Active Calibration of Cameras: Theory and Implementation

Anup Basu

Abstract— **The problem of calibrating a camera has been widely addressed in the past. Almost all techniques described in the literature use a known calibrating pattern and a static camera. We introduce a novel technique, based on an active camera, which does not need any predefined patterns. All that is required is a scene with some strong and stable edges. Two alternative algorithms are presented and analysed—the second method is shown to be more robust to noise and useful in practical situations. Using our methods, an active camera can automatically calibrate itself. Experimental results are shown, demonstrating the validity of the algorithms.**

I. Introduction

CAMERA calibration involves relating the optical features of a lens to the sensing device. The parameters of interest are the image center and the focal length along the x and y axes, all expressed in terms of image pixels. The focal lengths are usually different along the vertical and horizontal axes because the scaling factor is not necessarily the same in the two directions. Calibration is important for any process where a 2-D image needs to be related to the 3-D world. Examples of such applications include pose estimation [18], 3-D motion estimation [6] and automated assembly.

Camera calibration techniques can be broadly classified according to two criteria: non-linear vs. linear systems, and algorithms which consider lens distortions vs. those which do not. Even though linear techniques [14], [22], [28] are simpler to implement, most cannot model camera distortions. Non-linear methods are able to consider complicated imaging models with many associated parameters [13], [24], [25], [29]. However, they require computationally expensive search procedures, and a reasonably good initial guess for convergence of the solutions. Techniques which model lens distortions often need to make simplifying assumptions. For example, only radial distortion is considered in [27]. Such restrictions in modeling often limit the domain of application of many algorithms.

One major drawback of existing algorithms is that cameras are calibrated using predefined patterns [27], [17], [8], [1], [12], [23] by relating their image projections to the camera parameters. Even recent algorithms [15], [3], [9] suffer from this limitation. In order to avoid solving complicated equations, recently some researchers have started studying classes of problems that can be solved without camera calibration [20], [11]. Eyes in humans and other animals do not need any artificial assistance for calibration. How can we explain the difference? In this work we try to provide one plausible answer to the above question. Specifically, we show that eye movements simplify the calibration problem: an active camera capable of panning and tilting can automatically calibrate itself. Active vision systems and algorithms [2], [21] have received a great deal of attention in the recent past. Active machines allow us to keep track of objects of interest, without having a lens with a wide field of view. They also facilitate region-of-interest processing. Furthermore, vergence movements have been shown to simplify problems in stereo vision.

When a camera is only rotating (not translating), the position of image contours after rotation can be obtained as a transformation of the initial contours [16]. This result assumes that the rotation and the camera parameters are precisely known. We relate the lens parameters to the image boundaries before and after a camera rotation. For small pan and tilt displacements, simple solutions for various unknowns can be obtained. Initial estimates can be iteratively improved (by further camera movements) until desired accuracy is obtained.

Unlike several other techniques [12], [8], [23], we do not need a starting estimate of both focal length and image center. All we assume is an estimate of the center that is not too far from the true value. Reasonable results were obtained using synthetic data, even when the initial guess was up to 30 pixels away from the actual parameters. The active calibration procedure does not need any prior information about the focal length. In addition, the algorithm does not need to match points or features (such as corners or edge elements) between images. Only reasonably accurate localization of contours is assumed.

In this paper two methods are presented and analysed. The first technique [4] uses perspective distortions to measure calibration parameters. Since perpective changes are usually small, this algorithm does not produce reliable estimates for real data or synthetic data with low levels of noise. In order to circumvent this problem, a second strategy [5] which does not use perpective distortions, is introduced. It is shown that the alternative method gives reliable estimates for synthetic data with high levels of noise and real scenes.

The organization of this paper is as follows: Section II gives a survey of previous literature. Some mathematical derivations relevant to our method are presented in Section III. Section IV outlines an active camera calibration procedure. Experimental results demonstrating the validity of our algorithm are presented in Section V.

Manuscript received November 20, 1992; revised September 23, 1993 and March 9, 1994. This work was supported in part by the Canadian Natural Sciences and Engineering Research Council under an Operating and a Collaborative Research and Development grant, and by TR Labs.

The author is with the Telecommunications Research Laboratories, Edmonton, Canada T5K 2P7, and the Department of Computing Science, University of Alberta, Edmonton, Alberta, Canada T6G 2H1.

IEEE Log Number 9405713.

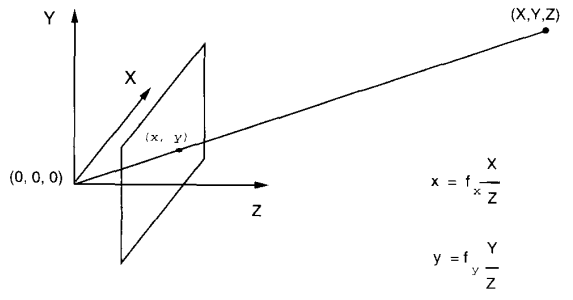


Fig. 1. Perspective projection.

II. Theoretical Derivations

We consider a perspective projection model (Fig. 1). The following notation is used:

- (x, y) Image coordinates.
- f_x Focal length in pixels in the x-direction.
- f_y Focal length in pixels in the y-direction.
- (δ_x, δ_y) Error in estimated lens center.
- $R = (r_{ij})_{3\times 3}$ Rotation matrix.
- (X, Y, Z) Point in the 3D scene.
- (x_p, y_p) Image coordinates after pan motion.
- (x_t, y_t) Image coordinates after tilt motion.
- θ_t Tilt angle.
- θ_p Pan angle.
- $T\begin{pmatrix}\Delta_X\\ \Delta_Y\\ \Delta_Z\end{pmatrix}$ Translation part of camera motion.

Lemma 1: Given a static scene, if the camera rotates by R and translates by T, the new image contours are given by:

$$x_n = \frac{r_{11}x + r_{21}\frac{f_x}{f_y}y + r_{31}f_x - f_x\frac{\Delta X}{Z}}{r_{13}\frac{x}{f_x} + r_{23}\frac{y}{f_y} + r_{33} - \frac{\Delta Z}{Z}}$$

$$y_n = \frac{r_{12}\frac{f_x}{f_y}x + r_{22}y + r_{32}f_y - f_y\frac{\Delta Y}{Z}}{r_{13}\frac{x}{f_x} + r_{23}\frac{y}{f_y} + r_{33} - \frac{\Delta Z}{Z}}$$

Proof: Follows from the fact that

$$\begin{pmatrix}X_n\\ Y_n\\ Z_n\end{pmatrix} = R^T\begin{pmatrix}X\\ Y\\ Z\end{pmatrix} - T$$

and the perspective projection equation. Note that camera rotation (translation) is equivalent to the inverse rotation (translation) of points in the scene. A similar result is proven in [16].

In this paper we consider f_x, f_y instead of a single focal length and the aspect ratio of the camera. Note that these two representations are equivalent.

Proposition 1: Suppose that depth (Z) is large compared to ΔX, ΔY, or ΔZ, and the camera moves by a small tilt angle θ_t. Then,

$$x_t \cong x\left(1 + \theta_t\frac{y}{f_y}\right)$$

$$y_t \cong (y + \theta_t f_y)\left(1 + \theta_t\frac{y}{f_y}\right)$$

Proof: For a small tilt angle θ_t the matrix R is approximately given by:

$$\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & -\theta_t\\ 0 & \theta_t & 1\end{bmatrix}$$

In addition if $\frac{\Delta X}{Z}$, $\frac{\Delta Y}{Z}$, & $\frac{\Delta Z}{Z}$ are negligible, it follows from Lemma 1 that

$$x_t = x\left(1 - \theta_t\frac{y}{f_y}\right)^{-1}$$

$$y_t = (y + \theta_t f_y)\left(1 - \theta_t\frac{y}{f_y}\right)^{-1}$$

Since θ_t is small, the right hand side of the equations can be expanded using the Taylor series, and higher order terms can be dropped to obtain the desired result. □

Remark: In our system the pan/tilt axes are within 6″ from the center of the optical axis (based on physical measurements). For a 2 degree pan/tilt rotation, the maximum value of $\Delta X/\Delta Y$ is around 0.2″, and ΔZ is negligible. If depth (Z) of contour observed is 40″ or more, the terms $\frac{\Delta X}{Z}$, $\frac{\Delta Y}{Z}$, & $\frac{\Delta Z}{Z}$ are less than 0.005, and thus can be ignored.

Corollary: Under the same assumptions as in Proposition 1, if the camera moves by a small pan angle θ_p, then:

$$x_p \cong (x + \theta_p f_x)\left(1 + \theta_p\frac{x}{f_x}\right)$$

$$y_p \cong y\left(1 + \theta_p\frac{x}{f_x}\right)$$

Proof: Follows from Proposition 1. □

III. Strategies for Active Calibration

Given some knowledge of stable image contours before and after a small camera motion, we would like to obtain a relation between the lens parameters and the image information. We will show that by using small pan/tilt movements, an active camera can easily estimate its center. Calibrated values of f_x and f_y can then be obtained by solving a system of linear equations.

As a first step, we will relate the focal length to other camera parameters and the pan/tilt angles. Then we will show how focal length can be eliminated from these equations if multiple independent contours are considered. The simplified formulae can be used to obtain the calibrated values of the lens center. Note that present active cameras (such as COHU MPC shown in Fig. 2) are equipped with potentiometers, which can be read for values that provide estimates of the pan/tilt angles. Fig. 3 describes the schematics of the device shown in Fig. 2.

We now prove a few propositions and corollaries, before outlining the active calibration algorithms.

Proposition 2: Under similar assumptions as in Proposition 1, if the center of the lens is estimated with a small error (δ_x, δ_y) then:

$$f_y \cong \frac{\theta_t}{(\overline{x}_t - \overline{x})}\{\overline{xy} - \overline{x}\delta_y - \overline{y}\delta_x\}$$

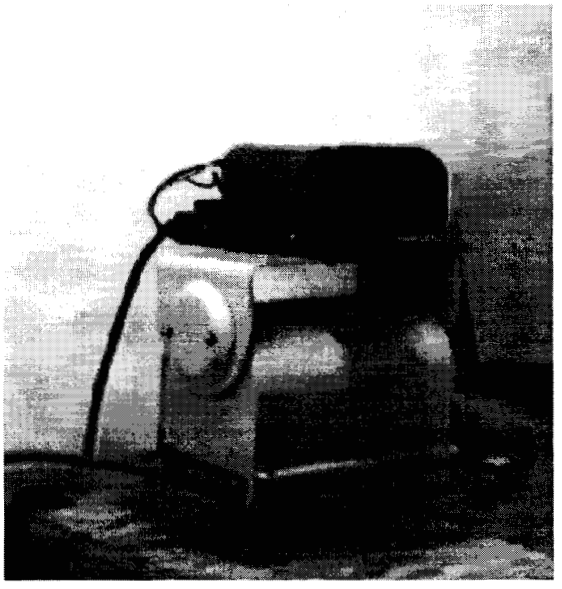


Fig. 2. An active camera.

where the symbol "$\bar{}$" denotes the average taken over a relevant image contour.

Proof: From Proposition 1,

$$x_t \cong x + \theta_t \frac{xy}{f_y}$$

When the image center is estimated with error (δ_x, δ_y) the above reduces to:

$$x_t - \delta_x \cong x - \delta_x + \frac{\theta_t}{f_y}(x - \delta_x)(y - \delta_y)$$

$$\Rightarrow \quad x_t - x \cong \frac{\theta_t}{f_y}(xy - x\delta_y - y\delta_x)$$

(Ignoring the terms in $\delta_x\delta_y$, since the error is small.) □

The result follows from the above equation, considering averages taken over specific contours.

Proposition 3: Using tilt (or pan) movements and considering three independent static contours, two linear equation in δ_x and δ_y can be obtained if negligible terms are ignored.

Proof: Let C_1 and C_2 be two different contours, and $(x^{(1)}, y^{(1)})$, $(x^{(2)}, y^{(2)})$ denote points lying on them. From Proposition 2 it follows that:

$$f_y \cong \frac{\theta_t}{(\overline{x^{(1)}}_t - \overline{x^{(1)}})}\left\{\overline{x^{(1)}y^{(1)}} - \overline{x^{(1)}}\delta_y - \overline{y^{(1)}}\delta_x\right\} \quad (1)$$

And:

$$f_y \cong \frac{\theta_t}{(\overline{x^{(2)}}_t - \overline{x^{(2)}})}\left\{\overline{x^{(2)}y^{(2)}} - \overline{x^{(2)}}\delta_y - \overline{y^{(2)}}\delta_x\right\} \quad (2)$$

Equating the right hand sides of the two equations and simplifying, we have:

$$\overline{x^{(2)}y^{(2)}} - K_1\overline{x^{(1)}y^{(1)}} = (\overline{x^{(2)}} - K_1\overline{x^{(1)}})\delta_y + (\overline{y^{(2)}} - K_1\overline{y^{(1)}})\delta_x \quad (3)$$

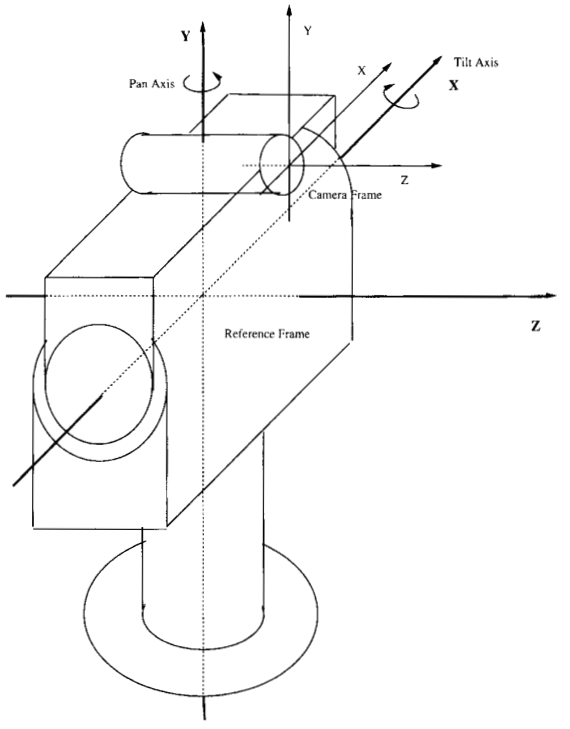


Fig. 3. Schematics of the active camera.

Where $K_1 = \frac{(\overline{x^{(2)}}_t - \overline{x^{(2)}})}{(\overline{x^{(1)}}_t - \overline{x^{(1)}})}$.

Let C_3 be the third contour, and $(x^{(3)}, y^{(3)})$ denote points lying on it. Then:

$$\overline{x^{(3)}y^{(3)}} - K_2\overline{x^{(1)}y^{(1)}} = (\overline{x^{(3)}} - K_2\overline{x^{(1)}})\delta_y + (\overline{y^{(3)}} - K_2\overline{y^{(1)}})\delta_x. \quad (4)$$

Where $K_2 = \frac{(\overline{x^{(3)}}_t - \overline{x^{(3)}})}{(\overline{x^{(1)}}_t - \overline{x^{(1)}})}$.

Once the lens center is estimated f_x and f_y can be obtained using:

$$f_x = \frac{\theta_p}{(\overline{y_p} - \overline{y})}\{\overline{xy} - \overline{x}\delta_y{}^e - \overline{y}\delta_x{}^e\} \quad (5)$$

$$f_y = \frac{\theta_t}{(\overline{x_t} - \overline{x})}\{\overline{xy} - \overline{x}\delta_y{}^e - \overline{y}\delta_x{}^e\} \quad (6)$$

Here the superscript "^{e}" denotes the estimate of a certain parameter.

The active calibration procedure discussed above can be briefly summarized as:

Strategy A:

- Using three distinct image contours, or three different positions of the camera, estimate δ_x and δ_y, using (3) and (4).
- Substituting these estimates into (5) and (6), obtain estimates for f_x and f_y.

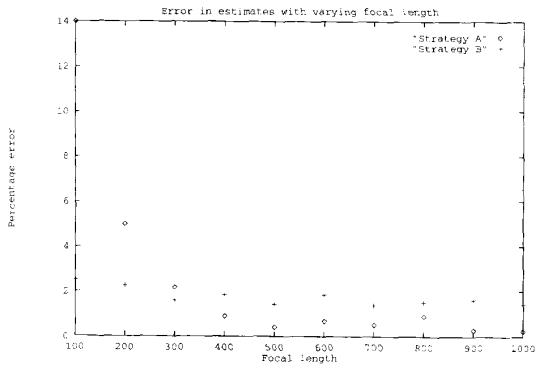


Fig. 4. Variation of error in focal length.

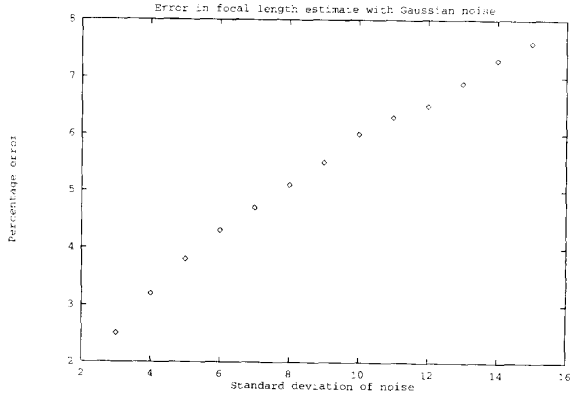


Fig. 5. Variation of error in focal length with noise (Strategy B).

A. Alternative Strategy

The major drawback of Strategy A is that the terms $(\overline{x_t} - \overline{x})$ and $(\overline{y_p} - \overline{y})$ make (5) and (6) unstable. When a camera tilts (moves up or down) the variation in the x-coordinate for any point is due to changes in perspective distortion. Similarly, when a camera pans (moves left or right) there is little change in the image y-coordinate corresponding to a given 3-d point. Thus $(\overline{x_t} - \overline{x})$ and $(\overline{y_p} - \overline{y})$ are small, usually a few pixels. In the presence of noise and inaccuracies in localization of a contour, the relative error in these terms can be large. Hence, the estimates in (5) and (6) are often unreliable.

An alternative strategy can be obtained by compromising the accuracy of the theoretical equations to obtain a more robust and simpler algorithm. Consider the following propositions:

Proposition 4: Using a single contour and pan/tilt camera movements f_x and f_y can be obtained if negligible terms are ignored.

Proof: From the second equation in Proposition 1, when δ_x, δ_y are non-zero:

$$(y_t - \delta_y) \cong (y - \delta_y + \theta_t f_y)\left(1 + \theta_t \frac{(y - \delta_y)}{f_y}\right)$$

After some simplifications we have:

$$y_t - y(1 + {\theta_t}^2) - \frac{\theta_t y^2}{f_y} \cong \theta_t f_y - \delta_y {\theta_t}^2 + \frac{\theta_t {\delta_x}^2}{f_y} - \frac{2\theta_t \delta_y y}{f_y}$$

Even if δ_x, δ_y are as large as 50 (that is, the optical center is 50 pixels away from the center of the sensor) the last three terms are negligible. This is true in our system since θ_t is less than 0.05 radians, and f_y is around 900 pixels. This gives us:

$$f_y \cong \frac{\overline{y_t} - \overline{y}(1 + {\theta_t}^2)}{\theta_t} - \frac{\overline{y^2}}{f_y} \tag{7}$$

Simplifying the above:

$${f_y}^2 - f_y \frac{\overline{y_t} - \overline{y}(1 + {\theta_t}^2)}{\theta_t} + \overline{y^2} = 0 \tag{8}$$

Similarly, f_x can be obtained by solving:

$${f_x}^2 - f_x \frac{\overline{x_p} - \overline{x}(1 + {\theta_p}^2)}{\theta_p} + \overline{x^2} = 0 \tag{9}$$

Corollary 2: Given two independent contours, pan/tilt camera movements, and estimates of f_x and f_y given by ${f_x}'$ and ${f_y}'$ respectively, δ_x and δ_y can be obtained by solving:

$${f_y}' \frac{(\overline{x^{(1)}}_t - \overline{x^{(1)}})}{\theta_t} = \{\overline{x^{(1)} y^{(1)}} - \overline{x^{(1)}}\delta_y - \overline{y^{(1)}}\delta_x\} \tag{10}$$

$${f_y}' \frac{(\overline{x^{(2)}}_t - \overline{x^{(2)}})}{\theta_t} = \{\overline{x^{(2)} y^{(2)}} - \overline{x^{(2)}}\delta_y - \overline{y^{(2)}}\delta_x\} \tag{11}$$

Proof: Follows from Proposition (2), considering two independent contours. □

Practical Considerations: Consider (8). For most practical systems the absolute value of y is less than 300, whereas f_y is greater than 500. This implies (8) is of the form $A{f_y}^2 + Bf_y + C = 0$, where $A = 1$ and B is negative, and C is small compared to B. Thus, the meaningful solution to (8) is:

$$f_y = \frac{\overline{y_t} - \overline{y}(1 + {\theta_t}^2)}{2\theta_t} + \frac{1}{2}\sqrt{\left(\frac{\overline{y}(1 + {\theta_t}^2) - \overline{y_t}}{\theta_t}\right)^2 - 4\overline{y^2}} \tag{12}$$

(The other root being close to zero can be ignored.)
Similarly, the solution for f_x is:

$$f_x = \frac{\overline{x_p} - \overline{x}(1 + {\theta_p}^2)}{2\theta_p} + \frac{1}{2}\sqrt{\left(\frac{\overline{x}(1 + {\theta_p}^2) - \overline{x_p}}{\theta_p}\right)^2 - 4\overline{x^2}} \tag{13}$$

Hence, the alternative strategy for active calibration can be summarized as:

Strategy B:

- Using a single image contour, estimate f_x and f_y from (12) and (13).
- Substituting these estimates into (10) and (11), and using another independent contour, solve for δ_x and δ_y.

B. Robustness Analysis

Proposition 5: When there is error in contour localization after pan/tilt movements, the ratio of the errors in Strategy A

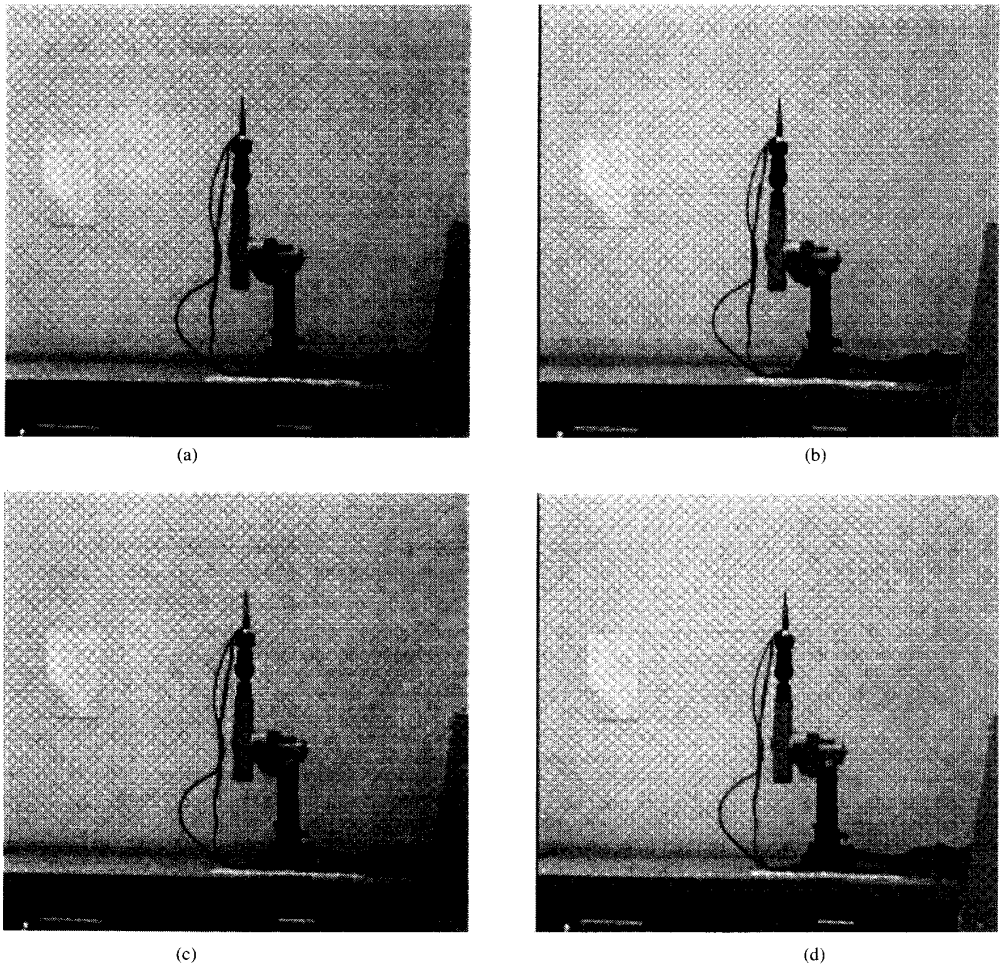


Fig. 6. Sequence of images.

compared to Strategy B for estimating f_x (f_y) *is approximately* $\frac{\overline{x}-\overline{x_p}}{\overline{y_p}-\overline{y}}$ $(\frac{\overline{y}-\overline{y_t}}{\overline{x_t}-\overline{x}})$.

Proof (Outline): The proof is outlined in the following steps:

i) Introduce similar error term in $\overline{x} - \overline{x_p}$ and $\overline{y_p} - \overline{y}$ in (13) and (5) respectively.
ii) Simplify the expressions and consider the approximate magnitude of error in both the expressions.
iii) Take the ratio of these two terms.

The above proposition has very important implications. For practical systems, the error in Strategy A can be as large as 30 times that of Strategy B, for estimating the focal lengths. Thus even though the first algorithm is theoretically more precise, it is not reliable for noisy real scenes. This point is further demonstrated by the experimental results in Section V. □

IV. Theoretical Error Analysis

In this section we study the effect of errors from various sources on the estimation of different parameters in Strategy A. The error analysis for Strategy B is similar, and thus not discussed here. First, errors in measurements of pan/tilt angles are taken into account; then the effect of noise, in the extraction of image contours, is analysed. The following remarks summarize the main results that were obtained.

Remark 1: Error in measurement of the pan (tilt) angle generates a proportional error in the estimate of f_x (f_y).

Proof: Consider (5). The estimate of f_x is proportional to the pan angle, hence any error in the measurement of this angle translates to a corresponding error in the estimate of f_x. Similarly, from (4) it follows that error in measurement of the tilt angle generates a proportional error in the estimate of f_y.□

Remark 2: Errors in measurement of the pan and tilt angles do not affect the estimate of the lens center, if and only if independent contours from the same image are considered.

Proof: Linear equations in δ_x and δ_y are obtained by equating the right hand sides of two equations. As an example, consider (1) and (2). Suppose contours are extracted from the same image, and let ϵ_1 denote the error in measurement of the tilt angle. Thus, precisely (3) is obtained as $(\theta_t + \epsilon_1)$

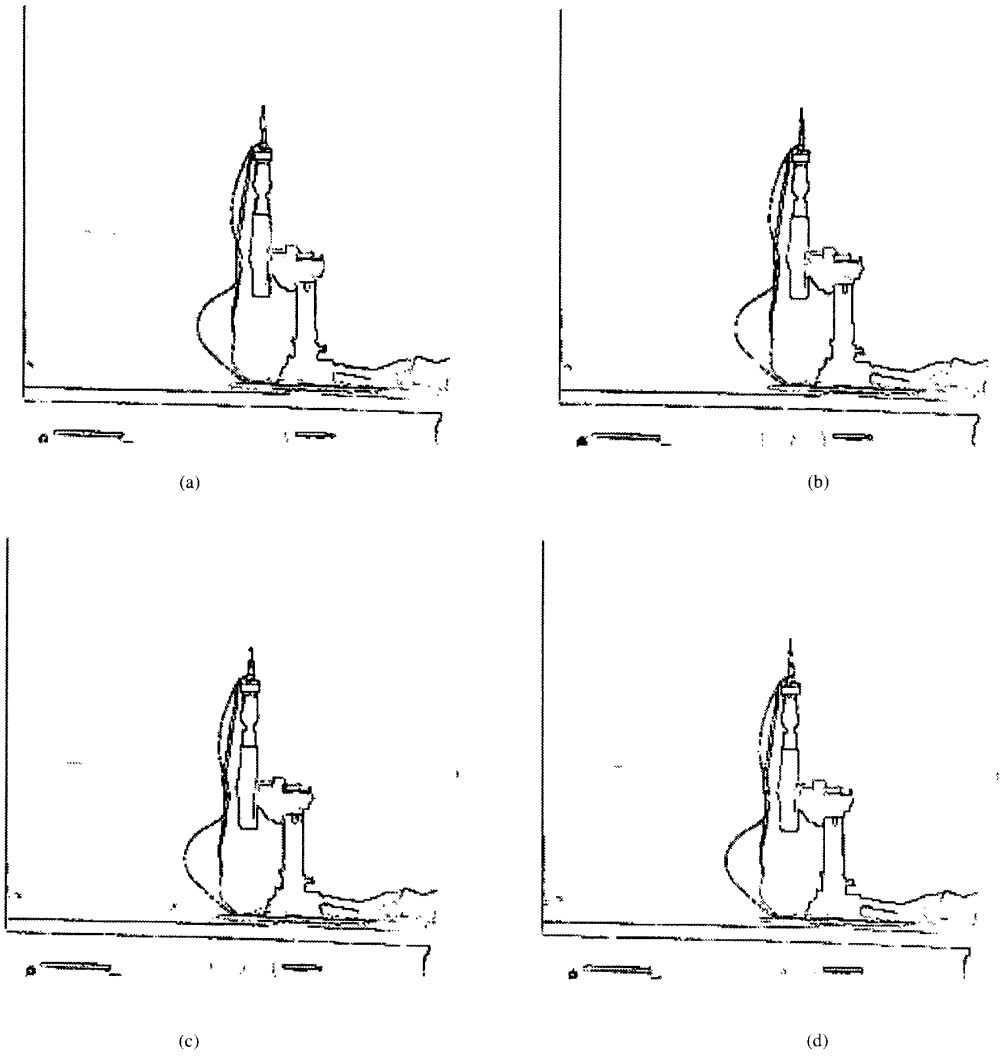


Fig. 7. Edges of image sequence.

cancels out from both sides. For this case the errors in the pan/tilt angle measurements do not affect the estimate of the lens center.

On the other hand, consider two independent images generating the contours in (1) and (2). Let ϵ_1, ϵ_2 be the errors in estimation of the tilt angles in (1) and (2) respectively. In this case K_1 in (3) gets modified to $K_1 \frac{(\theta_t+\epsilon_1)}{(\theta_t+\epsilon_2)}$. However, $\frac{(\theta_t+\epsilon_1)}{(\theta_t+\epsilon_2)}$ is not equal to 1 in general. Thus, if contours from independent images are considered, errors in angle measurements can change the estimates of the lens center. □

Remark 3: The coefficients of the linear (3)–(6) are unbiased in the presence of uncorrelated noise with zero mean.

Proof (outline): The coefficients involve a linear combination of terms of the form: $\overline{x}$, $\overline{y}$, $\overline{xy}$. These terms are unbiased in the presence of uncorrelated noise with zero mean. Thus any linear combination of terms like the above is also unbiased. □

Remark 4: When uncorrelated noise with zero mean is considered, the variance of the coefficients of (3)–(6) is inversely proportional to the number of points on a contour.

Proof: Follows from that fact that the variance of terms of the form: $\overline{x}$, $\overline{y}$, $\overline{xy}$, is inversely proportional to the number of points for which the averages were computed. [Further detail on similar analysis can be found in [6].] □

V. Experimental Results

A. Simulation

To test the validity of our algorithms, we first tested Strategy A using synthetic data. Three sets of 3D points were used to represent three independent contours. These points were projected onto the image plane and the values quantized to the nearest integer. When no noise was added Strategy

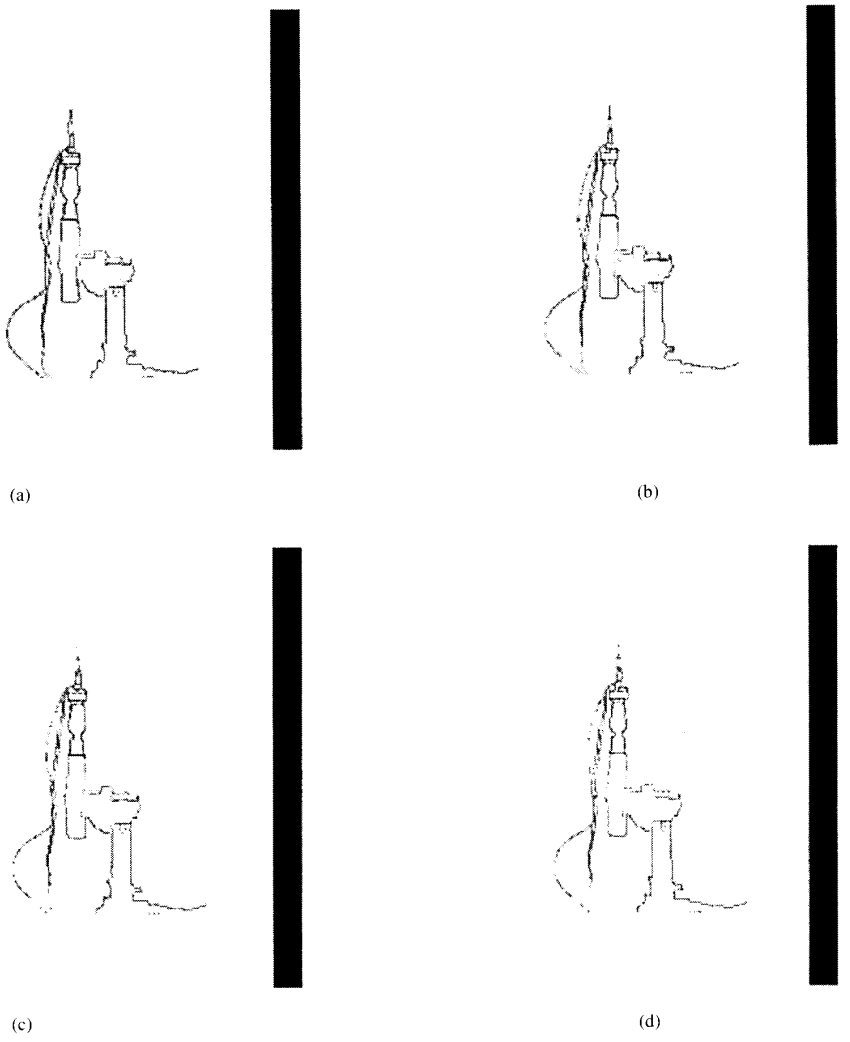


Fig. 8. Contours tracked over image sequence.

A produced more accurate estimates—**less than 1 percent relative error in focal length estimates**. Strategy B was slightly less accurate, giving estimates with a relative precision of around 1.5 percent.

In order to study the variation of error in focal length estimate, we changed f_x and f_y from 100 to 1000 by steps of 100, keeping all other parameters fixed. Fig. 4 shows the results obtained when the pan/tilt angles were fixed at 0.03 radians. The effect of discretization error seemed to influence the estimates of Strategy A more when the focal length was small. There is a simple explanation for this: Strategy A is not very robust to noise. The discretization error acts as random noise of up to half a pixel. When the focal length (in pixels) is large, this error is small relative to the focal length.[1] Thus Strategy A produces better estimates as the focal length increases. Note that the error in the estimates obtained by Strategy B does not drop off rapidly with increasing focal length—as is the case for the other method. This demonstrates that Strategy B is theoretically less accurate than Strategy A, because more terms need to be ignored to derive the equations.

When Gaussian noise was added to the image points, Strategy A performed quite poorly: it produced estimates with 20, 28, and 40 percent errors when the noise had standard deviation (σ) of 3, 4, and 5, respectively. For σ of 6 and higher, the estimates were quite unreliable (greater than 50 percent error). Strategy B, on the other hand, showed high robustness to noise. The variation of the error in the estimates with increasing noise is shown in Fig. 5. As can be seen from this figure, the second method generates reasonably good estimates (less than 8 percent error) even when σ is as large as 15.

The robustness of Strategy B follows from Proposition 5. When the camera pans by a small angle (e.g. 2–3 degrees)

[1] The inter-pixel distance measured in terms of the focal length is $\frac{1}{f_x}$ ($\frac{1}{f_y}$) in the x (y) direction. When focal length is increased the discretization error, relative to the focal length, is decreased.

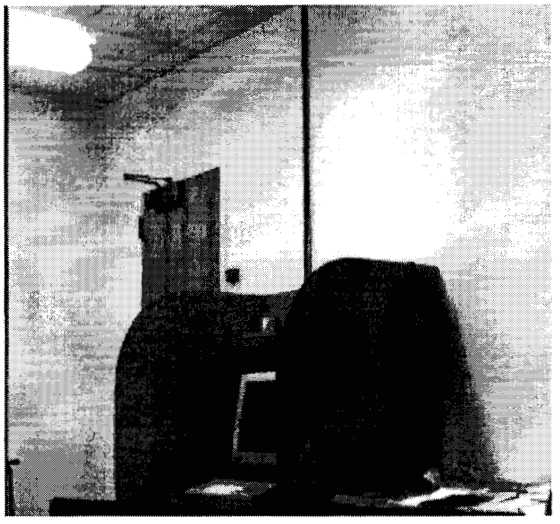


Fig. 9. Initial image.

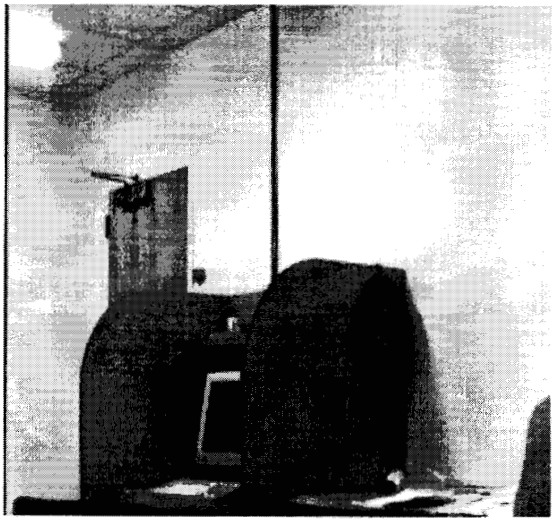


Fig. 10. Image after pan.

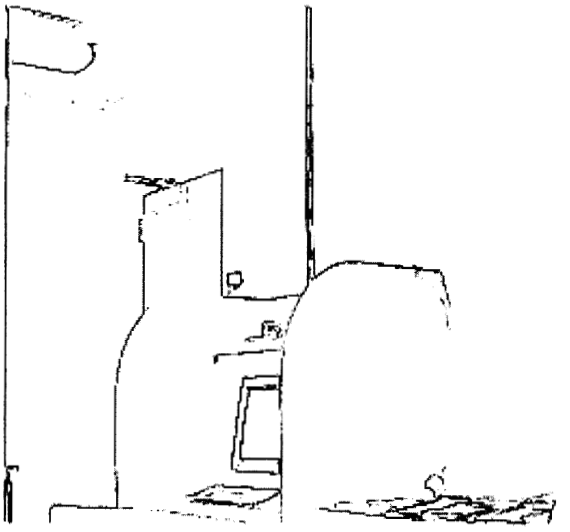


Fig. 11. Edges of initial image.

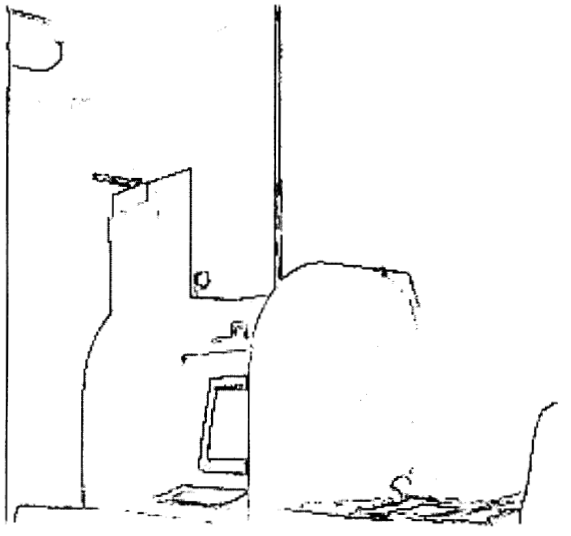


Fig. 12. Edges after pan.

$(\overline{y}_p - \overline{y})$ typically had values less than 2 pixels; $(\overline{x} - \overline{x}_p)$ on the other hand was large as 30–40 pixels. Strategy A uses terms like $(\overline{y}_p - \overline{y})$ which are affected much more by a few pixel noise than terms like $(\overline{x} - \overline{x}_p)$, that are used by Strategy B.

B. Tracking Contours

In order to perform the calibration procedure automatically, it is necessary to match contours of interest during pan/tilt motions of a camera. This was achieved by tracking image contours for small rotations. The tracking was done in two steps. First, edges in the original image, within a central window, was thickened using the morphological operation of "dilation" [19]. Then, edges after pan/tilt was AND-ed with the dilated image to extract corresponding contours after camera rotation. During pan (tilt) motion of a camera, contours move mainly horizontally (vertically). This observation was taken into account in designing accurate contour matching techniques. Fig. 6 shows a sequence of images for small pan movements of a camera. Fig. 7 shows the corresponding edge images. Contours within a pre-specified window in the initial image (top left) were considered, and tracked over the sequence of images (Fig. 8).

C. Calibration With Real Images

Finally, the calibration algorithm was tested on a real system. The active platform in our laboratory was used to

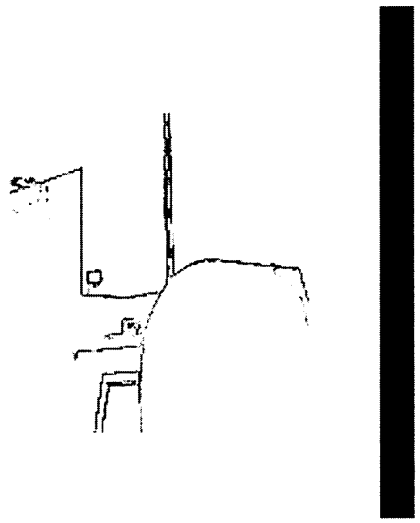


Fig. 13. Contours of initial image.

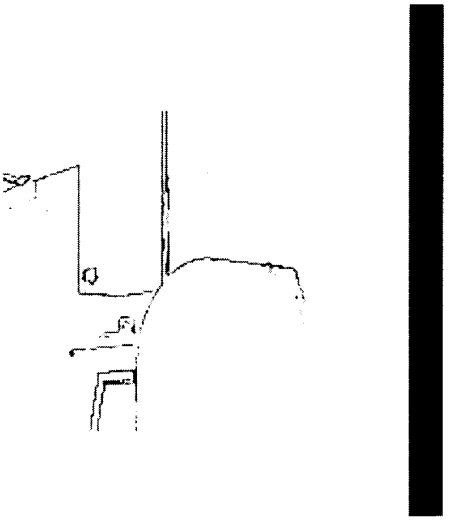


Fig. 14. Contours tracked for panned image.

acquire images of three different scenes. For each scene, three images (initial, after pan, and after tilt) were taken. Figs. 9 and 10 show the the initial and panned images for one scene. The corresponding edges are shown in Figs. 11 and 12. The other images are not shown here due to lack of space. Contours were tracked, for every step of pan/tilt camera rotation, using the procedure described above. The matching contours for the images in Figs. 9 and 10 are shown in Figs. 13 and 14. Using Strategy A, the estimates of f_x and f_y obtained by the active calibration procedure were 693 and 981. The corresponding values, obtained by presenting a known pattern and refining the initial estimates by trial and error, were 890 and 1109.[2] For Strategy B, the estimates for f_x and f_y were 917 and 1142 respectively. Thus the average relative error for these estimates was around 3 percent. This error can be attributed in part to the inaccuracies in measurements of the pan and tilt angles. The results demonstrate that the automatic active calibration procedure (using Strategy B) produces estimates fairly close to the true values.

Figs. 15 and 16 show two other environments for which the algorithms were tested (images after pan/tilt are not shown here). For the first example the estimates produced by Strategy B were 902 and 1123; in the other case 905 and 1099 were obtained as the calibration parameters. The average relative error in these examples was less than 1.5 percent. Strategy A on the other hand produced very inaccurate estimates. Note that $(\overline{x} - \overline{x}_p)$ was slightly over 30 pixels; and $(\overline{y} - \overline{y}_t)$ was around 40 pixels. On the other hand, both $(\overline{x} - \overline{x}_t)$ and $(\overline{y} - \overline{y}_p)$ were less than **1 pixel**. Thus Strategy B produced stable estimates, while Strategy A produced unstable ones. For the last two examples the pan and tilt angles were both around 2 degrees; the angles were measured by transforming a potentiometer reading. This produced a small error in the estimated angle—less than 0.05 degrees.

[2] These estimates were obtained by a variation of standard photogrammetric procedures outlined in [1]. Starting with an initial estimate, results were refined by corresponding a collection of points after pan/tilt movements. A least squares method was used to minimize the error between observed and predicted locations of the points.

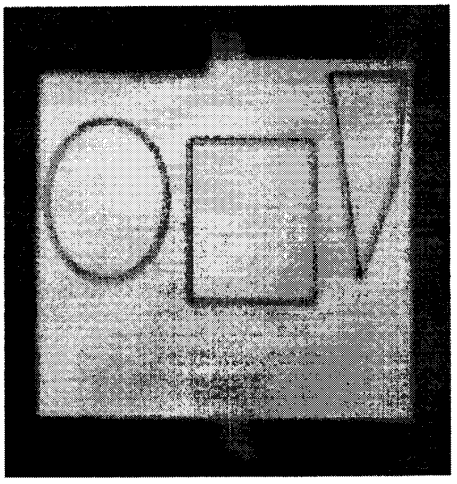


Fig. 15. Second environment.

Lens center estimates were not very accurate for both strategies since they depend on terms like $(\overline{x}_t - \overline{x})$ (i.e. purely perspective distortions) which are not robust to noise. At present we are working on developing a simple, reliable technique for estimating lens center using torsional (or roll) movements of the camera.

VI. Conclusion and Future Work

This work describes a unique way of calibrating a camera mounted on an active platform. The algorithms described do not require any predefined calibrating pattern. All that is needed are scenes with strong and stable edges. Not only are the techniques mathematically simple, they also produce

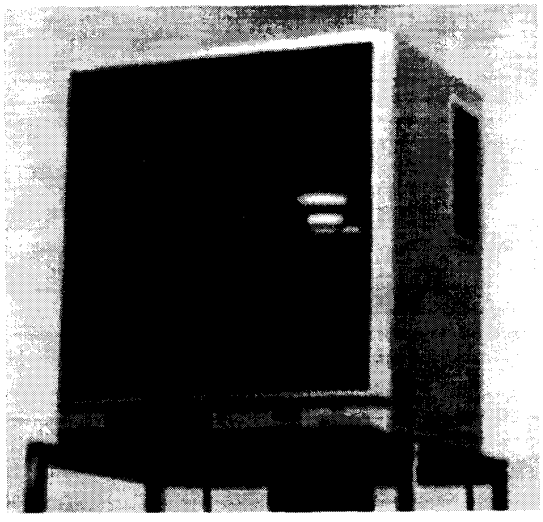


Fig. 16. Third environment.

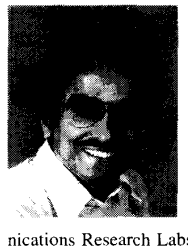
results with high accuracy when synthetic data are used. Of the two algorithms discussed, Strategy A gives almost perfect estimates in an ideal environment—contours detected without any noise, and a negligible effect of discretization error. For noisy synthetic images or real scenes, however, Strategy B is shown to be far superior.

In future work, we intend to simplify our algorithm further by considering roll movements of the camera (rotation around the z-axis) and designing a simple method to obtain the optical center of the lens. Recently, researchers have developed models for computer-controlled zoom lenses, and fish-eye lenses [26], [7]. It will be interesting to develop active calibration procedures for special lenses by integrating the theoretical models with the algorithm presented here. For some systems the pan (tilt) axis may not be parallel to the y (x) axis of the camera. Error resulting from such deviations needs to be modeled. Generalization of our method to solving the calibration problem for stereo, also needs to be studied [10].

Finally, the author wishes to thank Steve Sutphen and Sergio Licardie for developing the software involved in accurately controlling the active platform and obtaining estimates of the various parameters of the system.

REFERENCES

[1] *Manual of Photogrammetry*. American society of photogrammetry, 1980.
[2] J. Aloimons, A. Bandyopadhay and I. Weiss, "Active vision," *Int. Journal of Computer Vision*, vol. 1, no. 4, pp. 333–356, 1988.
[3] A. Bani-Hashemi, "A fourier approah to camera orientation," *IEEE Int. Conf. on Robotics and Automation*, pp. 1532–1538, 1992.
[4] A. Basu, "Active calibration," *IEEE Int. Conf. on Robotics and Automation*, pp. 764–769, 1993.
[5] ——, "Active calibration: Alternative strategy and analysis," *IEEE Int. Conf. on Computer Vision and Pattern Recongnition*, pp. 495–500, 1993.
[6] A. Basu and J. Aloimonos, "A robust, correspondenceless, translation-determining algorithm," *Int. Journal of Robotics Research*, vol. 9, no. 5, pp. 35–59, 1990.
[7] A. Basu and S. Licardie, "Modelling fish-eye lenses," *IEEE IROS Conference*, pp. 1822–1829, 1993.
[8] D. C. Brown, "Close-range camera calibration," *Photogrammetric Engineering*, vol. 37, pp. 855–866, 1971.
[9] G. Champleboux et al. "Accurate calibration of cameras and range image sensors: the npsb method," *IEEE Int. Conf. on Robotics and Automation*, pp. 1552–1557, 1992.
[10] O. Faugeras et al, "The calibration problem for stereo," *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, pp. 15–20, 1986.
[11] R. Mohr et al, "Relative positioning from geometric invariant," *IEEE Int. Conf. on computer Vision and Pattern Recongnition*, pp. 139–144, 1991
[12] W. Faig, "Calibration of close-range photogrammetry systems: mathematical formulation," *Photogrammetric Engineering*, vol. 41, pp. 1479–1486, 1975.
[13] S. Ganapthy, "Decomposition of transformation matrices for robot vision," *IEEE Int. Conf. on Robotics and Automation*, pp. 130–139, 1984.
[14] D. B. Gennery, "Stereo-camera calibration," *Proc. Image Understanding Workshop* pp. 101–108, 1979.
[15] R. Horaud, R. Mohr and B. Lorecki, "Linear-camera calibration," *IEEE Int. Conf. on Robotics and Automation*, pp. 1539–1544, 1992.
[16] K. Kanatani, "Camera rotation invariane of image characteristics," *Computer Vision, Graphics and Image Proc.*, vol. 39, no. 3, pp. 328–354, 1987.
[17] R. K. Lenz and R. Y. Tsai, "Techniques for calibration of the scale factor and image enter for high accuracy 3-d machine vision metrology," *IEEE Trans. PAMI*, vol. 10, pp. 713–720, 1988.
[18] S. Linnainmaa, D. Harwood and L. S. Davis, "Pose determination of a three-dimensinal object using triangle pairs," *IEEE Trans. PAMI*, vol. 10, pp. 634–647, 1988.
[19] P. Maragos, "Tutorial on advances in morphological image processing and analysis," *Optical Engineering*, vol. 26, no. 7, pp. 623–632, 1987.
[20] R. Mohr and E. Arbogast, "It can be done without camera calibration," *Pattern Recognition Letters*, vol. 12, pp. 39–43, 1991.
[21] U. Neisser, *Cognition and Reality*. W. H. Freeman, San Drancisco, 1976.
[22] A. Okamoto, "Orientation and construction of models, part I: The orientation problem in close-range photogrammetry," *Photogrammetric Engineering and Remote Sensing*, vol. 47, pp. 1437–1454, 1981.
[23] I. Soble, "On calibrating computer controlled cameras for perceiving 3-d scenes," *Artificial Intelligene*, vol. 5, pp. 185–198, 1974.
[24] T. M. Strat, "Recovering the camera parameters from a transformation matrix," *Proc. Image Understanding Workshop*, pp. 264–271, 1984.
[25] I. Sutherland, "Three-dimensional data input by tablet," *Proceedings of the IEEE*, vol. 62, pp. 453–461, 1974.
[26] K. Tarabanis, R. Y. Tsai and D. S. Goodman, "Modeling of a computer-controlled zoom lens," *IEEE Int. Conf. on Robotics and Automation*, pp. 1545–1551, 1992.
[27] R. Y. Tsai, A versatile camera calibration technique for high accuracy 3-d machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE J. of Robotics and Automation*, vol. 3, pp. 323–344, 1987.
[28] K. W. Wong, "Mathematical formulation and digital analysis is close-range photogrammetry," *Photogrammetric Engineering and Remote Sensing*, vol. 41, pp. 1355–1373, 1975.
[29] Y. Yakimsky and R. Cunningham, "A system for extracting three-dimensional measurements from a stereo pair of TV cameras," *Computer Graphics and Image Proessing*, vol. 7, pp. 195–210, 1978.



Anup Basu completed his B.S. in Math/Stat. and M.S. in computer science both from the Indian Statistical Institute. After working in the industry (Tata Consultancy Services, India and Strong Memorial Hospital, Biostatistics Division, Rochester, NY) for a few years, he returned to academics and completed a Ph.D. in computer science at the University of Maryland, College Park, MD.

At present he is an Assistant Professor in the Department of Computer Science at the University of Alberta, and an Adjunct Professor at Telecommunications Research Labs, Edmonton. His research interests include Computer Vision, Robotics, and Multimedia Teleconferencing.