2007 IEEE International Conference on
Robotics and Automation
Roma, Italy, 10-14 April 2007

Nearly Analytical Pose Estimation

John E. McInroy, *Senior Member, IEEE,*

Abstract— **This paper develops a new, nearly analytical method for pose estimation. Two steps are required: (1) Solution of a least squares matrix problem, followed by projection of the solution onto the nearest scaled subunitary matrix, and (2) Solution of a least squares vector problem, followed by projection onto the three dimensional Special Orthogonal group, SO(3). Although the method can be iterated when necessary, it achieves accuracy to standard stopping criteria with on average only one iteration; thus it can be considered nearly analytical. The method is compared to numerical methods and is shown to be much faster than earlier iterative methods. Both perspective and affine cameras models are treated.**

Index Terms— **pose estimation, visual servoing**

I. Introduction

Pose estimation is the calculation of a body's position and orientation from an image. Iterative, nonlinear numerical optimizations can provide fully optimal solutions which are therefore of the highest possible accuracy. There are many examples of such methods. Since this paper is concerned with rapid techniques, [1] is a fast iterative technique based on object, rather than image, iterations. Global convergence is found. Both [2] and [3] use iteratively re-weighted least squares techniques for tracking rapidly. These approaches are based on linearization using the Lie group's infinitestimal generator; thus they are suitable for tracking small changes in pose between images. Let SO(3) denote the three dimensional Special Orthogonal group modeling three dimensional rotation. Cyclic coordinate descent, which alternately finds the optimal rotation, $R \in SO(3)$, then Euclidean terms such as the translation, then recalculates R and so forth is used for camera calibration [4], [5]. Image moments are used to iteratively obtain fast pose measurements of planar objects in [6]. Closed form solutions which use only monocular images (two dimensional data) are available for $n = 3$ or $n = 4$ correspondence points. The roots of a fourth or fifth order polynomial contain the solution ([7] provides one example algorithm and references for other algorithms). For a small set of points when $n > 4$, non-iterative techniques methods are found in [8], [9], and [10]. The solution in [10] solves a quadratic problem through an over-parameterization, then multiple linear SVDs. The number of variables generated can be high. For instance, the constraint $R^T R = I$ requires 45 variables for the three dimensional element of R.

Despite the availability of many algorithms, there remains a need for a rapid, optimal, dependable algorithm. For both affine and perspective imaging, this paper develops a new method that is nearly closed form, yet provides optimal estimates even when $n > 4$. It does so by solving two distinct, unconstrained least squares problems: the first is for large rotations, while the second is for small rotations. These initial answers are then projected onto SO(3).

II. Camera Models

This section briefly presents both the affine and perspective camera models, then states the pose estimation problem for each model.

Let $p_i \in \mathbb{R}^3$ denote the i^{th} point on an object, and $d_i \in \mathbb{R}^2$ denote its image. Without loss of generality, assume the standard camera model with camera axis along the "z" dimension, so that the missing component of data is the third (or "z") component. The pin-hole model of a camera then yields the following perspective transformation [1]:

$$d_i = \frac{fP_2(Rp_i + t)}{\zeta^T(Rp_i + t)} \tag{1}$$

where f is the camera's focal length, $P_2 = [I_2\ 0]$, R is the rotation matrix from the object frame to the camera frame, t is the translation vector from the object frame to the camera frame, and $\zeta = [0\ 0\ 1]^T$.

Affine imaging systems [11] are a limiting case of perspective cameras for the situation wherein the depths of the correspondence points are much larger than the size of the object. All correspondence points are then scaled by approximately the same value, and the data loss becomes a linear loss of one dimension (along the sensing axis). This occurs when $\zeta^T Rp_i << \zeta^T t$, for all i. Then (1) becomes the affine camera model

$$d_i = \frac{fP_2(Rp_i + t)}{\zeta^T t} \tag{2}$$

The affine pose estimation problem can be phrased as follows:
Given $p_i \in \mathbb{R}^3$, $d_i \in \mathbb{R}^2$, $w_i \in \mathbb{R}_+$, $i = 1 \ldots n$, $P_2 = [I_2\ 0]$, find the minimum of

$$J = \sum_{i=1}^{n} w_i || \frac{fP_2[Rp_i + t]}{\zeta^T t} - d_i||^2 \tag{3}$$

over $g = (R, \vec{p}) \in$ SE(3) ($R \in$ SO(3), $t \in \mathbb{R}^3$). Here w_i are nonnegative real weights proportional to the quality of the i^{th} measurement.

The perspective pose estimation problem is similar:
Given $p_i \in \mathbb{R}^3$, $\vec{d_i} \in \mathbb{R}^2$, $w_i \in \mathbb{R}_+$, $i = 1 \ldots n$, $P_2 = [I_2\ 0]$, find the minimum of

$$J = \sum_{i=1}^{n} w_i ||fP_2[Rp_i + t] - \zeta^T[Rp_i + t]d_i||^2 \tag{4}$$

J. McInroy is with the Department of Electrical and Computer Engineering, University of Wyoming, Laramie, WY, 82071 E-mail:mcinroy@uwyo.edu, Phone: (307) 766-6137 , fax: (307)766-2248.

1-4244-0602-1/07/$20.00 ©2007 IEEE.

over $g = (R, \vec{p}) \in$ SE(3) ($R \in$ SO(3), $t \in \mathbb{R}^3$).
To date, minimization of either (3) or (4) has involved a nonlinear numerical optimization. This paper will derive a much faster method via a sequence of analytic solutions.

III. The closest scaled subunitary element to an arbitary matrix

This section derives methods for estimating a generalization of unitary matrices, termed scaled subunitary matrices. They will be used for estimating pose.

Definitions:

- $M_{l\times m}(\mathbb{R})$ denotes the set of $l \times m$ dimension matrices with real coefficients.
- $U_{sc}(\mathbb{R})$ denotes the set of scaled subunitary matrices. That is, $U_{sc}(\mathbb{R}) = \{sN | s \in \mathbb{R}, N \in M_{l\times m}(\mathbb{R}), l \leq m, NN^T = I\}$.
- $\hat{\cdot}$ is the cross product matrix

The following Lemma will be used extensively to find the closest element of SO(3) to a given arbitrary matrix. It is a generalization to non-square matrices of the basic concepts presented, for instance, in [12].

Lemma *1:* If $A \in M_{l\times m}(\mathbb{R})$ where $l \leq m$, the closest scaled subunitary matrix to A (in the least squares sense) is $cR \in U_{sc}(\mathbb{R})$, with

1) The SVD of A is $A = U\Sigma V^T$, where $\Sigma = [\Sigma_0 \ 0]$, Σ_0 is diagonal with elements $\Sigma_0 = \text{diag}(\sigma_1 \ \sigma_2 \ \dots \sigma_l)$, V_1 consists of the first l columns of V.
2) $c = \sigma_{avg} = (\sum_{i=1}^{l} \sigma_i)/l$
3) $R = UV_1^T$

The distance between the matrices is

$$||A - cR||_2 = \sqrt{\sum_{i=1}^{l}(\sigma_i - \sigma_{avg})^2}$$

, the standard deviation of the nonzero singular values of A. The proof is available upon request. Lemma 1 will now be used to find the scaled subunitary transformation which best relates two data matrices, P and D.

Proposition *2:* Let $P \in \mathbb{R}^{m\times n}, D \in \mathbb{R}^{l\times n}$ be given, where null(P^T) = 0 and $D = QP$ for some scaled subunitary Q. Then the minimizer of $||AP - D||_2^2$ over $A \in U_{sc}(\mathbb{R})$ is given by solving the linear, unconstrained equations $AP = D$ over $M_{m\times n}(\mathbb{R})$. That is, $A = DP^\dagger$, where $P^\dagger$ is the pseudo inverse of P. Then, find the closest scaled subunitary matrix to A from Lemma 1

Proof: The minimum of $||AP - D||_2^2$ is 0 when $AP = D$. Since $D = QP$, $AP = D = QP$. This implies that $A = QPP^\dagger$. Since null(P^T)=0, $PP^\dagger = I$, thus $A = Q$, and the global minimum of 0 can be attained. Let the SVD of $A = U[\Sigma_0 \ 0]V^T = Q$. Since $Q \in U_{sc}(\mathbb{R})$, $QQ^T = k^2 I$ for some $k \in \mathbb{R}$. Thus $U\Sigma_0^2 U^T = k^2 I$ and $\Sigma_0 = kI$. Then $A = Q = U\Sigma_0 V_1^T = kUV_1^T = kR$. $k = \sigma_1 = \dots = \sigma_l$, therefore $k = \sigma_{avg}$.

□

Proposition 2 simply shows that when no error is present (i.e. $D = QP$, $Q \in U_{sc}(\mathbb{R})$), minimization in the unconstrained linear space of matrices $M_{l\times m}(\mathbb{R})$ and then finding its closest element in $U_{sc}(\mathbb{R})$ is equivalent to the far more difficult problem of solving the *constrained* minimization. This involves a little sleight of hand–since no error is present, the initial unconstrained solution is already in $U_{sc}(\mathbb{R})$. The theorem presents this temporary unnecessary step to set the stage for realistic, imperfect data matrices, $D \neq QP$. Theorem 3 shows that, when error is present, this method yields a **second order** approximation.

Theorem *3:* Let $P \in \mathbb{R}^{m\times n}, D \in \mathbb{R}^{l\times n}$ be given, where null(P^T) = 0. A second order approximation to the problem

$$J_* = min_{B\in U_{sc}(\mathbb{R})}||BP - D||_2^2 \tag{5}$$

can be found by analytically solving the unconstrained, linear problem

$$J_{lb} = min_{A\in M_{l\times m}(\mathbb{R})}||AP - D||_2^2 \tag{6}$$

A second order approximation to (5) is then the closest element in $U_{sc}(\mathbb{R})$ to the A minimizing (6).

Proof: From Lemma 1, the closest element in $U_{sc}(\mathbb{R})$ to the minimizing A, $A = DP^\dagger$, is $cR = \sigma_{avg}UV_1^T$, where the SVD of $A = U[\Sigma_0 \ 0]V^T$, and V_1 contains the first l columns of V. The cost when using this estimate is

$$J_{est} = ||cRP - D||_2^2 \tag{7}$$

To complete the proof, it must be shown that this cost is a second order or better approximation of the optimal cost, J_*. That is, for some constant matrix G, $J_{est} \leq J_* + ||EG||_2^2$ where E is an error matrix modelling the difference between A and cR.

Since null(P^T)=0, Σ_0^{-1} exists and cR can be written as $cR = \sigma_{avg}U\Sigma_0^{-1}U^T A$. Let $I + E = \sigma_{avg}U\Sigma_0^{-1}U^T$. Then

$$E = \sigma_{avg}U[\Sigma_0^{-1} - I]U^T \tag{8}$$

Since $cR = (I + E)A$, (7) becomes

$$J_{est} = ||(I + E)AP - D||_2^2$$

or

$$J_{est} = J_{lb} + 2tr([AP - D][EAP]^T) + ||EAP||_2^2 \tag{9}$$

Since $A = DP^\dagger$, $AP = DP^\dagger P = DV_{p1}V_{p1}^T$, where V_{p1} contains the nonzero input directions to P (these can be obtained from the SVD of P). Since $V_{p1}^T V_{p1} = I$, $(AP - D)(AP)^T = 0$, and the middle, first order term of (9) becomes zero. Thus (9) becomes

$$J_{est} = J_{lb} + ||EAP||_2^2 \tag{10}$$

Because the minimum of (6) is found over the entire vector space of $l \times m$ matrices, while (5) is restricted to scaled subunitary $l \times m$ matrices, $J_{lb} \leq J_*$. Therefore $J_{est} \leq J_* + ||EAP||_2^2$

□

Section V will show that these techniques when applied to pose estimation yield very accurate answers that are, for many applications, sufficiently accurate to eliminate the need for any further search. This surprising result is explained by Prop. 2 and Theorem 3: (1) The method yields an exact solution for exact data (by Prop. 2), and (2) The approximation error is small when noise is present, because it is a second order approximation, with zero first order error (by Theorem 3).

IV. Estimation of Pose Using Affine and Perspective Cameras

These new results will now be used to find second order approximations to the affine and perspective pose estimation problems.

Theorem *4:* Let $p_i \in \mathbb{R}^3$, $i = 1, \dots, n$ be known points on an object, and $d_i \in \mathbb{R}^2$, $i = 1, \dots, n$ be their images taken by an affine camera with focal length f. Let w_i be real positive weights proportional to the quality of the i^{th} measurement ($w_i \in \mathbb{R}_+$, $i = 1 \dots n$), and $P_2 = [I_2\ 0]$. Let

$$\tilde{p_i} = \sqrt{w_i}[p_i - p_{avg}],\ \tilde{d_i} = \sqrt{w_i}[d_i - d_{avg}]$$

where

$$p_{avg} = \frac{\sum_{i=1}^n w_i p_i}{\sum_{i=1}^n w_i},\ d_{avg} = \frac{\sum_{i=1}^n w_i d_i}{\sum_{i=1}^n w_i}$$

and

$$\tilde{P} = [\tilde{p_1}\ \tilde{p_2}\ \cdots \tilde{p_n}],\ \tilde{D} = [\tilde{d_1}\ \tilde{d_2}\ \cdots \tilde{d_n}]$$

Then, if null($\tilde{P}^T$) = 0, a second order, analytical approximation to the homogeneous transformation from the object frame to the camera frame $g = (R, t) \in$ SE(3) ($R \in$ SO(3), $t \in \mathbb{R}^3$) minimizing

$$J = \sum_{i=1}^n w_i || \frac{fP_2[Rp_i + t]}{\zeta^T t} - d_i||^2 \tag{11}$$

where $\zeta = [0\ 0\ 1]^T$ can be found by the following steps:

1) Calculate $A = \tilde{D}\tilde{P}^\dagger$, and its SVD, $A = U[\Sigma_0\ 0][V_1\ V_2]^T$. Then $\sigma_{avg} = (\sigma_1 + \sigma_2)/2$ where
$$\Sigma_0 = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix}$$
2) $P_2R = UV_1^T$, $t_z = \zeta^T t = \frac{f}{\sigma_{avg}}$.
3) $P_2 t = \frac{d_{avg}}{\sigma_{avg}} - P_2 R p_{avg}$,
$$t = \begin{bmatrix} P_2 t \\ t_z \end{bmatrix}$$
4) Let $P_2 R = \begin{bmatrix} x^T \\ y^T \end{bmatrix}$, $z = \hat{x}y$, then
$$R = \begin{bmatrix} P_2 R \\ z^T \end{bmatrix}$$

If no noise is present, the solution is exact.

Proof: Let $P_2 t = t_a$. Then $J = \sum_{i=1}^n w_i || \frac{fP_2Rp_i + ft_a]}{\zeta^T t} - d_i||^2$. Minimizing J with respect to t_a gives

$$t_a = \frac{t_z d_{avg}}{f} - P_2 R p_{avg} \tag{12}$$

Substituting (12) into J gives $J = \sum_{i=1}^n ||\frac{f}{t_z}P_2R\tilde{p_i} - \tilde{d_i}||^2$. Writing this in terms of matrix norms gives $J = ||\frac{f}{t_z}P_2R\tilde{P} - \tilde{D}||^2$. A second order approximation to the minimum of J is given by Theorem 3: Find $A = \tilde{D}\tilde{P}^\dagger$, and its SVD $A = U[\Sigma_0\ 0][V_1\ V_2]^T$. Then $P_2R = UV_1^T$ and $\sigma_{avg} = \frac{f}{t_z}$, thus $t_z = \frac{f}{\sigma_{avg}}$. Use (12) to find $P_2 t = t_a$. This solution will be exact for noise free data by Prop. 2. Finally, since $R \in$ SO(3), the missing row can be found from the cross product of the two known rows, i.e. let the rows of P_2R be denoted as x^T and y^T. Then $z = \hat{x}y$ and

$$R = \begin{bmatrix} P_2 R \\ z^T \end{bmatrix}$$

Prop. 2 implies that the solution is exact when the data is noise free.

□

The perspective transform destroys the affine structure, but the same essential concepts are used in Theorem 5 to find second order approximations in the perspective case.

Theorem *5:* Let $p_i \in \mathbb{R}^3$, $i = 1, \dots, n$ be known points on an object, and $d_i \in \mathbb{R}^2$, $i = 1, \dots, n$ be their images taken by a perspective camera with focal length f. Let positive weights, proportional to how good the i^{th} measurement is, be denoted $w_i \in \mathbb{R}_+$, $i = 1 \dots n$, and $P_2 = [I_2\ 0]$.

When the data is not over-constrained, a second order approximation of the homogeneous transformation from the object frame to the camera frame $g = (R, t) \in$ SE(3) ($R \in$ SO(3), $t \in \mathbb{R}^3$) minimizing

$$J = \sum_{i=1}^n w_i ||fP_2[Rp_i + t] - \zeta^T(Rp_i + t)d_i||^2$$

where $\zeta = [0\ 0\ 1]^T$ can be found by the following steps:

1) Calculate the matrices:
$$B_i = \sqrt{w_i}[fP_2 - d_i\zeta^T],\ \beta = (\sum_{i=1}^n B_i^T B_i)^{-1}$$
$$N_i = \begin{bmatrix} p_i^T & 0 & 0 \\ 0 & p_i^T & 0 \\ 0 & 0 & p_i^T \end{bmatrix},\ M = \beta \sum_{i=1}^n B_i^T B_i N_i$$
2) Stack these matrices as follows:
$$F = \begin{bmatrix} B_1(N_1 - M) \\ B_2(N_2 - M) \\ \vdots \\ B_n(N_n - M) \end{bmatrix}$$
3) Find the input direction to F, v_{F9}, which gives the minimal norm output. Since F is typically a tall matrix, with dimension $2n \times 9$, this can be done by the SVD of $F^T F = V_F \Sigma_F V_F^T$. Then v_{F9} is the last column of V_F.
4) Partition v_{F9} into three, three element vectors as
$$v_{F9} = \begin{bmatrix} x_e \\ y_e \\ z_e \end{bmatrix}$$

5) Rearrange these vectors into a matrix, A, and find its SVD:
$$A = \begin{bmatrix} x_e^T \\ y_e^T \\ z_e^T \end{bmatrix} = U\Sigma V^T$$
6) The rotation matrix, R is given by $R = UV^T$.
7) The translation vector, t is given by $t = -\beta\sum_{i=1}^{n}(B_i^T B_i R p_i)$

Proof: Minimizing J with respect to t gives
$$t = -\beta\sum_{i=1}^{n}(B_i^T B_i R p_i) \tag{13}$$
The i^{th} error term of J is
$$e_i = \sqrt{w_i}(fP_2 - d_i\zeta^T)[Rp_i + t]$$
Substituting in (13) yields
$$e_i = \sqrt{w_i}(fP_2 - d_i\zeta^T)[Rp_i - \beta\sum_{j=1}^{n}(B_j^T B_j R p_j)] \tag{14}$$
This equation is linear in the unknown matrix, R. Let the rows of R be denoted as
$$R = \begin{bmatrix} x^T \\ y^T \\ z^T \end{bmatrix} \text{ and form } v = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \tag{15}$$
Note that
$$||R||_2^2 = ||v||_2^2 = 3 \tag{16}$$
Equation (14) can then be rewritten as
$$e_i = B_i[N_i - \beta\sum_{j=1}^{n}(B_j^T B_j R p_j)]v = B_i[N_i - M]v$$
Stacking these individual errors into a vector gives the equation
$$\begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} = \vec{e} = Fv \tag{17}$$
J can now be rewritten as $J = ||\vec{e}||_2^2$. It is therefore minimized when v equals the minimum input direction of F, v_{F9}. Finding the SVD or Schur decomposition of $F^TF = V_F\Sigma_F V_F^T$ provides v_{F9}, which is the last column of V_F. This vectorized estimate will now be formed back into a matrix, then its closest unitary neighbor will be found. Let
$$v_{F9} = \begin{bmatrix} x_e \\ y_e \\ z_e \end{bmatrix}$$
Rearrange these vectors into a matrix, A, and find its SVD:
$$A = \begin{bmatrix} x_e^T \\ y_e^T \\ z_e^T \end{bmatrix} = U\Sigma V^T$$
The rotation matrix, R is given by $R = UV^T$, while the translation vector is given from (13), $t = -\beta\sum_{i=1}^{n}(B_i^T B_i R p_i)$.

It will now be shown that this approximation has zero first order additional cost when the data is not overconstrained. From (17), the cost is $J = ||Fv||_2^2$. The minimizing input direction v_{F9} is found by minimizing over the entire real vector space $\mathbb{R}^9$, while the optimum must enforce the additional orthonormality constraints on R. Therefore a lower bound on the cost is given by $J_{lb} = ||Fv_{F9}||_2^2$. To make the cost computations consistent, the length of v_{F9} should be scaled from (16) to be $||v_{F9}||_2^2 = 3$. Since R does not depend on the length of v_{F9}, this does not change the value of R (or t, which is calculated from R).

The orthogonality constraints are imposed by finding the closest unitary matrix to A, $R = UV^T$. Vectorizing R by the process in (15), the cost when using this estimate is then $J_{est} = ||Fv||_2^2$. Let $v = \tilde{v} + v_{F9}$, where $\tilde{v}$ models the difference between v and v_{F9}. Then
$$J_{est} = ||F(\tilde{v} + v_{F9})||_2^2 = J_{lb} + 2\tilde{v}^T F^T F v_{F9} + ||F\tilde{v}||_2^2$$
Since v_{F9} is the minimum input direction of F,
$$J_{est} = J_{lb} + 2\sigma_{min}\tilde{v}^T F^T u_{F9} + ||F\tilde{v}||_2^2 \tag{18}$$
where σ_{min} is the minimum singular value of F and u_{F9} is the minimum output direction of F (by convention the 9^{th} column of U_F). Is is well known that four appropriate points are sufficient to determine pose, therefore consider the case where $n = 4$. F is then of dimension 8×9, which implies $\sigma_{min} = 0$. This is the case where the data is not overconstrained. Then
$$J_{est} = J_{lb} + ||F\tilde{v}||_2^2 \tag{19}$$
Thus the approximation is second order.

□

Both Theorems 4 and 5 minimize J when noise free (perfect) data is present, which in this case gives J=0. In this sense, they can be considered to be analytical solutions to the pose problem. They are not, on the other hand, guaranteed to minimize J in the more realistic case when noise is present. Fundamentally, this occurs because minimizing in an unconstrained space and then finding the closest point in a constrained space is not in general equal to minimizing within the constrained space directly. What we have found very surprising, however, is the high level of accuracy–thousands of different noise scenarios, many with very large noise levels, have demonstrated experimentally that the new method reliably gives rotation matrices very close to the optimum. Section V will provide more details, but for many applications, Theorems 4 and 5 yield sufficient accuracy to consider them a nearly analytical solution: no other calculations are needed. This stems from the second order nature of the approximation.

When higher accuracy is desired, the basic concept of minimizing in the unconstrained space and then projecting onto the constrained space can be modified to find rapid iterative solutions. A first order approximation of R is parameterized as $A = \alpha I + \hat{w}$. The four variables ($w \in \mathbb{R}^3$,

$\alpha \in \mathbb{R}$) are then found by an unconstrained least squares minimization of J. R is then estimated by finding the closest element in SO(3) to $\alpha I + \hat{w}$.

Theorem *6:* Pose estimates found using Theorem 5 can be improved (i.e. J can be reduced) by the following steps:

1) Using Theorem 5, find an initial estimate of the homogeneous transformation from the object frame to the camera frame,

$$T^{(0)} = \begin{bmatrix} R^{(0)} & t^{(0)} \\ 0 & 1 \end{bmatrix}$$

Let $k = 1$, and let superscripts in parenthesis indicate the k^{th} estimate. The initial data is: $p_i^{(0)} = p_i$.

2) Transform the object data points to the k^{th} estimate of the camera frame, $p_i^{(k)} = R^{(k-1)} p_i^{(k-1)} + t^{(k-1)}$.

3) Calculate the matrices:

$$C^{(k)} = -\beta[\sum_{i=1}^{n} B_i^T B_i p_i^{(k)}],\ E^{(k)} = \beta[\sum_{i=1}^{n} B_i^T B_i \hat{p}_i^{(k)}]$$

B_i and β are already available from the Theorem 5 calculations.

4) Stack these matrices as follows:

$$G^{(k)} = \begin{bmatrix} B_1(-\hat{p}_2^{(k)} + E^{(k)} \quad p_2^{(k)} + C^{(k)}) \\ B_2(-\hat{p}_2^{(k)} + E^{(k)} \quad p_2^{(k)} + C^{(k)}) \\ \vdots \\ B_n(-\hat{p}_n^{(k)} + E^{(k)} \quad p_n^{(k)} + C^{(k)}) \end{bmatrix}$$

5) Calculate the minimum input direction of $G^{(k)}$, $q^{(k)}$. Let $w^{(k)}$ equal the first three elements of $q^{(k)}$, and $\alpha^{(k)}$ equal the last.

6) Calculate $A^{(k)} = \alpha^{(k)} I + \hat{w}^{(k)}$, and its SVD $A^{(k)} = U\Sigma V^T$.

7) $R^{(k)} = UV^T$, $t^{(k)} = -\beta \sum_{i=1}^{n} (B_i^T B_i R^{(k)} p_i^{(k)})$

8) Update the homogeneous transform estimate as

$$T^{(k)} = \begin{bmatrix} R^{(k)} & t^{(k)} \\ 0 & 1 \end{bmatrix} T^{(k-1)}$$

9) If $||w^{(k)}|| \to 0$, stop. Otherwise, $k = k + 1$ and go to step 2.

Proof: The essence of this proof is to find an estimate, update the data to contain that estimate, then estimate again. It will use the same concept used in Theorem 5, but let A have only the four variables contained in $A = \alpha I + \hat{w}$. The data is updated by letting $p_i^{(k)} = R^{(k-1)} p_i^{(k-1)} + t^{(k-1)}$. A new cost function which uses this data is then

$$J^{(k)} = \sum_{i=1}^{n} w_i ||f P_2[R^{(k)} p_i^{(k)} + t^{(k)}] - \zeta^T (R^{(k)} p_i^{(k)} + t^{(k)}) d_i||^2$$

Minimizing $J^{(k)}$ with respect to $t^{(k)}$ gives

$$t^{(k)} = -\beta \sum_{i=1}^{n} (B_i^T B_i R^{(k)} p_i^{(k)}) \qquad (20)$$

The i^{th} error term of $J^{(k)}$ is

$$e_i^{(k)} = \sqrt{w_i}(f P_2 - d_i \zeta^T)[R^{(k)} p_i^{(k)} + t^{(k)}]$$

Substituting in (20) yields

$$e_i^{(k)} = \sqrt{w_i} B_i [R^{(k)} p_i^{(k)} - \beta \sum_{j=1}^{n} (B_j^T B_j R^{(k)} p_j^{(k)})] \qquad (21)$$

This equation is linear in the unknown matrix, $R^{(k)}$. Let $R^{(k)} = \alpha^{(k)} I + \hat{w}^{(k)}$, then Equation (21) can be rewritten as

$$e_i^{(k)} = \sqrt{w_i} B_i [(\alpha^{(k)} I + \hat{w}^{(k)}) p_i^{(k)}$$
$$-\beta \sum_{j=1}^{n} (B_j^T B_j (\alpha^{(k)} I + \hat{w^{(k)}}) p_j^{(k)})]$$

or

$$e_i^{(k)} = B_i[-\hat{p}_i^{(k)} + E^{(k)} \ p_i^{(k)} + C^{(k)}] q^{(k)}$$

where

$$C^{(k)} = -\beta[\sum_{i=1}^{n} B_i^T B_i p_i^{(k)}],\ E^{(k)} = \beta[\sum_{i=1}^{n} B_i^T B_i \hat{p}_i^{(k)}]$$

and

$$q^{(k)} = \begin{bmatrix} w^{(k)} \\ \alpha^{(k)} \end{bmatrix}$$

Stacking these individual errors into a vector gives the equation

$$\begin{bmatrix} e_1^{(k)} \\ e_2^{(k)} \\ \vdots \\ e_n^{(k)} \end{bmatrix} = \vec{e}^{(k)} = G^{(k)} q^{(k)} \qquad (22)$$

$J^{(k)}$ can now be rewritten as $J^{(k)} = ||\vec{e}^{(k)}||_2^2$. It is therefore minimized when $q^{(k)}$ equals the minimum input direction of $G^{(k)}$. Lemma 1 can then be used to estimate the rotation, closest to $A^{(k)} = \alpha^{(k)} I + \hat{w}^{(k)}$. Finally, the translation can be estimated from (20) to be $t^{(k)} = -\beta \sum_{i=1}^{n} (B_i^T B_i R^{(k)} p_i^{(k)})$. The estimate of the rotation will be accurate if the singular values are all nearly equal. Since the singular values are $[\sqrt{||w^{(k)}||^2 + (\alpha^{(k)})^2} \ \sqrt{||w^{(k)}||^2 + (\alpha^{(k)})^2} \ \alpha^{(k)}]$, they become equal as $||w^{(k)}|| \to 0$.

□

This proof shows how the iterative scheme arises as a first order approximation of rotation matrices, but no formal proof yet exists to show why it converges. The next section will provide thousands of simulations illustrating its convergence.

V. Experimental Results

This section now presents simulation results comparing the new algorithms to the method of Lu, Hager, and Mjolsness [1]. Their method is chosen for comparison due to several reasons. First, its performance did out-shine other methods we implemented. Second, it uses some similar concepts (SVDs in particular). Third, it is emerging as a popular technique. Finally, the authors offer downloadable code from their website, therefore it can serve as a benchmark.

The results from two types of data sets will be presented. All data sets pass through the perspective transformation in (1). The focal length is $f = 1$, since [1] fixes it there.

Characteristic	Affine	Perspect	AF+iter.	Lu,Hager,Mjol.
% Erroneous	0	71.9	0	0
Mean Num. Iter.	0	0	1.00	32.7
RMS θ Error [^{0}]	1.93	40.5	1.19	1.19
RMS X error [m]	0.136	1.15	0.081	0.081
RMS Y error [m]	0.129	1.43	0.084	0.084
RMS Z error [m]	13.7	119.	8.12	8.14

TABLE I

THE Z COMPONENT OF THIS DATA SET IS ON THE ORDER OF 100 TIMES LARGER THAN THE X AND Y COMPONENTS, MAKING THE DATA AFFINE.

Characteristic	Affine	Perspect	Per+iter.	Lu,Hager,Mjol.
% Erroneous	49.1	9.30	6.95	14.2
Mean Num. Iter.	0	0	1.54	42.2
RMS θ Error [^{0}]	41.5	15.2	0.507	1.69
RMS X error [m]	2.85	0.945	0.035	0.045
RMS Y error [m]	2.76	0.873	0.033	0.038
RMS Z error [m]	2.04	1.15	0.061	0.069

TABLE II

THE Z COMPONENT OF THIS DATA SET IS POSITIVE BUT OTHERWISE STATISTICALLY THE SAME AS THE X AND Y COMPONENTS

The object points, p_i, translation, t, and axis of rotation , w, ($R = exp(\hat{w})$) are chosen with Gaussian distributions. Four different techniques will be compared: (1) The affine method (Theorem 4); (2) The perspective method (Theorem 5); (3) The new iterative method (Theorem 6), and (4) Lu, Hager, and Mjolsness. In this order, they are sorted from least to most required calculations. Heavy levels of normally distributed noise are added. Answers more than 60^0 off are considered to be erroneous answers. Root Mean Square (RMS) errors in angle and translation are calculated excluding these erroneous answers, with the percentage of erroneous answers listed separately. The stopping criteria for the new iterative method (3) is chosen to match the default accuracy of method (4).

The differences in the data sets arises by changing the range of the objects, which appears in the Z component of t, t_z. Table I summarizes the results when t_z is on the order of 100 times larger than the X and Y components, making the affine camera model fit well even though the actual data comes from the perspective transformation (1). Using the affine method as an initial starting point, then iterating (AF+iter) emerges as the clearly superior technique, as it produces slightly better estimates than Lu's (4), but requires on average only 1.00 iterations to equal the accuracy obtained using an average of 32.7 iterations of Lu's (4).

Second, Table II summarizes the results when t_z is first chosen from the same distribution as the X and Y components, but then the absolute value of that t_z is taken to ensure positivity. Due to the high noise levels, each of the algorithms yield erroneous answers, the affine technique almost half of the times. This is to be expected, since the affine camera model is no longer a good approximation for this data. The new iterative method again is the clear winner, using on average 1.54 iterations to equal the accuracy obtained using an average of 42.2 iterations of Lu's (4). However, its initial starting point is provided by the perspective method (2), since that camera model applies. Averaging the results of these two simulations, the new iterative method requires on average only 1.27 iterations to match the accuracy of 37.4 iterations using the Lu, Hager, and Mjolsness technique.

VI. CONCLUSIONS

A new pose estimation method which produces an exact solution for exact data, and a second order approximation sufficiently accurate for many applications with imperfect data is derived. A powerful new iterative method relying on the new concepts is developed and compared to other iterative schemes. Simulations show that, for the same level of accuracy, on average only 1.27 iterations are needed for the new method, versus 37.4 for the popular method in [1].

REFERENCES

[1] C. Lu, G. Hager, and E. Mjolsness, "Fast and globally convergent pose estimation from video images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 6, pp. 610–622, June 2000.

[2] A. Comport, D. Kragic, E. Marchand, and F. Chaumette, "Robust real-time visual tracking: Comparison, theoretical analysis and performance evaluation," in *IEEE International Conference on Robotics and Automation*, Barcelona, Spain, April 2006, pp. 2852–2857.

[3] T. Drummond and R. Cipolla, "Real-time visual tracking of complex structures," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 932–946, July 2002.

[4] F. Park, J. Kim, and C. Kee, "Geometric descent algorithms for attitude determination using the global positioning system," *Journal of Guidance, Control, and Dynamics*, vol. 23, no. 1, pp. 850–863, January-February 2000.

[5] S. Gwak, J. Kim, and F. Park, "Numerical optimization on the euclidean group with applications to camera calibration," *IEEE Transactions on Robotics and Automation*, vol. 19, no. 1, pp. 65–74, February 2003.

[6] O. Tahri and F. Chaumette, "Complex objects pose estimation based on image moment invariants," in *IEEE International Conference on Robotics and Automation*, Barcelona, Spain, April 2006, pp. 438–443.

[7] D. DeMenthon and L. Davis, "Exact and approximate solutions of the perspective three-point problem," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 11, pp. 1100–1105, November 1992.

[8] L. Quan and Z. Lan, "Linear n-point camera pose determination," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, pp. 774–780, 1999.

[9] P. Fiore, "Efficient linear solution of exterior orientation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, pp. 140–148, 2001.

[10] A. Ansar and K. Daniilidis, "Linear pose estimation from points or lines," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 5, pp. 578–589, May 2003.

[11] R. Vidal, "Multi-subspace methods for motion segmentation from affine, perspective, and central panoramic cameras," in *IEEE International Conference on Robotics and Automation*, Barcelona, Spain, April 2005, pp. 1228–1233.

[12] B. Horn, H. M. Hilden, and S. Negahdaripour, "Closed-form solution of absolute orientation using orthonormal matrices," *Journal of the Optical Society of America*, vol. 5, pp. 1127–1135, 1988.