Please verify that (1) all pages are present, (2) all figures are correct, (3) all fonts and special characters are correct, and (4) all text and figures fit within the red margin lines shown on this review document. Complete formatting information is available at http://SPIE.org/manuscripts

Return to the Manage Active Submissions page at http://spie.org/app/submissions/tasks.aspx and approve or disapprove this submission. Your manuscript will not be published without this approval. Please contact author_help@spie.org with any questions or concerns.

Digital ruler: real-time object tracking and dimension measurement using stereo cameras

James Nash, Kalin Atanassov, Sergio Goma, Vikas Ramachandra, Hasib Siddiqui
Qualcomm Inc.

ABSTRACT

Stereo metrology involves obtaining spatial estimates of an object's length or perimeter using the disparity between boundary points. True 3D scene information is required to extract length measurements of an object's projection onto the 2D image plane. In stereo vision the disparity measurement is highly sensitive to object distance, baseline distance, calibration errors, and relative movement of the left and right demarcation points between successive frames. Therefore a tracking filter is necessary to reduce position error and improve the accuracy of the length measurement to a useful level. A Cartesian coordinate extended Kalman (EKF) filter is designed based on the canonical equations of stereo vision. This filter represents a simple reference design that has not seen much exposure in the literature. A second filter formulated in a modified sensor-disparity (DS) coordinate system is also presented and shown to exhibit lower errors during a simulated experiment.

Keywords: 3-D tracking, stereo vision, filter design, extended Kalman filter

INTRODUCTION

The addition of a second camera, or indeed several cameras, to consumer mobile devices has been made possible by the continued cost reduction of CMOS cameras. The placement of multiple cameras on mobile devices allows the perception of an object's distance as well as its location on the sensor plane. Using the principles of stereo vision the functionality of the mobile device may be extended to sensing and interpreting in a 3-D world. Example applications include measuring the true sizes of objects, gesture recognition, and object cued context aware behavior.

The perception of depth arises by measuring the disparity of an object's location in the reference camera with its corresponding location at an auxiliary camera. By measuring the pixel offset in the camera buffer, knowing the calibration parameters, and measuring the location of the object keypoint in the reference camera, the X, Y, Z coordinates may be determined. The value of the pixel coordinates corresponding to the desired target object will be corrupted by noise due to imperfect ability to identify the object's keypoints in the image. Furthermore, perfect calibration cannot be relied upon therefore errors in roll, pitch, and yaw offsets between cameras will cause additional deviation of the depth measurement from its true value [1].

Small errors in pixel values translate to large errors in depth calculations [2] therefore it is imperative to use a tracking filter to reduce the errors below the measurement noise floor. A tracking filter with predictive capacity can be used to reduce errors due to hand panning motion, jitter, or to continue tracking when the object intermittently resides outside the field of view. A tracking filter may also provide additional useful information such as velocity measurements which may be useful in applications. The filter may also be combined with onboard sensors such as accelerometers to increase estimation accuracy.

In this paper two tracking filters are developed and it is shown that the design of the filter impacts performance. An EKF in Cartesian coordinates is designed based on the geometry of stereo vision and represents a standard reference design for tracking in stereo vision systems. The formulation of this filter is not ideal because the measurements are coupled and thus the noise covariance matrix is not diagonal. A second filter is designed based on a modified coordinate system related to the standard EKF through an invertible nonlinear transformation. When formulating filter problems it is desirable to decorrelate input measurements so that the maximum information is conveyed in the minimum dimension measurement space. The modified coordinate system which we call disparity-sensor (DS) coordinates uses measurements which correspond exactly to the three coordinates and therefore in theory should be independent for independent motion.

SPIE USE: ____ DB Check, ____ Prod Check, Notes:

Please verify that (1) all pages are present, (2) all figures are correct, (3) all fonts and special characters are correct, and (4) all text and figures fit within the red margin lines shown on this review document. Complete formatting information is available at http://SPIE.org/manuscripts

Return to the Manage Active Submissions page at http://spie.org/app/submissions/tasks.aspx and approve or disapprove this submission. Your manuscript will not be published without this approval. Please contact author_help@spie.org with any questions or concerns.

In the following $\mathbf{x}^*$ denotes the transpose of vector $\mathbf{x}$, and $\dot{\mathbf{x}}$ and $\dot{x}_k$ denote the time derivatives of the vector and its k[th] component, respectively. Also, the k[th] component of vector function $\mathbf{f}(\mathbf{x})$ is $\mathrm{f}_k(\mathbf{x})$.

DEVELOPMENT OF THE FILTER EQUATIONS

Let *(X,Y,Z) ,(X',Y',Z')* and *(x, y), (x', y'),(i, j), (i', j')* represent points in the Cartesian coordinate systems of the reference camera, auxiliary camera, reference sensor plane, and auxiliary sensor plane, reference image, auxiliary, image, respectively. The basic equations of a canonical stereo system are [3]

$$X = \frac{x}{d} \tag{1.1}$$

$$Y = \frac{y}{d}$$

$$Z = \frac{1}{d} \tag{1.2}$$

where the normalized disparity is $d = D/B$, D is the disparity, B is baseline separation distance, and Z is the object distance normalized by the focal length f. The sensor frame disparity is related to the pixel disparity by

$$D = x' - x = w(j' - j) \tag{1.3}$$

where w is the sensor pitch. The linear constant velocity state equation describing the object kinematics is

$$\mathbf{r}(t) = \mathbf{F}\mathbf{r}(t-1) + \mathbf{q}(t) \tag{1.4}$$

where

$$\mathbf{r}(t) = \begin{bmatrix} X(t) & Y(t) & Z(t) & \dot{X}(t) & \dot{Y}(t) & \dot{Z}(t) \end{bmatrix}^* \tag{1.5}$$

$$\mathbf{F} = \begin{bmatrix} 1 & 0 & 0 & T & 0 & 0 \\ 0 & 1 & 0 & 0 & T & 0 \\ 0 & 0 & 1 & 0 & 0 & T \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{1.6}$$

$\mathbf{q}(t)$ represents a white noise acceleration model[4] with maximum acceleration q, and T is the exposure time. The measurements are the pixel values of the object's keypoint and the disparity between the corresponding points in both cameras. These are related to the states by the measurement equation

Please verify that (1) all pages are present, (2) all figures are correct, (3) all fonts and special characters are correct, and (4) all text and figures fit within the red margin lines shown on this review document. Complete formatting information is available at http://SPIE.org/manuscripts

Return to the Manage Active Submissions page at http://spie.org/app/submissions/tasks.aspx and approve or disapprove this submission. Your manuscript will not be published without this approval. Please contact author_help@spie.org with any questions or concerns.

$$\mathbf{m}(t) = \mathbf{h}_r[\mathbf{r}(t)] + \mathbf{v}(t)$$
$$\begin{bmatrix} x \\ y \\ d \end{bmatrix} = \begin{bmatrix} h_1(\mathbf{r}) \\ h_2(\mathbf{r}) \\ h_3(\mathbf{r}) \end{bmatrix} = \begin{bmatrix} X/Z \\ Y/Z \\ 1/Z \end{bmatrix} + \mathbf{v}(t) \qquad (1.7)\backslash$$

The zero mean measurement noise **v**(t) has non-diagonal covariance matrix **R** due to the correlation between disparity and the *x* position in the sensor plane through (1.3).

Thus the state propagation is linear and the measurement is nonlinear in this formulation. A straightforward application of the extended Kalman filter (EKF) may now be used to track the 3D position. For convenience the EKF equations following the notation in [5] are mentioned briefly below.

$$\hat{\mathbf{r}}(t \mid t-1) = \mathbf{F}\hat{\mathbf{r}}(t-1), \text{ state prediction} \qquad (1.8)$$

$$\mathbf{P}(t \mid t-1) = \mathbf{F}\mathbf{P}(t)\mathbf{F}^* + \mathbf{Q}, \text{ predicted error covariance} \qquad (1.9)$$

$$\mathbf{K}(t) = \mathbf{P}(t \mid t-1)\hat{\mathbf{H}}^*(t)\left[\hat{\mathbf{H}}(t)\mathbf{P}(t \mid t-1)\hat{\mathbf{H}}^*(t) + \mathbf{R}\right]^{-1}, \text{ optimum gain} \qquad (1.10)$$

$$\hat{\mathbf{r}}(t) = \hat{\mathbf{r}}(t \mid t-1) + \mathbf{K}(t)\left[\mathbf{m}(t) - \mathbf{h}(\hat{\mathbf{r}}(t \mid t-1))\right], \text{ state correction} \qquad (1.11)$$

$$\mathbf{P}(t) = \left(\mathbf{I} - \mathbf{K}(t)\hat{\mathbf{H}}(t)\right)\mathbf{P}(t \mid t-1), \text{ error covariance update} \qquad (1.12)$$

where the Jacobian matrix $\mathbf{H}(\mathbf{r})$ with entries $H_{ij} = \partial h_i / \partial r_j$ is

$$\mathbf{H}(\mathbf{r}) = \begin{bmatrix} 1/Z & 0 & -X/Z^2 & 0 & 0 & 0 \\ 0 & 1/Z & -Y/Z^2 & 0 & 0 & 0 \\ 0 & 0 & -1/Z^2 & 0 & 0 & 0 \end{bmatrix} \qquad (1.13)$$

and $\hat{\mathbf{H}}(t) = \mathbf{H}(\hat{\mathbf{r}}(t \mid t-1))$. Initialization is to the first measurement and zero velocities

$$\hat{\mathbf{r}}(1) = \begin{bmatrix} \mathbf{h}^{-1}(\mathbf{m}(1)) & 0 & 0 & 0 \end{bmatrix} \qquad (1.14)$$

If another coordinate system can be found such that an invertible transformation exists, the filter equations may be reformulated in terms of the new coordinate system [6]. By formulating the filter equations in the new coordinate system advantages in stability and asymptotic range estimation often result because the noise covariance matrices are better posed. In the following we will develop such a transformation by algebraic methods which avoids the solution to a complicated system of nonlinear coupled differential equations.

Letting $\mathbf{s}(t)$ denote the new coordinate system we seek invertible transformations

Please verify that (1) all pages are present, (2) all figures are correct, (3) all fonts and special characters are correct, and (4) all text and figures fit within the red margin lines shown on this review document. Complete formatting information is available at http://SPIE.org/manuscripts

Return to the Manage Active Submissions page at http://spie.org/app/submissions/tasks.aspx and approve or disapprove this submission. Your manuscript will not be published without this approval. Please contact author_help@spie.org with any questions or concerns.

$$\mathbf{r} = \mathbf{f}_r(\mathbf{s}) \tag{1.15}$$

$$\mathbf{s} = \mathbf{f}_s(\mathbf{r}) \tag{1.16}$$

We may then transform (1.4) from Cartesian to SD coordinates by substituting (1.15) for $\mathbf{r}(t-1)$ and then applying (1.16) to both sides which results in

$$\mathbf{s}(t) = \mathbf{f}(\mathbf{s}(t-1)) \tag{1.17}$$

where

$$\mathbf{f}(\mathbf{s}) \square \mathbf{f}_s[\mathbf{F}\mathbf{f}_r(\mathbf{s}) + \mathbf{q}(t)] \tag{1.18}$$

This technique has many similarities with the well-known method for preserving unbiasedness in any coordinate system by converting measurements [7]. Define the sensor-disparity (SD) coordinates

$$\mathbf{s} = \frac{1}{d}\begin{bmatrix} \dot{x} & x & \dot{d} & 1 & \dot{y} & y \end{bmatrix}^* \tag{1.19}$$

In this coordinate system the differential equations of constant velocity motion are

$$\dot{\mathbf{s}} = \begin{bmatrix} -s_1 s_3 & s_1 - s_2 s_3 & -s_3^2 & -s_4 s_3 & -s_5 s_3 & s_5 - s_6 s_3 \end{bmatrix}^* \tag{1.20}$$

Differentiating(1.19), (1.5) and substituting into(1.5), (1.19), respectively, yields

$$\mathbf{s} = \mathbf{f}_s(\mathbf{r}) = \begin{bmatrix} r_4 - \frac{r_6}{r_3} r_1 & r_1 & -\frac{r_6}{r_3} & r_3 & r_5 - \frac{r_6}{r_3} r_2 & r_2 \end{bmatrix}^* \tag{1.21}$$

$$\mathbf{r} = \mathbf{f}_r(\mathbf{s}) = \begin{bmatrix} s_2 & s_6 & s_4 & s_1 - s_2 s_3 & s_5 - s_6 s_3 & -s_3 s_4 \end{bmatrix}^* \tag{1.22}$$

Applying (1.22) to the right hand side of (1.4) gives

$$\mathbf{r}(t) = \begin{bmatrix} s_2(1 - Ts_3) + Ts_1 \\ s_6(1 - Ts_3) + Ts_5 \\ s_4(1 - Ts_3) \\ s_1 - s_2 s_3 \\ s_5 - s_6 s_3 \\ -s_3 s_4 \end{bmatrix}(t-1) + \mathbf{q}(t) \tag{1.23}$$

and substituting the result into (1.21) results in the SD state equation

Please verify that (1) all pages are present, (2) all figures are correct, (3) all fonts and special characters are correct, and (4) all text and figures fit within the red margin lines shown on this review document. Complete formatting information is available at http://SPIE.org/manuscripts

Return to the Manage Active Submissions page at http://spie.org/app/submissions/tasks.aspx and approve or disapprove this submission. Your manuscript will not be published without this approval. Please contact author_help@spie.org with any questions or concerns.

$$\mathbf{s}(t) = \mathbf{f}[\mathbf{s}(t-1)]$$
$$= \begin{bmatrix} \frac{s_1}{1-Ts_3} & s_2(1-Ts_3)+Ts_1 & \frac{s_3}{1-Ts_3} & s_4(1-Ts_3) & \frac{s_5}{1-Ts_3} & s_6(1-Ts_3)+Ts_5 \end{bmatrix}(t-1)^* \quad (1.24)$$

The measurement equation is

$$\mathbf{z}(t) = \begin{bmatrix} x/d \\ y/d \\ 1/d \end{bmatrix} = \mathbf{H}\mathbf{s}(t) + \mathbf{v}_s(t) \quad (1.25)$$

with

$$\mathbf{H} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

and $\mathbf{v}_s(t)$ represents the measurement noise associated with the disparity normalized measurements. Since the measurement vector elements correspond to the to X, Y, and Z positions of the model, if the object motion is independent in all three directions, the measurement noise becomes decoupled.

We may now apply the EKF to (1.24) and (1.25) this time with the nonlinearity residing in the state equation. The equations are

$$\hat{\mathbf{s}}(t \mid t-1) = \mathbf{f}(\hat{\mathbf{s}}(t \mid t-1)), \text{ state prediction} \quad (1.26)$$

$$\mathbf{P}(t \mid t-1) = \hat{\mathbf{F}}(t)\mathbf{P}(t)\hat{\mathbf{F}}(t)^*, \text{ error covariance prediction} \quad (1.27)$$

$$\mathbf{K}(t) = \mathbf{P}(t \mid t-1)\mathbf{H}\left[\mathbf{H}\mathbf{P}(t \mid t-1)\mathbf{H} + \mathbf{R}_s\right]^{-1}, \text{ optimum gain} \quad (1.28)$$

$$\hat{\mathbf{s}}(t) = \hat{\mathbf{s}}(t \mid t-1) + \mathbf{K}_s(t)\left[\mathbf{z}(t) - \mathbf{H}_s\hat{\mathbf{s}}(t \mid t-1)\right], \text{ state correction} \quad (1.29)$$

$$\mathbf{P}(t) = \left(\mathbf{I} - \mathbf{K}(t)\mathbf{H}\right)\mathbf{P}(t \mid t-1), \text{ error covariance update} \quad (1.30)$$

where

$$\mathbf{F}(\mathbf{s}) = \begin{bmatrix} 1/(1-Ts_3) & & Ts_1/(1-Ts_3)^2 & & & \\ T & 1-Ts_3 & -Ts_2 & & & \\ & & 1/(1-Ts_3)^2 & & & \\ & & -Ts_4 & 1-Ts_3 & & \\ & & Ts_5/(1-Ts_3)^2 & & 1/(1-Ts_3) & \\ & & -Ts_6 & & T & 1-Ts_3 \end{bmatrix} \quad (1.31)$$

Please verify that (1) all pages are present, (2) all figures are correct, (3) all fonts and special characters are correct, and (4) all text and figures fit within the red margin lines shown on this review document. Complete formatting information is available at http://SPIE.org/manuscripts

Return to the Manage Active Submissions page at http://spie.org/app/submissions/tasks.aspx and approve or disapprove this submission. Your manuscript will not be published without this approval. Please contact author_help@spie.org with any questions or concerns.

with has elements $F_{ij} = \partial \mathrm{f}_i / \partial s_j$ and $\hat{\mathbf{F}}(t) = \mathbf{F}(\hat{\mathbf{s}}(t \mid t-1))$. Initialization uses the first measurement and zero velocities

$$\hat{\mathbf{s}}(1) = \mathbf{z}(1) \tag{1.32}$$

In this coordinate system the length of an object can more naturally be accommodated. If the length is placed as a state in the Cartesian filter a nonlinear state equation will result which destroys the advantage of simplicity. If the length is linear in the state, then the measurement is synthetic which causes correlations in the noise covariance. It is best to have the smallest number of uncorrelated measurements thus maximizing the information content and minimizing the dimensionality of the problem. In the DS system if we let $s_7 = l(t) = \sqrt{X(t)^2 + Y(t)^2 + Z)t)^2}$ and $s_8 = \dot{l}(t)$ then

$$s_7 = \sqrt{s_2^2 + s_6^2 + s_4^2} \tag{1.33}$$

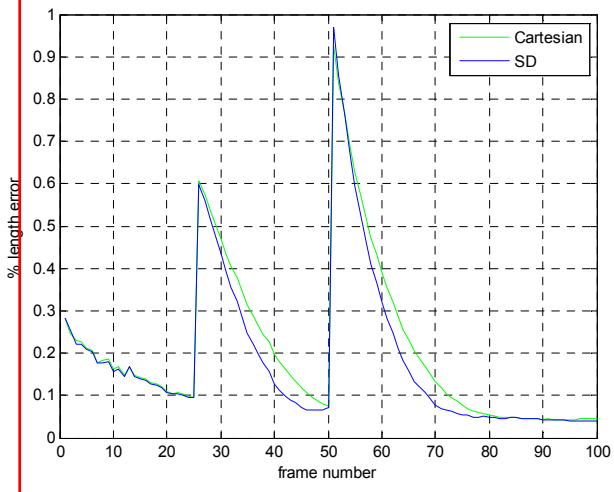
$$s_8 = \frac{1}{\sqrt{s_7}}\left[s_2\left(s_1 - s_2 s_3\right) + s_6\left(s_5 - s_6 s_3\right) - s_4^2 s_3 \right] \tag{1.34}$$

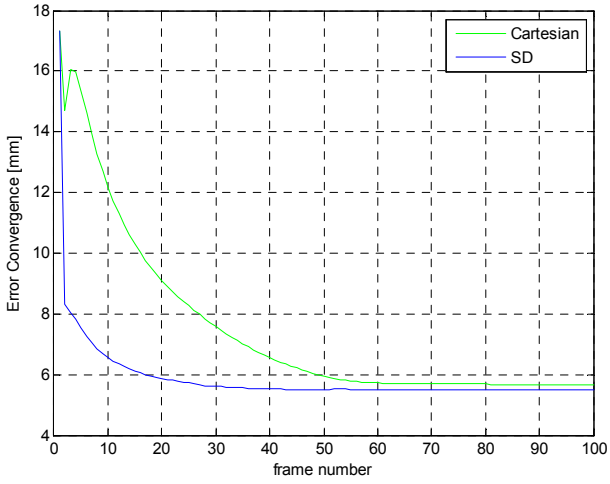
and an object's radial distance can be easily added to the state model.

SIMULATION RESULTS

A Monte-Carlo simulation was performed for 100 trials of an object whose *x* and *z* velocities reverse from 1 mm/frame at intervals of 25, and 50 frames. The same measurement data is input to both filters over 100 frames of *x, y*, and *d* measurements with errors of 4.5 and 30 pixels, respectively. The baseline is 10 cm, the frames are 5MP and the sensor is 1/2.5 format. Note that there are two sources of related errors. The errors in the location of the keypoint in both sensor frames and the error in disparity which results from those errors.

In Figure 1 the length error converges more quickly using the SD coordinate system. In Figure 2 the internal convergence of the error covariance occurs more quickly and the asymptotic error is lower.





Please verify that (1) all pages are present, (2) all figures are correct, (3) all fonts and special characters are correct, and (4) all text and figures fit within the red margin lines shown on this review document. Complete formatting information is available at http://SPIE.org/manuscripts

Return to the Manage Active Submissions page at http://spie.org/app/submissions/tasks.aspx and approve or disapprove this submission. Your manuscript will not be published without this approval. Please contact author_help@spie.org with any questions or concerns.

Figure 1. Percent error comparison in distance of two filters **Figure 2. Internal filter convergence comparison**

SYSTEM DESCRIPTION

A "virtual ruler" algorithm employs a similar filter to perform real-time object length calculations on a Qualcomm mobile development platform. The algorithm performs cued feature extraction, triangulation, position measurement, state estimation and real-time display of the overlaid spatial information. The feature extraction is general enough to permit extraction *in vivo* as well as on regular rigid bodies. Unique to this approach is a multiscale refinement procedure which iteratively improves the location of feature points responsible for object demarcation. The output of the computationally inexpensive coarse extraction procedure is passed to a fine detection algorithm which responds to normalized intensity variation in two directions. An extended Kalman filter is used to resolve the nonlinear relationship between the triangulated z-coordinate and the disparity. The filter helps reduce the error caused by random camera jitter experienced during panning and provides true 3D position, velocity, and the dimension of the user defined object. A constrained least squares triangulation procedure reduces the error caused by inconsistent motion of stereo key points relative to motion of the object. The ability of the algorithm to precisely identify extremal points combined with the tracking function overcomes a key limitation of many systems in which small changes in the position estimates induce large errors in the measurement.

SUMMARY AND CONCLUSIONS

A tracking filter is required to reduce position errors which arise due to erroneous keypoint pixel value measurements. Because depth calculation is highly sensitive to disparity measurements, slight fluctuations in estimated values can translate to large errors in position estimates. Downstream algorithms which require position estimates to compute lengths and sizes of objects will be adversely effected. A reference design for an extended Kalman filter is provided which is based on the canonical stereo camera geometry. The reference design is compared to another filter formulated in a modified coordinate system which transforms the nonlinearity in the measurements into the states equation with some benefit in noise statistics and some increase in complexity. The design of the tracking filter effects performance and for optimal results the state and measurement equations should be formulated in a coordinate system which decouples measurement errors. Future work will analyze the actual form of the covariance matrices to validate the simulation results. Application of the tracking filter for joint tracking points in both the reference and auxiliary reference frames or tracking multiple objects may also be described.

REFERENCES

[1] Zhao, W. "Effects of camera alignment errors on stereoscopic depth estimates," Pattern Recognition 29(12), 2115-2136, (1996)
[2] Blostein, S., Huang, T., "Error Analysis in Stereo Determination of 3-D Point Positions," IEEE Trans. Pattern Analysis and Machine Intelligence, 9(6), (1997)
[3] Jain, R., [Machine Vision], McGraw Hill, (1995)
[4] Bar-Shalom, Y., Kirubarajan, L., [Estimation with Applications to Tracking and Navigation], John Wiley, (2001)
[5] Mendel, J., [Lessons in Estimation Theory for Signal Processing, Communications, and Control], Prentice Hall, (1995)
[6] Aidala, V., Hammel, S., "Utilization of Modified Polar Tracking Coordinates for Bearings-Only Tracking," IEEE Trans. Automatic Control, (1983)

SPIE USE: ____ DB Check, ____ Prod Check, Notes:

Please verify that (1) all pages are present, (2) all figures are correct, (3) all fonts and special characters are correct, and (4) all text and figures fit within the red margin lines shown on this review document. Complete formatting information is available at http://SPIE.org/manuscripts

Return to the Manage Active Submissions page at http://spie.org/app/submissions/tasks.aspx and approve or disapprove this submission. Your manuscript will not be published without this approval. Please contact author_help@spie.org with any questions or concerns.

[7] Longbin, X., Bar-Shalom, Y., "Unbiased converted measurements for tracking," Aerospace and Electronic Systems, IEEE Transactions, 34(3), 1023 – 1027, (1998)

SPIE USE: ____ DB Check, ____ Prod Check, Notes: