17 shows a typical picture of the detected nine stereo correspondences. The image location accuracy of the centroid detection for the nine calibration disks is about 0.1 pixel, and the total time for image acquisition and processing is approximately 0.5 second, 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 to the nonlinear recursive least-square estimator described in section IV-E. The distribution ranges of the generated joint values are listed in Table I.

D. Real Experimental Results

Figures 18–20 show some results of real experiments. Figure 18 shows the RMS (root mean square) 2D prediction error based on both the left and right 2D predictions (eighteen

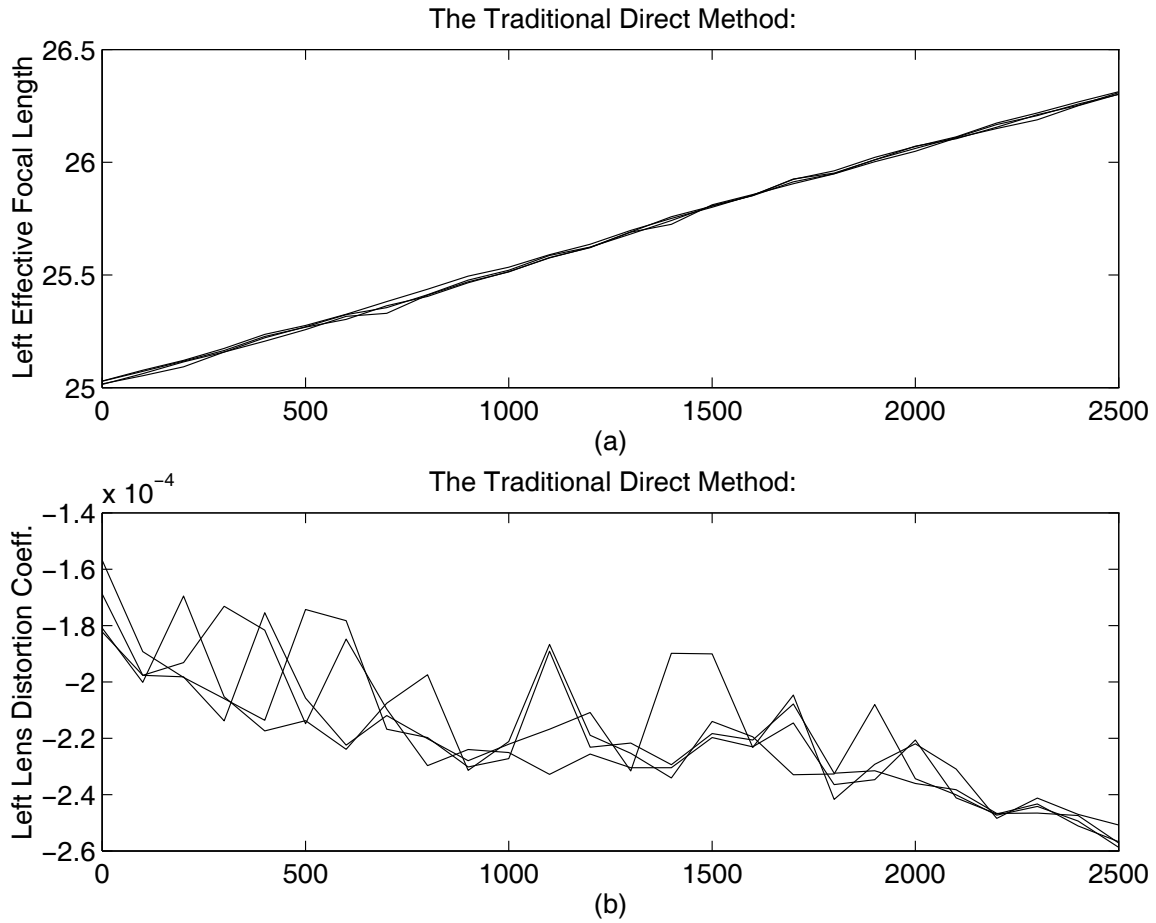

Fig. 16. Variations of the left effective focal length and the left lens distortion coefficient subject to the focus motor position estimated by using a traditional direct method. Four curves represent for the results of four different real experiments. The horizontal axes are the motor position and the vertical axes are (a) the effective focal length, f, (in millimeters), (b) the radial lens distortion coefficient, κ_L.

2D 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 39 kinematic parameters of the binocular head. Notice that since the effective focal length of the IIS head is about 25 millimeters and the pixel dimension is about ten micrometers, one pixel of RMS prediction error is approximately equal to 0.4 milli-radian of RMS orientation error, which represents an extremely high accuracy calibration result for an active binocular head.

Figure 19 shows another important performance index of an active binocular head, i.e,

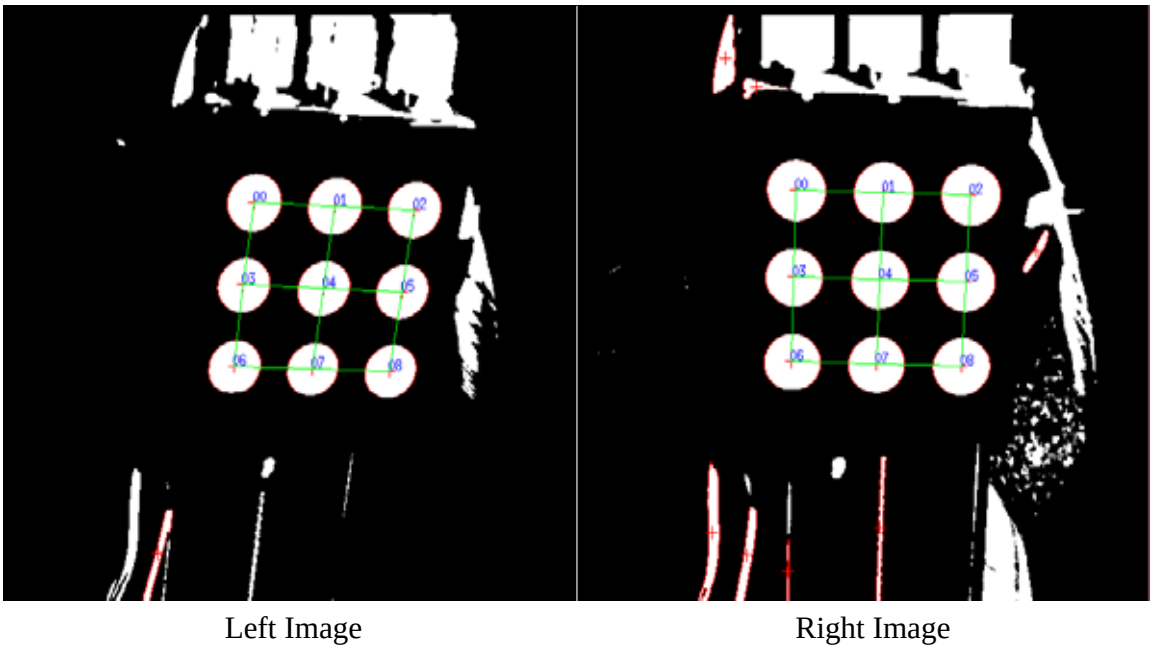

Fig. 17. A typical picture of the detected stereo correspondences.

the RMS epipolar error. As shown in this figure, the initial value of the extrinsic and intrinsic parameters would cause an epipolar error of about ten pixels, and after several hundreds of iterations, the epipolar error were reduced to several fractions of a pixel on the average. Notice that the 2D location error of the calibration points are approximately 0.1 pixel and the converged epipolar error is approximately 0.2 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 stero cameras at the calibration object were determined by using the inverse kinematics module described in section III-C. Here, the mission of the inverse kinematics module is to compute a set of feasible joint values such that the given interest point will be fixated at the center of the left and right images. In this experiment, the interest point was selected as the centroid of the nine disks on the calibration plate. Figure 20 shows distributions of the image locations of the nine calibration points on the left and right image planes, respectively, corresponding to the

TABLE I

THE 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 Vergence	0°-80°	0°-43°
5R	Right Vergence	0°-80°	0°-44°
6L	Left Effective Focal Length	25.01-26.34 mm	25.11-25.52 mm
6R	Right Effective Focal Length	25.25-26.50 mm	25.21-25.65 mm

1000 iterations in the experiment. It can be seen that because the center calibration point was chosen to be fixated, its footprints in the image were confined in a much smaller region than those of the other eight calibration point.

VI. CONCLUSION

Active stereo vision is attracting more and more research interest. More and more 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 few work relying on well-calibrated heads has been accomplished in the past. The main difficulties of the calibration problem for a binocular head are the high complexity, ultra-high accuracy requirement and 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 all moved simultaneously. In this paper, we demonstrate how an active stereo vision system can be easily calibrated with very high accuracy. Consequently, many 3D applications can directly make full use of an active binocular head. In addition to the proposal and demonstration of an four-stage

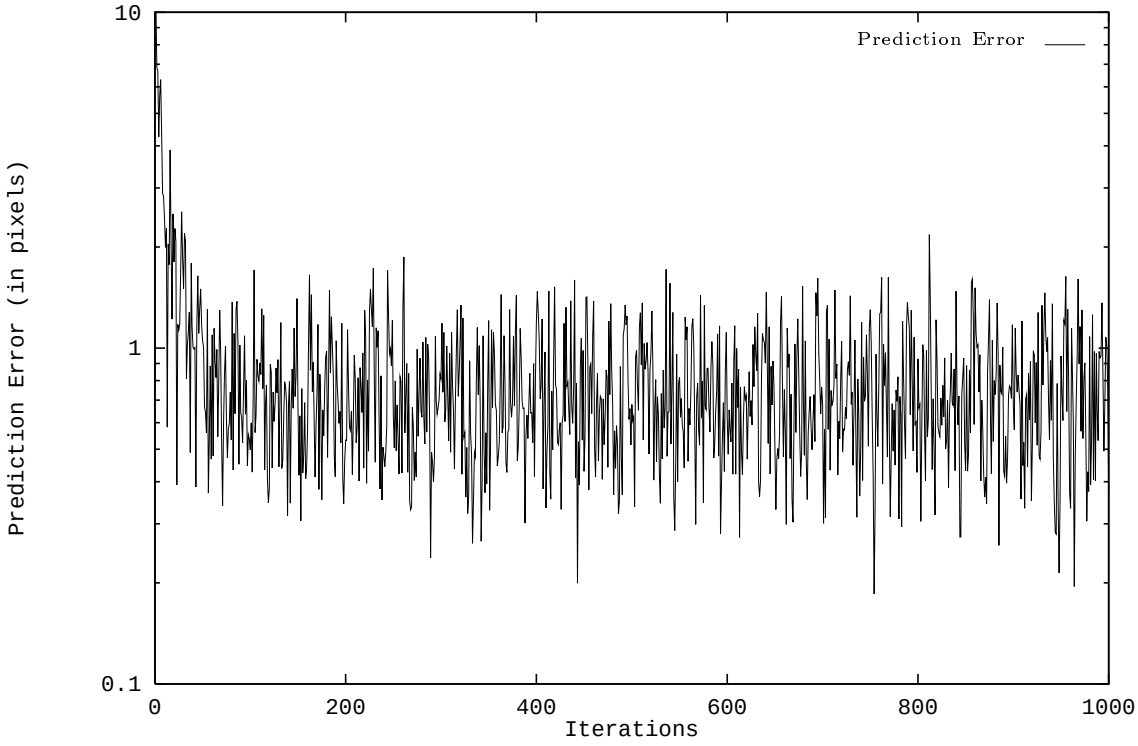

Fig. 18. The decreasing of the 2D prediction error by iterative refinement

calibration process, there are three major contributions in this paper:

1. We proposed a MFL camera model for cameras using motorized lens for adjusting 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 make the full use of the extra-degrees of freedom in the estimation process (i.e, the linear dependency of some camera parameters in small variations) by fixing the extrinsic parameters. This model is simple and especially suitable for cameras having low distortion lens, because the linear dependencies in small variations will become invalid when the amount of lens distortion is large (refer to [29]).
2. We proposed a new method for calibrating a MFL camera model to map the value of 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

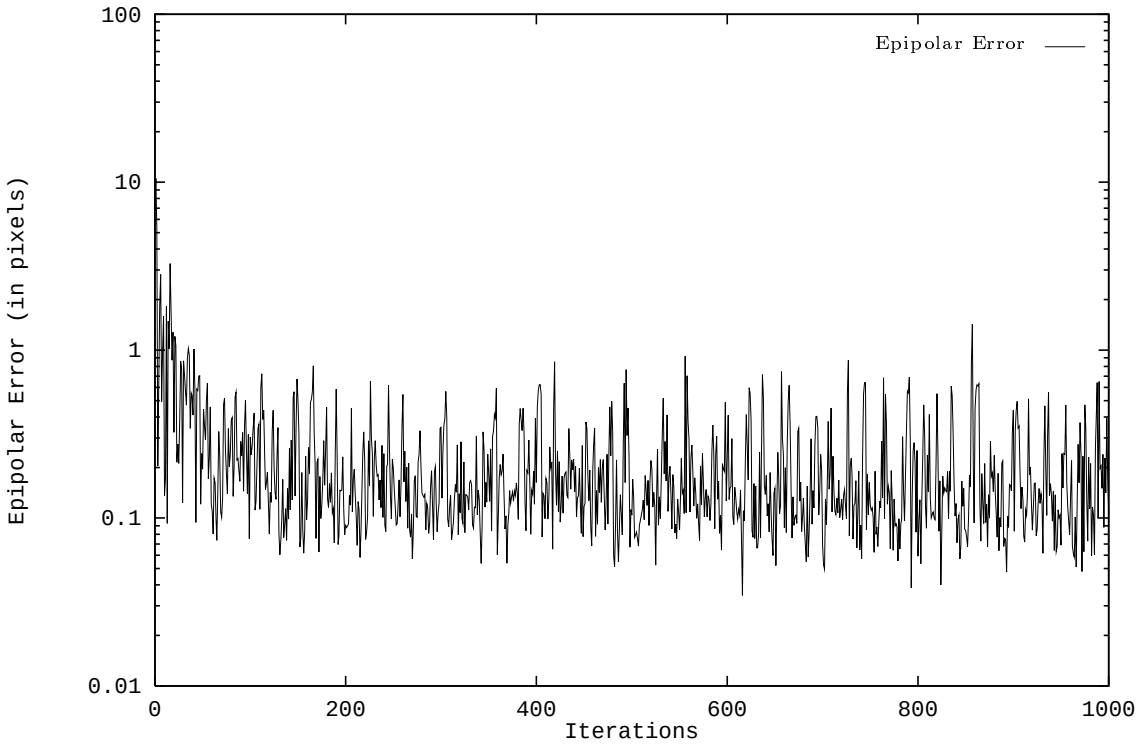

Fig. 19. The decreasing of the epipolar error by iterative refinement

nonlinear optimization method. The second reason is that the estimation of image center and effective focal length are made independent to the estimation of 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 in small variations.

3. We showed 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 2D image location and the epipolar error. Totally, 39 parameters are estimated simultaneously which includes three joint level parameters. An inverse kinematics module was 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.

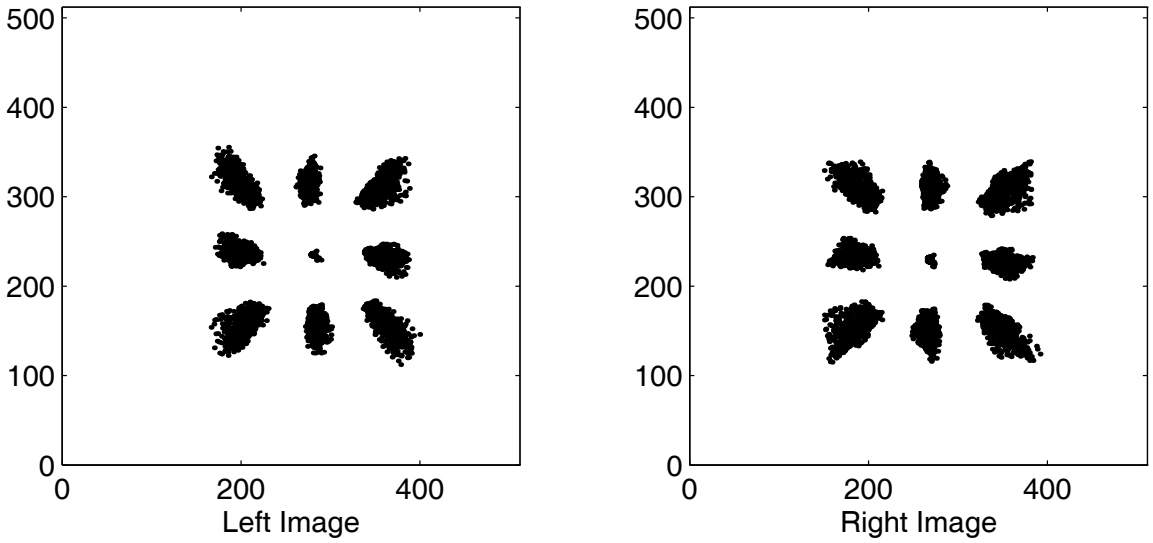

Fig. 20. Footprints of the nine calibration points in the left and right images

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 3D vision applications, such as 3D navigation, 3D object tracking, or even 3D reconstruction. Based on the highly accurate calibration, we have successfully done some experiments on panoramic stereo imaging, and are now working on reconstruction of 3D environment model using our active binocular system.

Acknowledgments

This work is supported in part by the National Science Council under grant NSC-85-2213-E-001-016.

Appendix

A. Estimation of Gear Backlash

In general, the amounts of gear backlash are different for different motor positions. For any motor position, the actual motor position (denoted as p_{act}), the nominal motor position (denoted as p_{nom}) and the backlash constant (denoted as g_{bl}) can be related as

follows:

$$p_{act} = \begin{cases} p_{nom}, & \text{if the last move is in forward direction,} \\ p_{nom} + g_{bl}, & \text{otherwise.} \end{cases} \tag{21}$$

The following procedures can be used to estimate the backlash quantity corresponding to the nominal motor position p_{nom}:

1. Mount a diode laser on the link corresponding to the joint to be calibrated.
2. Mark the projected laser spot on a sheet of paper posted on the wall which is far away from the joint under calibration.
3. Move the motor forward and backward for the same amount of motion and record the motor position which will be referred to as the hysteresis position denoted as p_{hys}.
4. Move the motor in backward direction until the laser spot align with the mark on the sheet of paper posted on the wall and record the current motor position denoted as p_{cur}.
5. The backlash constant can be obtained by computing the absolute value of the difference between the current motor position and the hysteresis position, i.e, $g_{bl} = |p_{hys} - p_{cur}|$.

B. Computation of the Epipolar Error

Suppose that $p_L = (u_{IL}, v_{IL})$ in the left image and $p_R = (u_{IR}, v_{IR})$ in the right image form a pair of stereo correspondence. Normally, such a pair of stereo correspondence should fall on their corresponding epipolar lines. However, due to the feature extraction error and the epipolar line error introduced by incorrect intrinsic and extrinsic camera parameters, the epipolar line may not pass through the corresponding image location. This kind of epipolar error can be defined and computed in the following way. By back projecting the two image feature point into 3D rays, we have

$$\begin{pmatrix} U_L \\ V_L \\ f_L \end{pmatrix} = \begin{pmatrix} \mathcal{D}_L \left(u_{IL} - u_{0L}\right) s_{uL} \\ \mathcal{D}_L \left(v_{IL} - v_{0L}\right) s_{vL} \\ f_L \end{pmatrix}, \tag{22}$$

and

$$\begin{pmatrix} U_R \\ V_R \\ f_R \end{pmatrix} = \begin{pmatrix} \mathcal{D}_R \left(u_{IR} - u_{0R}\right) s_{uR} \\ \mathcal{D}_R \left(v_{IR} - v_{0R}\right) s_{vR} \\ f_R \end{pmatrix}, \tag{23}$$

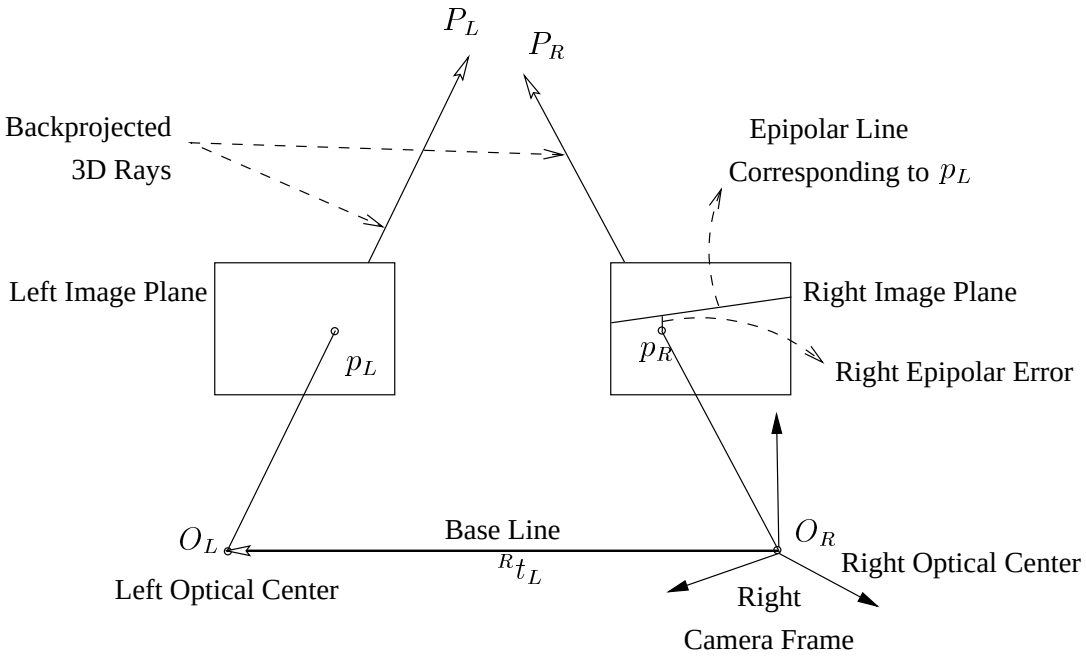

Fig. 21. The epipolar plane of a pair of stereo correspondences.

where

$$\mathcal{D}_L = 1 - \kappa_L \left[(u_{IL} - u_{0L})^2 \, s_{uL}^2 + (v_{IL} - v_{oL})^2 \, s_{vL}^2 \right]$$

and

$$\mathcal{D}_R = 1 - \kappa_R \left[(u_{IR} - u_{0R})^2 \, s_{uR}^2 + (v_{IR} - v_{oR})^2 \, s_{vR}^2 . \right]$$

For convenience, let P_L and P_R denote the back-projected 3D vectors, $[U_L \; V_L \; f_L]'$ and $[U_R \; V_R \; f_R]'$, respectively. Let the relative pose between the left and right camera frame be

$${}^R T_L = \begin{pmatrix} {}^R R_L & {}^R t_L \\ 0 & 1 \end{pmatrix} = \left({}^0 T_{6R} \, {}^{6R} T_{CR} \right)^{-1} {}^0 T_{6L} \, {}^{6L} T_{CL}, \tag{24}$$

where ${}^R R_L$ and ${}^R t_L$ are the rotation matrix and the translation vector, respectively. Notice that an epipolar plane can be defined by a feature point on the left or right image plane and both the origins of the reference frames corresponding to the stereo cameras. The normal vector of the epipolar plane corresponding to a left 2D feature point, p_L, can be determined from the vector of base line, i.e, ${}^R t_L$, and the back-projected 3D rays, i.e, P_L (see Figure 21). The normal vector of the epipolar plane defined by P_L, with respective

to the right camera reference frame, can be computed as follows:

$$n_R = {}^R t_L \times \left({}^R R_L P_L \right), \tag{25}$$

where '$\times$' denotes the cross-product operation.

The equation of the right image plane can be represented in the following:

$$z = f_R \tag{26}$$

By using equations (25) and (26), the epipolar line on the right image plane is

$$n_R \begin{pmatrix} x \\ y \\ f_R \end{pmatrix} = 0. \tag{27}$$

The right epipolar error, $h_{eR}\left(p_L(i), p_R(i), q(t); \Theta\right)$, is then defined as the distance between the image location (U_R, V_R) and the epipolar line described in equation (27).

Similarly, the left epipolar error, $h_{eL}\left(p_L(i), p_R(i), q(t); \Theta\right)$, can be computed by using the following epipolar line equation on the left image plane:

$$n_L \begin{pmatrix} x \\ y \\ f_L \end{pmatrix} = 0, \tag{28}$$

where

$$n_L = \left({}^R R_L^{-1} {}^R t_L \right) \times \left({}^R R_L^{-1} P_R \right). \tag{29}$$

Notice that since both of the left and right epipolar lines are close to horizontal lines for our lateral stereo camera configuration, we can simply divide the computed epipolar error by the vertical pixel size of the camera to roughly describe the epipolar error in the unit of pixels.

References

[1] J. Aloimonos and A. Badyopadhyay, "Active vision," in *the first International Conference on Computer Vision*, pp. 35–54, 1987.

[2] A. L. Abbott, *Dynamic Integration of Depth Cues for Surface Reconstruction from Stereo Images.* PhD thesis, University of Illinois as Urbana-Champaign, 1990.

[3] H. I. Christensen, "A low-cost robot camera head," *International Journal of Pattern Recognition and Artificial In telligence*, vol. 7, no. 1, pp. 69–87, 1993.

[4] N. J. Ferrier and J. J. Clark, "The harvard binocular head," *International Journal of Pattern Recognition and Artificial In telligence*, 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," *International Journal of Pattern Recognition and Artificial In telligence*, vol. 7, no. 1, pp. 51–68, 1993.

[6] K. Pahlavan and J. O. Eklundh, "Heads, eyes and head-eye systems," *International Journal of Pattern Recognition and Artificial In telligence*, vol. 7, no. 1, pp. 33–49, 1993.

[7] J. R. G. Pretlove and G. A. Parker, "The surrey attentive robot vision system," *International Journal of Pattern Recognition and Artificial In telligence*, vol. 7, no. 1, pp. 89–107, 1993.

[8] P. M. Sharkey, D. W. Murray, S. V. 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 3d machine vision metrology using off-the-shelf tv cameras and lenses," *IEEE Journal of Robotics and Automation*, 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 Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 10, pp. 965–980, 1992.

[11] G. Q. Wei and S. D. Ma, "Implicit and explicit camera calibration: Theory and experiments," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 5, pp. 469–480, 1994.

[12] U. R. Dhond and J. K. Aggarwal, "Structure from stereo — a review," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 19, no. 6, pp. 1489–1510, 1989.

[13] S. B. Marapane and M. M. Trivedi, "Multi-primitive hierarchical (mph) stereo analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 3, pp. 227–240, 1994.

[14] E. P. Krotkov, *Active Computer Vision by Cooperative Focus and Stereo*. Springer-Verlag, New York Inc., 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 in Computer Science, Computer Vision - ECCV'94*, vol. 800, pp. 565–576, Jan-Olof Eklundh (Ed.), 1994.

[16] F. Du and M. Brady, "Self-calibration of the intrinsic parameters of cameras for active vision systems," in *Computer Vision Pattern Recognition*, pp. 477–482, 1993.

[17] S. D. Ma, "A self-calibration technique for active vision systems," *IEEE Transactions on Robotics and Automation*, vol. 12, no. 1, pp. 114–120, 1996.

[18] Z. Zhang, Q. T. Luong, and O. Faugeras, "Motion of an uncalibrated stereo rig: Self-calibration and metric reconstruction," *IEEE Transactions on Robotics and Automation*, vol. 12, no. 1, pp. 103–113, 1996.

[19] M. Li, "Camera calibration of the kth head/eye system," Tech. Rep. CVAP147, Computational Vision and Active Perception Laboratory, Department of Numerical Analysis and Computing Science, Royal Institute of Technology (KTH), S-100 44, Stockman, Sweden, 1994.

[20] M. Li, D. Betsis, and J. M. Lavest, "Kinematic calibration of the kth head-eye system," Tech. Rep. CVAP171, Computational Vision and Active Perception Laboratory, Department of Numerical Analysis and Computing Science, Royal Institute of Technology (KTH), S-100 44, Stockman, Sweden, 1994.

[21] M. Li and J. M. Lavest, "Some aspects of zoom-lens camera calibration," Tech. Rep. CVAP172, Computational Vision and Active Perception Laboratory, Department of Numerical Analysis and Computing Science, Royal Institute of Technology (KTH), S-100 44, Stockman, Sweden, 1995.

[22] P. F. McLauchlan and D. W. Murray, "Active camera calibration for a head-eye platform using the variable state-dimension filter," *To appear in IEEE Transaction on Pattern Analysis and Machine Intelligence*.

[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," in *SPIE Intelligent Robots and Computer Vision XI*, vol. 1825, pp. 643–657, 1992.

[24] S. W. Shih, Y. P. Hung, and W. S. Lin, "Kinematic parameter identification of a binocular head using stereo measurements of single calibration point," in *IEEE International Conference on Robotics and Automation*, (Nagoya, Japan), pp. 1796–1801, 1995.

[25] G. S. Young, T. H. Hong, M. Herman, and J. C. S. Yang, "Kinematic calibration of an active camera system," in *IEEE International Conference on Computer Vision Pattern Recognition*, pp. 748–751, 1992.

[26] R. Y. Tsai and R. K. Lenz, "A new technique for fully autonomous and efficient 3d robotics hand/eye calibration," *IEEE Transactions on Robotics and Automation*, vol. 5, no. 3, pp. 345–358, 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 Transactions on Pattern Analysis and Machine Intelligence*, vol. 10, no. 5, pp. 713–720, 1988.

[28] R. K. Lenz and R. Y. Tsai, "Calibrating a cartesian robot with eye-on-hand configuration independent of eye-to-hand relationship," *IEEE Transactions on Robotics and Automation*, vol. 11, no. 9, pp. 916–928, 1989.

[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 (ftp://140.109.23.242/pub/TR-IIS-96-006), Institute of Information Science, Acdemia Sinica, Nankang, Taipei, Taiwan, 1996.

[30] K. A. Tarabanis, R. Y. Tsai, and D. S. Goodman, "Modeling of a computer-controlled zoom lens," in *IEEE International Conference on Robotics and Automation*, pp. 1545–1551, 1992.

[31] R. G. Willson, *Modeling and Calibration of Automated Zoom Lenses*. PhD thesis, Department of Electrical and Computer Engineering, Carnegie Mellon University, 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 Transactions on Robotics and Automation*, vol. 11, no. 5, pp. 777–780, 1995.

[33] S. W. Shih, Y. P. Hung, and W. S. Lin, "Error analysis on closed-form solutions for kinematic calibration," Tech. Rep. TR-IIS-96-00x (in preparation, ftp://140.109.23.242/pub/TR-IIS-96-00x), Institute of Information Science, Acdemia Sinica, Nankang, Taipei, Taiwan, 1996.

[34] C. C. Wang, "Extrinsic calibration of a vision sensor mounted on a robot," *IEEE Transactions on Robotics and Automation*, vol. 8, no. 2, pp. 161–175, 1992.

[35] H. Zhuang, K. Wang, and Z. S. Roth, "Simultaneous calibration of a robot and a hand-mounted camera," *IEEE Transactions on Robotics and Automation*, vol. 11, no. 5, pp. 649–660, 1995.

[36] P. Y. Tsai, "Calibration of a camera with motorized lens and its application to depth from focusing," Master's thesis, Department of Computer Science and Information Enginerring, National Taiwan University, Taipei, Taiwan, 1995.

[37] Z. S. Roth, B. W. Mooring, and B. Ravani, "An overview of robot calibration," *IEEE Journal of Robotics and Automation*, vol. RA–3, no. 5, pp. 377–385, 1987.

[38] S. Umeyama, "Least-squares estimation of transformation parameters between two point patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 4, pp. 376–380, 1991.

[39] J. M. Mendel, *Lessons in Digital Estimation Theory*. Englewood Cliffs, New Jersey 07632, U.S.A.: Prentice-Hall, Inc., 1987.