

# Robust Active Stereo Calibration

J. Neubert, and N. J. Ferrier\*

Mechanical Engineering, University of Wisconsin, Madison, WI, USA 53706

neubert@robios6.me.wisc.edu ferrier@mechatron.me.wisc.edu

## Abstract

*We present a calibration procedure to determine the kinematic parameters of an active stereo system in a robot-centric frame of reference. Our goal was to obtain a solution of sufficient accuracy that the kinematic model information can be used to estimate scene structure given the measured motion/position of the eye and target object's image location. We formulate our problem using canonical coordinates of the rotation group, which enables a particularly simple closed form solution. Additionally, this formulation and solution provides quantitative measures of the resulting solutions. Experiments verify that the solutions are accurate, 3D structure can be estimated from the kinematic models, and the algorithm can indicate when image errors are large enough to produce un-reliable results.*

**Keywords:** Active Vision, Stereo Vision, Calibration, 3D Reconstruction

## 1 Introduction

Active stereo camera systems, often called “head-eye” systems, typically consist of a pair of cameras whose motion can be controlled. Control over, and measurement of, the motion of the eyes can facilitate many vision algorithms and is a central tenet of many active vision solutions. For example, the ability of a system to fixate can yield useful information about the environment. In order to extract *metric* information from controlled eye movements, a precise kinematic model parameterizing the motion of the visual sensor can be used. Unlike classical camera calibration, the extrinsic parameters in a dynamic stereo camera system change continuously and new calibration procedures must be developed.

Many approaches have been proposed for active stereo calibration. For a system mounted on a mobile robot we sought a solution that: 1) determines the kinematic parameters so that they can be used in later computations, 2) uses a robot-centric frame of reference, 3) is

accurate and provides quantitative assessment of the accuracy, 4) performs well when restricted to small rotations of the sensor, and 5) exploits information that is known (e.g. the motion of the pan-tilt units, PTUs). Brooks *et. al.* [3] assumes that the pan and tilt angles are unknown and solves for them. Knight and Reid [8] present approaches based on projective geometry, however neither provides a solution as functions of the controlled movement (PTU angles). Tsai and Lenz [14] transform the problem to a linear form in terms of the sines and cosines of the rotations sought. Similarly, Shiu and Ahmad [12] re-formulated the vision hand-eye problem as  $AX = XB$  then solved the problem by transforming it into a linear problem. The re-formulation into the form  $AX = XB$  allows the use of solutions to similar problems in other robot calibration tasks. Li [9] used this formulation of the problem and developed a non-linear optimization technique to solve it. His approach uses the motion of the PTU from a home position to a set of discrete positions and performs a classical camera calibration at each position, and then an optimization step computes the camera transformation.

Our approach is similar to that of Li [9] in that we utilize a classical camera calibration at a set of discrete pan-tilt positions. However, unlike Li, we use the relative motion between positions (not the motion from home to each position) and thus errors in each pair of configurations are independent (not biased towards errors in the home position). Unlike our method, Li's approach does not afford a robot-centric solution.

Here we present our active stereo calibration method that uses the  $AX = XB$  formulation. Instead of using optimization we use a Lie theory approach for kinematic modeling [1]. One advantage of this approach is that the use of canonical coordinates enables the formulation of the solution as a linear least squares fit [11]. Martin and Park [11] derive a closed form solution to the  $AX = XB$  problem as a linear least squares fit and this result is applied here in determining the “hand-eye” transformation. Additionally, our formulation provides a *quantitative* measure of the

\*This research supported in part by NSF IRI-9703352.

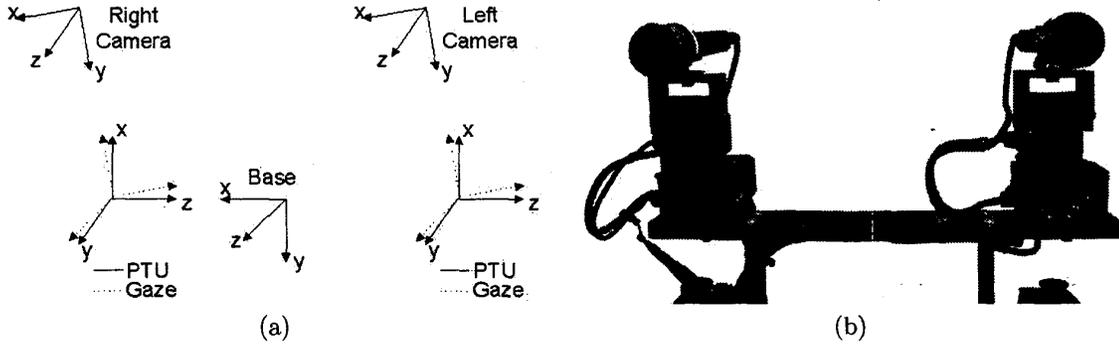


Figure 1: (a) The coordinate frames that are present on the dynamic stereo head. (b) The active stereo head.

calibration results, indicating whether further data is needed to ensure an accurate estimate of the kinematic model.

In the remainder of the paper we present our formulation of the active stereo calibration problem for a system with two pan-tilt units mounted on a mobile robot base. Some background is presented and the calibration procedure is described. We present results showing the accuracy of the method in the presence of noise, and the ability of the procedure to quantify possible inaccuracies (when the noise is too high). The kinematic model that results from this procedure enables us to estimate depth using the measured/controlled position of the cameras, along with image data, as demonstrated in section 5.

## 2 Head Geometry

Our physical system is similar to numerous others. Two pan-tilt are mounted on a mobile robot base. The PTUs have two degrees of freedom, both of them rotational. In order to describe the physical set-up, it is convenient to assign frames of reference. Here we present the mathematical notations and definitions for the transformations used.

In the following discussions, we will outline the coordinate systems used throughout the rest of the paper (see figure 1):

- (*b*)*ase* the robot-centric coordinate system located midway between the two pan-tilt units. We denote the base coordinate system with a “b” ( $x_b, y_b, z_b$ ).
- (*p*)*tu* fixed on the pan/tilt unit when in “home” position (both cameras parallel to each other facing “forward”). The pan/tilt angles are zero at this position. (i.e.  $T_{gp}(\phi = 0, \tau = 0)$ )

- (*g*)*aze* the coordinate system attached to the pan/tilt unit for the current gaze position (i.e. it changes with the pan/tilt angle). When the pan/tilt angles are both zero, the axes of the gaze and ptu frames coincide. (i.e.  $T_{gp}(\phi, \tau)$ )

- (*c*)*amera* the camera coordinate frame located at the optical center of the camera with positive  $z$  axis along the optical axis.

- (*w*)*orld* a frame of reference (often attached to a calibration grid or some physical landmark) - the robot moves relative to this frame.

The coordinate systems *ptu*, *gaze*, and *camera* are defined for both the right and the left camera.

We express a point,  ${}^bq$  with respect to the camera by modeling the transformation

$${}^c q = [T_{cb}] {}^b q \quad (1)$$

where  $[T_{cb}]$  is decomposed into 3 transformations

$$[T_{cg} T_{gp} T_{pb}]. \quad (2)$$

The transformation from system *base* to *camera* has been decomposed into three transformations.  $T_{pb}$  is the transformation from the robot base coordinates to the home position of the pan-tilt unit.  $T_{cg}$  is the transformation from the pan-tilt unit to the optical center of the camera. Assuming a fixed focus and rigidly attached pan-tilt units,  $T_{cg}$  and  $T_{pb}$  are assumed to be fixed (constant) transformations.  $T_{cg}$  and  $T_{pb}$  can be found using the **Robust Active Stereo Calibration (RASC)** method described in section 3.  $T_{gp}$  represents the pan/tilt action and is a function of those two angles. The transformation  $T_{gp}(\phi, \tau)$  describes the motion of the unit from its home position (0,0) to the current “gaze” position  $(\phi, \tau)$ . Our Directed Perception pan-tilt unit [4] rotates about a fixed origin, thus

we can model the PTU to gaze transformation as

$$T_{gp} = \begin{bmatrix} \cos \tau & -\sin \tau \cos \phi & -\sin \tau \sin \phi & 0 \\ \sin \tau & \cos \tau \cos \phi & \cos \tau \sin \phi & 0 \\ 0 & -\sin \phi & \cos \phi & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (3)$$

where  $\tau$  is the PTU tilt angle and  $\phi$  is the pan angle. The value  $\tau$  is a negative rotation (with respect to the chosen right hand coordinate system) but a positive value as commanded on the physical unit (see figure 2). With this notation, our goal is re-stated as:

Given the ability to control  $T_{gp}$ , find  $T_{cg}$  and  $T_{pb}$  for each of the “eyes”.

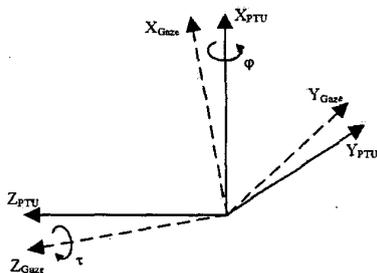


Figure 2: Pan-tilt transformation is a pure rotation for our hardware.

### 3 Calibration Procedure

First we present an outline of our calibration procedure, then the details of each step will be presented. The kinematic calibration of an active stereo head is broken down into the following steps:

1. PTU-camera calibration. For each “hand-eye” system (PTU/camera)
  - (a) Move each eye to a set of discrete positions. Record the PTU position and the image. For each position:
    - Perform camera calibration to determine intrinsic and extrinsic parameters at each position.
    - [Optional] Find the best intrinsic parameters (use an average of those found, do robust statistics to reject outliers) then re-do the extrinsic parameter estimation using the new (fixed) intrinsic parameters.
  - (b) Calculate the transformation from gaze to image,  $T_{gc}$  using equations (6) and (7).

#### 2. Base-PTU calibration.

- (a) Use physical measurements between the PTUs to determine the distance and angle between the PTUs.
- (b) Construct PTU to base transformations from the transformation between the PTUs.

### 3.1 Camera Calibration

The RASC method requires both intrinsic and extrinsic parameters. This requires some type of 3D-2D correspondence. Using a known 3D model or calibration grid: Determine correspondences between known model points and image projections of those points, solve for the projection matrix, and decompose the projection matrix. Procedures for this are laid out in detail in [5] and [13].

### 3.2 Gaze to Camera Calibration

We can measure a point with respect to the world coordinate frame (determined in the individual extrinsic calibration step). The relationship between the position of a point expressed with respect to the camera at position/frame  $\{i\}$  and at frame  $\{j\}$  depends on the relative motion of the PTU between the frames,  $(R_{gp}^i (R_{gp}^j)^T)$  and the other (fixed) transformations (equation (4)). Additionally one can use the extrinsic calibration estimated at each position  $(R_{cw}^i, t_{cw}^i)$  and  $(R_{cw}^j, t_{cw}^j)$  to determine the “change of coordinates” (equation (5)).

$${}^cP^i = [R_{cg} (R_{gp}^i (R_{gp}^j)^T) R_{cg}^T] {}^cP^j + [I - R_{cg} R_{gp}^i (R_{gp}^j)^T (R_{cg})^T] t_{cg} \quad (4)$$

$${}^cP^i = [R_{cw}^i (R_{cw}^j)^T] {}^cP^j + t_{cw}^i - [R_{cw}^i (R_{cw}^j)^T] t_{cw}^j \quad (5)$$

where  $(R_{cg}, t_{cg}) \in SE(3)$  is the transformation from the gaze to the camera (which we wish to determine). Equating equation (4) and (5) separates the rotation and translation components to produce:

$$[R_{cg} (R_{gp}^i (R_{gp}^j)^T) (R_{cg})^T] = [R_{cw}^i (R_{cw}^j)^T] \quad (6)$$

$$[I - R_{cg} R_{gp}^i (R_{gp}^j)^T (R_{cg})^T] t_{cg} = t_{cw}^i - [R_{cw}^i (R_{cw}^j)^T] t_{cw}^j \quad (7)$$

The first is easily rearranged to be of the form  $AX = XB$ , for  $X = R_{cg}$ . Referring to the solution outlined by Park and Martin [11], if  $A = XBXT^T$  then there exists

vectors  $\alpha$  and  $\beta$  and corresponding skew-symmetric matrices<sup>1</sup>,  $[\alpha]$  and  $[\beta]$  such that  $A = e^{[\alpha]}$  and  $B = e^{[\beta]}$  and  $A = XBXT^T$  can be re-written as

$$e^{[\alpha]} = Xe^{[\beta]}X^T = e^{X[\beta]X^T} = e^{[X\beta]}.$$

Thus  $AX = XB$  implies the linear relationship

$$\alpha = X\beta \quad (8)$$

Using canonical coordinates of  $SE(3)$  to parameterize the coordinate transformations, the equation  $AX = XB$  is transformed into a linear equation. Park and Martin [11] first made this observation for solving robot sensor calibration problems. The linear form of  $AX - XB$  allows the use of earlier results in least squares fitting on matrix subgroups [2] to derive a closed form solution to  $\alpha = X\beta$  for  $X \in SE(3)$ ,  $\alpha, \beta \in \mathfrak{R}^3$  (and hence to  $AX = XB$ , for  $A, B, X \in SO(3)$ ). The calibration procedure yields a set of estimates  $(A_k, B_k)$ ,  $k = 1, \dots, n$  (where  $A_k$  and  $B_k$  are elements of  $SO(3)$  of the form given in equation (6)). Define  $(\alpha_k, \beta_k)$  where  $\alpha_k = \log A_k$  and  $\beta_k = \log B_k$ , where

$$\log \Theta = \frac{\phi}{2 \sin \phi} (\Theta - \Theta^T) \quad (9)$$

for  $\Theta \in SO(3)$  such that  $\text{Tr}(\Theta) \neq 1$ . Then the solution  $X$  that minimizes

$$\eta = \sum_k d(XA_k - BX_k)$$

where  $d(\cdot, \cdot)$  is a suitable metric on  $SO(3)$  is given by

$$X = R_{cg} = (M^T M)^{-1/2} M^T$$

where

$$M = \sum_i^n \beta_i \alpha_i^T$$

(see [2] for details) and we take the symmetric, positive definite square root of  $(M^T M)^{-1}$  (see [7]). The translation  $t_{cg}$  that satisfies equation (7)

$$t_{cg} = (C^T C)^{-1/2} C^T d$$

where

$$C = \begin{bmatrix} I - A_1 \\ I - A_2 \\ \vdots \\ I - A_n \end{bmatrix} \quad \text{and} \quad d = \begin{bmatrix} b_{A_1} - Xb_{B_1} \\ b_{A_2} - Xb_{B_2} \\ \vdots \\ b_{A_n} - Xb_{B_n} \end{bmatrix}.$$

<sup>1</sup>Formally,  $\alpha$  is the element in the Lie algebra that maps, via the exponential map, to the element  $A$  of the Lie group, details are given in [6].

The solution for  $t_{cg}$  is unique for a given  $X$ . The choice of  $X$  is unique if  $M^T M$  is nonsingular and has no repeated eigenvalues. This clearly shows the advantage of using canonical coordinates because our nonlinear problem is transformed to a linear one for which a closed form least squares solutions [2] can be applied. We use this to solve equation (8) and obtain the rotation component of the gaze to camera transformation. Our rotation component is computed explicitly without the use of optimization/minimization techniques. The quality of the solution can be verified by calculating the noise in the rotation solution. This is a good indication of the overall solution quality. The amount of noise in the rotation solution can be obtained by averaging the Frobenious norm of

$$[R_{cg} (R_{gp}^i (R_{gp}^j)^T) (R_{cg})^T] - [R_{cw}^i (R_{cw}^j)^T] \quad (10)$$

for all  $i$  and  $j$  where  $i \neq j$ . Our experience indicated that if Frobenious norm exceeds unity then the solution is unusable.

### 3.3 Robot base to PTU calibration

The relationship between the PTUs can be found explicitly using loop closure. First we write the equation of the loop,

$$[T_{pg}^L \ T_{gc}^L \ T_{cw}^L]^{-1} T_{RL} [T_{pg}^R \ T_{gc}^R \ T_{cw}^R] = I \quad (11)$$

Then solve for  $T_{RL}$  to obtain

$$T_{RL} = [T_{pg}^L \ T_{gc}^L \ T_{cw}^L] [T_{pg}^R \ T_{gc}^R \ T_{cw}^R]^{-1} \quad (12)$$

The values of  $T_{pg}^L$ ,  $T_{pg}^R$ ,  $T_{gc}^L$ , and  $T_{gc}^R$  are known through parameterization or calibration and  $T_{cw}$  can be found using the methods described in [5, 14]. These values are substituted into equation (12) and  $T_{RL}$  is calculated. This procedure provided reasonable results (rotation error of 2 - 3 degrees and baseline distance found within about a centimeter).

A more practical and accurate way to determine the transformation between the PTUs is to physically measure the angles and translation (because the PTUs were placed using precision machined mounts we have a good estimate of the transformation and measurement confirms the estimate). This method eliminates the noise associated with  $T_{gc}$  and  $T_{cw}$  that can lead to large errors in  $T_{RL}$ . When the transformation was constructed using physical measurements the angular error was less one degree and the error in the baseline was on the order of 2-3 mm.

There is an ambiguity in the translation placement of the base frame "symmetrically" with respect to each

PTU: an entire plane of solutions exists. The cyclopean frame assumes translations equal and opposite in  $X$  (for our particular choice of axes). Choosing this form above allows for a unique solution for  $b$ .

Although there is flexibility in its exact placement, note that the base frame of reference is *on the robot*. Hence employing this kinematic model, say using triangulation to produce an estimate of scene depth, will produce a measurement relative to the current pose of the robot. This proves very useful for navigation/localization. In contrast to others who report the need for localization [9], we can use image information to solve the localization problem.

## 4 Evaluation

The procedure outlined in section 3.2 was implemented along with the method outlined by Li [9]. These methods were evaluated using simulated data to determine the effect of various levels of noise.

The simulated data was created using a set of known transformations. This set of transformations was constructed to model the active vision system described in section 2. Both methods require as input the pan/tilt angles and  $T_{cw}$ . Because the pan/tilt angles are known within a few seconds of a degree, error was found to arise predominantly from the calibration procedure (image noise and errors in corner detection), and hence in our simulation noise was added to the  $T_{cw}$  transformation. The following formula was used to add noise to the true  $T_{cw}$ :

$$R_{cw}^{error} = R_{cw}^{true} \begin{bmatrix} 1 & -\omega_3 & \omega_2 \\ \omega_3 & 1 & -\omega_1 \\ -\omega_2 & \omega_1 & 1 \end{bmatrix} \quad (13)$$

$$t_{cw}^{error} = t_{cw}^{true} + [t_1 \quad t_2 \quad t_3]^T \quad (14)$$

where the noise in the rotation is described by the parameters  $\omega_1$ ,  $\omega_2$ , and  $\omega_3$  and  $t_1$ ,  $t_2$ , and  $t_3$  characterizes the error in the translation component. The noise parameters were Gaussian with mean zero. Equation (14) was used to create data sets with eight different levels of noise. The data sets were defined by the radius of the Gaussian noise in the angle and translation parameters. In order to get a good statistical representation of the performance of both methods 30 samples were created for each of the noise levels: (0.005 rad, 1 mm), (0.010 rad, 2 mm), (0.015 rad, 3 mm), (0.020 rad, 4 mm), (0.025 rad, 5 mm), (0.030 rad, 6 mm), (0.035 rad, 7 mm), and (0.040 rad, 8 mm).

These eight data sets were used with each method and the resulting error in the transformation is plotted in

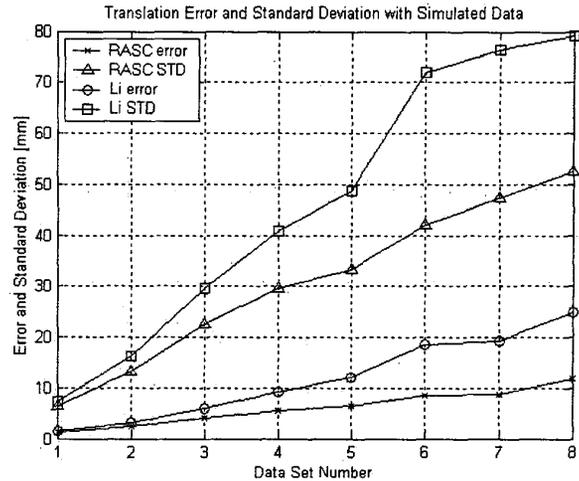


Figure 3: Error in the translation vector found with varying amounts of input data error. Noise in the input increases with set number. (See below for level of noise)

figures 3 and 4. The error in the rotation matrix, or rotation error, is defined as  $\|R_{calc} - R_{true}\|_F$ , where  $\|\cdot\|_F$  is the Frobenious norm of the matrix. The translation error is simply the  $L_2$  norm of the difference between  $t_{calc}$  and  $t_{true}$ . The standard deviation was defined in a similar fashion with the standard deviation of the rotation matrix defined as the Frobenious norm of the deviations of each term. The standard deviation in the translation was the  $L_2$  norm of the deviations in each of the terms in the translation. Referring to the results (figures 3 and 4) of each method some interesting observations are made. The RASC method produced results with lower error and standard deviation in both the rotation and translation components on average, but the difference in the rotation error was not significant. The translation component of the  $T_{gc}$  found with the RASC method contains significantly less error than that obtained with Li's method.

Another one of the major advantages of the method described by this paper is its robustness to noise in data obtained from any one of the camera positions. The method outlined by Li [9] is dependent on the PTU home position, thus if data from this particular position contains a large amount of error the results will be *significantly* effected. Using only a moderate amount of noise, (0.030 rad, 6 mm), in the image at this home position, the method described by Li resulted in an order of magnitude more error than those

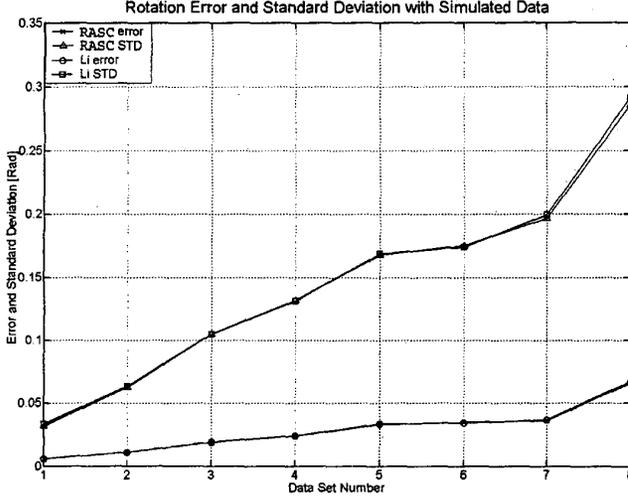


Figure 4: Error in the rotation matrix found with varying amount of noise in the the input. The noise in the data set increases with set number.

obtained with the RASC method (and hence results are not plotted for this experiment).

## 5 Experimental Results

The stereo vision system seen in figure 1 was calibrated using the RASC method described in section 3. The method outlined by [13] was used to find the intrinsic and extrinsic parameters for each camera at nine positions(pan, tilt)[deg]:  $(\pm 8, \pm 8)$ ,  $(\pm 8, 0)$ ,  $(0, \pm 8)$ , and  $(0, 0)$ . Then the extrinsic parameters were used to find the transformation from the gaze to camera frames with the method described in section 3.2. The transformation between the PTUs was constructed via physical measurement. The robot-centric coordinate system was arbitrary placed in the middle of the PTUs.

Given the kinematic parameters, and assuming that we have identified a point in each eye corresponding to an object of interest. The location of this point, relative to the base frame, is computed using a direct kinematic model approach.

The  $uv$  coordinates of the left and right image points can be expressed as projections of points in the *base* coordinate system. Applying equation (1) to both cameras results in:

$$\begin{bmatrix} u \\ v \end{bmatrix} \equiv \begin{bmatrix} \tilde{u} \\ \tilde{v} \\ \lambda \end{bmatrix} = \underbrace{{}^c\mathcal{P} T_{cg} T_{gw} T_{wb}}_A \begin{bmatrix} x_b \\ y_b \\ z_b \\ 1 \end{bmatrix} \quad (15)$$

	x	y	z
error	1.02	0.57	18.3
std	8.8	12.8	53.7

Table 1: Error and standard deviation of the reconstructed points (in mm). Data is based on the reconstruction of 2772 points.

The projective equivalence relationship yields

$$u = \frac{\tilde{u}}{\lambda}, \quad v = \frac{\tilde{v}}{\lambda} \quad (16)$$

and we solve for the components in equation (15):

$$u = \frac{a_{11}x_b + a_{12}y_b + a_{13}z_b + a_{14}}{a_{31}x_b + a_{32}y_b + a_{33}z_b + a_{14}} \quad (17)$$

$$v = \frac{a_{21}x_b + a_{22}y_b + a_{23}z_b + a_{24}}{a_{31}x_b + a_{32}y_b + a_{33}z_b + a_{14}} \quad (18)$$

where  $a_{ij}$  are the elements of the matrix  $A$  in equation (15). Combining the image projective relationship for each camera (and using superscripts,  $L$  and  $R$  to represent the left and right cameras) results in the linear system of equations

$$\underbrace{\begin{bmatrix} u_R a_{31}^R - a_{11}^R & u_R a_{32}^R - a_{12}^R & u_R a_{33}^R - a_{13}^R \\ v_R a_{31}^R - a_{21}^R & v_R a_{32}^R - a_{22}^R & v_R a_{33}^R - a_{23}^R \\ u_L a_{31}^L - a_{11}^L & u_L a_{32}^L - a_{12}^L & u_L a_{33}^L - a_{13}^L \\ v_L a_{31}^L - a_{21}^L & v_L a_{32}^L - a_{22}^L & v_L a_{33}^L - a_{23}^L \end{bmatrix}}_B \underbrace{\begin{bmatrix} x_b \\ y_b \\ z_b \end{bmatrix}}_{\tilde{x}} = \underbrace{\begin{bmatrix} a_{14}^R - u_R a_{34}^R \\ a_{24}^R - v_R a_{34}^R \\ a_{14}^L - u_L a_{34}^L \\ a_{24}^L - v_L a_{34}^L \end{bmatrix}}_{\tilde{q}} \quad (19)$$

which can be solved for  $x_b$ ,  $y_b$  and  $z_b$  using least squares (or total least squares). These four equations relate the four  $uv$  values ( $u_R$ ,  $v_R$ ,  $u_L$ ,  $v_L$ ) to the base coordinates  $x_b$ ,  $y_b$  and  $z_b$ .

The calibrated model obtained using the method outlined in this paper was used to map the pan/tilt angles and image position of various points to there corresponding 3D base coordinates. The calculated 3D base coordinates were then compared to the measured values to determine the performance of the calibrated head. The target locations observed by the stereo head occurred in 1 by 1 meter planes at three different depths with respect to the robot base, 1.622m, 1.771m, and 1.8615m. Data was collected using the method outlined in Neubert *et. al.* [10].

The model performed well on the data gathered, especially in the x and y directions (see table 1). The average error in the x and y directions was on the order of 1mm, but produced a standard deviation of about 1cm. As expected, the results contained significantly higher average error and standard deviation in the z direction. Depth estimates from the closer two planes produced about 6mm of error on average with half the standard deviation of the data at the larger (1.8615m) depth. The increase of error as depth increases was expected due to the nature of triangulation. The accuracy of the measurement, combined with the simplicity of the data acquisition and depth computation, will enable solutions (based on multiple measurements) to be developed to compensate for the level of precision obtained. This is consistent with many active vision approaches.

## 6 Discussion

A method for kinematic calibration of a dynamic stereo head has been presented. Unlike previous approaches, the solution is simple and explicit (i.e. truly in closed form without requiring analysis of "cases", nor having to perform any optimization or non-linear analysis). This simple approach is made possible through the use of canonical coordinates for Lie Groups describing the motion of the device. The method uses information "equally" (no bias from errors in any "special" zero or home position), and the derivation automatically provides a qualification of the "quality" of fit - thus highly uncertain results can be rejected automatically.

We have presented simulation results to verify RASC's benefits with respect to a previous method. The simulation also demonstrated the robustness of the RASC method to noise. We verified the accuracy of the model by calibrating an active stereo vision system, then using the model we mapped the pan/tilt angles and image coordinates to 3D coordinates in the robot-centric frame. The model produced good estimates of the measured coordinates. The demonstrated application of the calibrated kinematic stereo head model for depth computation shows that this method is simple and accurate, and thus can be readily incorporated into active vision solutions.

## References

- [1] Roger Brockett. Some mathematical aspects of robotics. In *Robotics*, volume 41 of *Proceedings of Symposia in Applied Mathematics*, pages 1-19. American Mathematical Society, 1990.
- [2] R.W. Brockett. Least squares matching problems. *Linear Algebra and its Applications*, 124:761-777, 1989.
- [3] M.J. Brooks, L. de Agapito, D.Q. Hyunh, and L. Baumela. Towards robust metric reconstruction via a dynamic uncalibrated stereo head. *Journal of Image and Vision Computing*, 16:989-1002, 1998.
- [4] Directed Perception. Directed perception ptu model 46-17.5. URL://www.dperception.com.
- [5] Olivier Faugeras. *Three-Dimensional Computer Vision - A Geometric Viewpoint*. MIT Press, 1993.
- [6] N.J. Ferrier and J. Neubert. A closed form solution for active stereo kinematic calibration, 2002. (in preparation) available at: <http://robios.me.wisc.edu/papers>.
- [7] G.H. Golub and C.F. Van Loan. *Matrix Computations*. Johns Hopkins Press, 1983.
- [8] J. Knight and I. Reid. Active visual alignment of a mobile stereo camera platform. In *IEEE International Conference on Robotics and Automation*, pages 3203-3208, San Francisco, CA, April 2000. IEEE.
- [9] Mengxiang Li. Kinematic calibration of an active head-eye system. *IEEE Transactions on Robotics and Automation*, 14(1):153-157, February 1998.
- [10] J. Neubert, A. Hammond, N. Guse, Y. Do, Y. Hu, and N. Ferrier. Automatic training of a neural net for active stereo 3d reconstruction. In *IEEE Int'l Conf. on Robotics and Automation*, pages 2140-2146, 2001.
- [11] Frank C. Park and Bryan J. Martin. Robot sensor calibration: Solving  $AX = XB$  on the euclidean group. *IEEE Transactions on Robotics and Automation*, 10(5):717-721, October 1994.
- [12] Yiu Chiung Shiu and Shaheen Ahmad. Calibration of wrist-mounted robotic sensors by solving homogenous transform equations of the form  $AX = XB$ . *IEEE Transactions on Robotics and Automation*, 5(1):16-29, February 1989.
- [13] Roger 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*, RA-3(4):323-344, August 1987.
- [14] Roger Y. Tsai and Reimar K. Lenz. A new technique for fully autonomous and efficient 3d robotics hand/eye calibration. *IEEE Transactions on Robotics and Automation*, 5(3):345-358, June 1989.