# POSE AND MOTION ESTIMATION USING EXTENDED KALMAN FILTER

***J Narendra Babu 1 ****

*1. Associate Professor, Dept of ECE, SIETK, PUTTOR, INDIA.*

**ABSTRACT:**

In this paper measurements from a monocular vision system are fused with inertial/magnetic measurements from an Inertial Measurement Unit (IMU) rigidly connected to the camera. Two Extended Kalman filters (EKFs) were developed to estimate the pose of the IMU/camera sensor moving relative to a rigid scene (ego-motion), based on a set of fiducials. The two filters were identical as for the state equation and the measurement equations of the inertial/magnetic sensors. The DLT-based EKF exploited visual estimates of the ego-motion using a variant of the Direct Linear Transformation (DLT) method; the error-driven EKF exploited pseudo-measurements based on the projection errors from measured two-dimensional point features to the corresponding three-dimensional fiducials. The two filters were off-line analyzed in different experimental conditions and compared to a purely IMU-based EKF used for estimating the orientation of the IMU/camera sensor. The DLT-based EKF was more accurate than the error-driven EKF, less robust against loss of visual features, and equivalent in terms of computational complexity. Orientation root mean square errors (RMSEs) of 1° (1.5°), and position RMSEs of 3.5 mm (10 mm) were achieved in our experiments by the DLT-based EKF (error-driven EKF); by contrast, orientation RMSEs of 1.6° were achieved by the purely IMU-based EKF.

**KEYWORDS:** sensor fusion; extended Kalman filtering; inertial/magnetic sensing; monocular vision; ego-motion.

## I.INTRODUCTION

Estimation of relative three-dimensional (3–D) position and orientation (pose) and structure as well as relative motion between two reference frames is an important problem in robotic guidance, manipulation, and assembly as well as in other areas such as photogrammetry, tracking, object recognition, and camera calibration. Pose is defined for an object in 3–D cartesian space consisting of an object reference frame and a base reference frame. The pose with respect to the base frame comprises the three position coordinates of the object reference frame origin and the three orientation angles of the object frame. Remote estimation of the relative pose and motion of a 3–D object without physically touching the object or without human intervention is the fundamental problem examined here. For example, an autonomous robot that moves along a path may need to determine its position relative to an obstacle. Similarly, a mechanical hand grasping a moving object requires that the gripper motion be matched to the object and then placed at the correct position and orientation. Automated spacecraft docking requires the measurement of pose relative to the docking port. The pose can be expressed relative to the sensing element frame or with respect to the base frame. This pose estimation or determination problem has been a subject of considerable research for many years in computer vision, photogrammetry, and robotics with many solutions having been proposed. The methods available are generally deterministic and use single-vision cameras, stereo-vision cameras, or more direct 3–D measuring techniques such as range images from laser, ultrasonic, and structured lighting devices. Two general classifications of sensors can be made: active and passive. Active devices (e.g., laser range finders) are those which emit energy and sense its reflection to determine the sensor response, while passive devices such as video cameras sense the natural or background energy (e.g., visible light or infrared) reflected off an object. This distinction is important in an application such as a battlefield environment, where a sensor that emits energy could be detected by enemy surveillance. However, relying on background energy alone may be insufficient in providing a good signal-to-noise ratio for feature detection and object recognition. Poor visibility caused by fog, smoke, or rain can produce images that are highly variable in quality with difficult to recognize features. Noise considerations, therefore, must be taken into account to a much greater extent than would be the case in a laboratory environment where illumination can be easily controlled. Even in a wellcontrolled setting, sensor noise must be considered since intrinsic sources can introduce a substantial amount of measurement noise. Depending on the accuracy required for pose measurement, this noise can either be ignored or the effects can be minimized. Generally, the highest accuracy is needed for camera calibration while model matching or grasping may permit larger errors in the estimate. Accuracy requirements vary for docking and tracking applications. The general problem developed here is to locate an object and measure its motion in three dimensions based on three position coordinate parameters and three rotation coordinate parameters either relative to an observer or with respect to a fixed-reference frame. Figure 1.1 shows the basic problem with each reference frame and transformation shown. Known as the exterior orientation problem in photogrammetry [1], this question has been addressed for photographs using a number of manual methods dating back to 1879 [2]. More recently, beginning in the 1960's, methods using computer vision techniques have been developed. Most measurement techniques for pose determination or estimation are image-based and can be classified into two major categories. These categories are point-based methods and model-based methods using higher-order geometric primitives. Each type involves acquiring an image, either two-dimensional (2–D) or 3–D, and then processing that image to arrive at a value for the pose. Noise effects may also be analyzed for either type. Specific techniques and applications that have been proposed are described in Section 1.3. The particular sensor used can determine the ultimate accuracy achieved. The precise layout of the elements in the sensing charge coupled device (CCD) array along with the use of low distortion lenses gives low geometric distortion for CCD imaging cameras. Calibration for lens distortion can reduce the error still further. Feature detection and extraction software can then locate an object's feature in 2–D spatial coordinates to subpixel accuracy provided the signal-to-noise ratio is high. Noise and uncertainty are present due to the response nonuniformity of the individual CCD pixels along with photon and thermal noise. Laser range finders likewise have uncertainty in both range as well as geometric location correspondence. The imaged scene also influences the accuracy achieved. Extracting features from an object of low contrast with its surroundings is not as straightforward nor as accurate as one with high contrast. To overcome some of the limitations of natural objects, artificial markings may be placed on the object whose position relative to the object is known. These markings can consist of a high-contrast pattern that maximizes signal-to-noise ratio and optimizes a particular pose determination algorithm. Other factors to be considered for pose and motion determination are the computational complexity and robustness of a method. The algorithm chosen may require a large computational overhead due to an optimization solution technique requiring many iterations. The choice of method usually involves a trade-off between accuracy and noise rejection versus computation time. Robustness in this context is defined as the ability of an algorithm to provide a result within a known error bound or to indicate that the error condition cannot be met. Measurement data outside a known error bound relative to a working model may be ignored or given low weight. Obtaining accurate results for autonomous pose determination in everyday environments in which people operate with ease is a difficult problem and one that has not yet been fully solved. Variable conditions such as lighting and visibility, along with the presence of noise from various sources, are principle difficulties that need to be examined. This work is an attempt to model these uncertainties and imprecisions in such environments as applied to pose and motion estimation. The term estimation is used as opposed to determination since the modeling is done in a statistical sense where estimated results are computed. Other methods may determine the pose by directly calculating a set of equations or by calculating a least squared error solution that does not model the noise or process stochastically.

## II.RELATED WORK

This section provides a summary of previous work relating to noncontact position and orientation measurement methods as described in the published literature. Existing research shows a large number of position determination methods applicable to a variety of applications. While not as numerous, pose estimation methods have also been described, and these are summarized as well. Variations in the methods include type and location of the sensor, illumination requirements, the object or scene feature on which the pose is calculated, relative motion of the robot or object, iterative versus direct solution, and modeling of uncertainty or noise in an attempt to improve the robustness and accuracy of the results. Applications proposed are for research programs or for commercial prototype development in the general areas of robot location, manufacturing, camera calibration, and tracking of a moving object.

Point-based methods rely on the identification and location of feature points on a target object from a 2–D image of the scene. A rigid body is generally assumed but no explicit geometric model is given. Information concerning the geometric shape other than size is not used in calculating the pose. Coordinates of the points in a local or world reference frame may or may not be known. Methods of this class, referred to as N-point perspective, were the first to be studied and, as a result, have been more extensively developed than model-based methods [3]. A perspective model that assumes the projection of a 3–D object onto a 2–D image plane through a pinhole camera model is generally used [1]. Both single-image and stereo methods have been reported; however, single-vision techniques have, by far, the greatest number of solutions. One reason for this is that point correspondence with an object from a single image is easier to determine than correspondences between two images and the object as required in stereo. The general framework is, given N corresponding points in the object and in the image, to solve for the relative pose between the camera and the object. The minimum N that produces a finite number of solutions is three, although, up to four solutions are possible. Four coplanar, noncollinear points give a unique solution. Four or five noncoplanar, noncollinear points may result in two solutions. For N greater than five noncollinear points, the result is unique and consists of an overdetermined set that can be solved using least squared error methods [4]. In general, as N increases, the accuracy of the results increases [5]. These overdetermined solutions are used for camera calibration in which a large number of points are needed through minimization of an error criteria to achieve the desired accuracy and to calculate both the external and internal camera parameters [5]. Three- and four-point coplanar targets have been directly used for pose determination. With stereo cameras, three corresponding points on an object are sufficient to uniquely identify the relative pose of the object although uncertainty may be reduced through a larger number of points [6]. Range images, similarly, can determine pose with a minimum of three points. An advantage of 3–D range images over stereo is that the correspondence problem is not present since the three coordinates of an object point are determined directly. Algorithms for these techniques are primarily iterative. For the three- and four-point special cases, however, closed-form solutions have been demonstrated [7, 4, 8, 9]. Trabasso and Zielinski [10] describe an approximate calibration method for calculating exterior orientation parameters and the x and y scaling factors. A calibration block consisting of four coplanar points is assumed to be perpendicular to the optical axis of the camera. This method has been tested in an experimental work cell for robot grasping and placement of automobile bodyshells whose plane is approximately perpendicular to the camera axis. Placement accuracies of one millimeter were achieved. Tsai has described a technique for high-accuracy, 3–D camera calibration using standard television cameras and lenses [5]. This paper gives a good survey of the calibration and pose determination techniques at the time of its publication. It also proposes a new two-stage method based on the radial alignment constraint that determines the six extrinsic and intrinsic parameters including focal length, two radial lens distortion coefficients, and the x coordinate scale factor. The first stage is a direct linear solution neglecting lens distortion. This result is used as the initial estimate in the nonlinear second stage that takes into account the lens distortion. An overdetermined set of points is used to achieve high accuracies using least squared error fitting. Using the radial alignment constraint, solutions are given for both coplanar calibration points and noncoplanar points. Test results are given where the number of calibration points used is 60. Total accuracy in 3–D is about one part in 2000. Fewer points gave higher errors while more points did not give a significant improvement in accuracy. The error obtained is approximately one-half the total theoretical predicted error.

**III.PROPOSED METHODOLOGY**

The perspective projection equations for lines from 3-D to 2-D are nonlinear. Calculation of relative position and orientation from the 3-D coordinates of corresponding points in two reference frames is also n nlinear. For the minimum three points needed to calculate a pose, up to four solutions are possible. Solutions with four or more points are overconstrained with a minimum error criterion used to determine the result. Motion estimation requires a dynamic system model that is nonlinear in the rotational dynamics. Direct linear estimation techniques, such as the standard Kalman filter, are therefore not applicable. The IEKF, however, is highly suitable as a nonlinear estimator.

**3.1 SYSTEM MODEL:**

The state assignment estimates the transformation between the camera and the object reference frames and the first derivatives of this transformation. The assignment is based on the dual quaternion representation of the 3-D transformation. Thirteen state variables are present: t, and z', terms are the linear translation and linear velocity, respectively; q, is the rotational quaternion; and w, is the rotational velocity in each axis. Translation, rather than the dual part of the dual quaternion, is estimated in the state vector since the dual part can readily be calculated from the translation and the rotational real quaternion.

### 3.2 Iterated Extended Kalman Filter Representation

The system and measurement models of the IEKF have been given above. An initial state estimate is required based on prior knowledge. An initial error covariance matrix, PO, that is dependent on the prior knowledge must also be specified. Process and measurement noise covariance matrices, Qk: and Rk, respectively, are required. After each iteration, the derived partial derivatives with respect to each state variable are calculated to determine the linearized equations. With each updated state estimate, a new linearization is then performed about this state. Several iterations about the updated state may then be needed to achieve convergence.

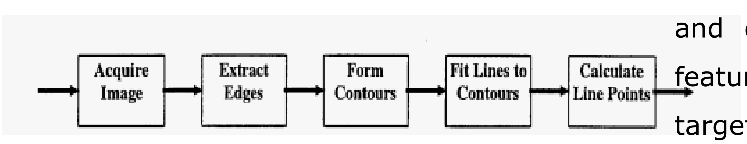


Fig1: Estimation of motion values

## IV.EXPERIMENTAL RESULTS

This section provides an evaluation for the line-based method through both simulation and actual physical system testing. A dynamic vision application for the approach is defined where a camera provides the visual feedback for a robot performing a task. Simulation testing measures the accuracy of the estimation under an assumed noise distribution and magnitude. Speed of convergence, mean square error, and stability are presented. Actual test results measuring relative motion and position from a robot arm are also shown. Relative accuracy as well as noise characterization is provided. Simulation testing was performed to evaluate the performance of the pose estimation method under a variety of conditions. The results are compared with corresponding results from a point-based extended Kalman filtering method using an identical target to illustrate the performance differences. An ideal camera model is used in the simulation. Perspective projection is assumed for the camera with a known effective focal length. Noise of an assumed magnitude and distribution is added to the image feature locations before processing. A target object consisting of four coplanar points in a rectangular pattern is simulated with individual feature points. Pairs of these points, when extracted from the image plane, are connected together to form lines. Initial conditions requiring specification include the initial state, .TO, and the error covariance matrix, PO. The state vector may be considered a collection of Gaussian random variables with covariance Po. The initial state is a sample ta.ken from each random variable. Process noise given by the covariance matrix Q is also specified as an initial condition for the simulations, remaining constant throughout. Similarly, measurement noise given by the covariance matrix R is initially specified as a constant.

## V.CONCLUSION

This paper proposes a Dual Quaternion Multiplicative Extended Kalman Filter (DQMEKF) for pose estimation that is an extension of the well-known and widely used Quaternion Multiplicative Extended Kalman Filter (Q-MEKF) for spacecraft attitude estimation. By using the dual quaternion multiplication and the concept of error unit dual quaternion, the two algebraic constraints of unit dual quaternions are automatically satisfied during the measurement update of the DQ-MEKF and the number of states is reduced from eight to six. Three different forms of the DQ-MEKF are presented, each with a different application in mind. Experimental results show that the DQ-MEKF does not encounter singularities and is accurate, precise, and fast enough for operational use. Moreover, when compared with two other EKF formulations, experimental results and Monte-Carlo simulations suggest that the DQ-MEKF might be the best formulation if the measurements are expressed in a different reference frame than the variable to be estimated. This is the case, for example, when one needs the inertial position of a satellite expressed in the body frame, e.g., to implement a control law, but the measurements are expressed in the inertial frame, like the inertial position measurements produced by a GPS. Finally, it should be mentioned that whereas the derivations presented in this paper do not rely on a model of the system dynamics, as they may be hard to model accurately enough, it is relatively straightforward to do so, if desired.

## VI.REFERENCES

[1] J. S.-C. Yuan, "A general photogrammetric method for determining object position and orientation," IEEE Transactions on Robotics and Automation, vol. 5, no. 2, pp. 129–142, April 1989.

[2] R. M. Haralick and H. Joo, "2d-3d pose estimation," in Proceedings of the International Conference on Pattern Recognition, 1988, pp. 385–391.

[3] R. J. Holt and A. N. Netravali, "Camera calibration problem: Some new results," CVGIP: Image Understanding, vol. 54, no. 3, pp. 368–383, November 1991.

[4] M. A. Abidi and T. Chandra, "Pose estimation for camera calibration and landmark tracking," in Proceedings of the IEEE International Conference on Robotics and Automation, May 1990, pp. 420–426.

[5] R. Y. Tsai, "A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses," IEEE Journal of Robotics and Automation, vol. RA-3, no. 4, pp. 323–344, August 1987.

[6] V. Torre, A. Verri, and A. Fiumicelli, "Stereo accuracy for robotics," in Robotics Research, The Third International Symposium, G. Giralt O. D. Faugeras, Ed., pp. 5–9. MIT Press, 1986.

[7] D. F. DeMenthon and L. S. Davis, "Model-based object pose in 25 lines of code," in Computer Vision– ECCV '92, G. Sandini, Ed., pp. 335–343. Springer-Verlag, 1992.

[8] M. A. Abidi and T. Chandra, "A new efficient and direct solution for pose estimation using quadrangular targets: Algorithm and evaluation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 17, no. 5, pp. 534–538, May 1995.

[9] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," Communications of the ACM, vol. 24, no. 6, pp. 381–395, June 1981.

[10] L. G. Trabasso and C. Zielinski, "Semi-automatic calibration procedure for the visionrobot interface applied to scale model decoration," Robotica, vol. 10, pp. 303–308, 1992.

[11] M. W. Walker, L. Shao, and R. A. Volz, "Estimating 3-d location parameters using dual number quaternions," CVGIP: Image Understanding, vol. 54, no. 3, pp. 358–367, November 1991.

[12] T. Q. Phong, R. Horaud, A. Yassine, and D. T. Pham, "Optimal estimation of object pose from a single perspective view," in International Conference on Computer Vision, February 1993, pp. 534–539.

[13] T. Phong, R. Horaud, A. Yassine, and D. Pham, "Object pose from 2-d to 3-d point and line correspondences," International Journal of Computer Vision, vol. 15, pp. 225–243, 1995.

[14] M. A. Abidi and R. C. Gonzalez, "The use of multisensor data for robotic applications," IEEE Transactions on Robotics and Automation, vol. 6, no. 2, pp. 159–177, April 1990.

[15] N. Chen, J. R. Birk, and R. B. Kelley, "Estimating workpiece pose using the feature points method," IEEE Transactions on Automatic Control, vol. AC-25, no. 6, pp. 1027–1041, December 1980.

[16] R. M. Haralick, "Pose estimation from corresponding point data," IEEE Transactions on Systems, Man, and Cybernetics, vol. 19, no. 6, pp. 1426–1446, November 1989.