[9] E. Chong and S. H. Zak, *An Introduction to Optimization*, 2nd ed. New York: Wiley, 2001, ISBN-10: 0471391263.
[10] S. Durola, P. Danès, D. Coutinho, and M. Courdesses, "Rational systems and matrix inequalities to the multicriteria analysis of visual servos," in *Proc. IEEE Int. Conf. Robot. Autom.*, Kobe, Japan, May 2009, pp.1504–1509.
[11] P. Danès, D. Bellot, "Towards an LMI approach to multicriteria visual servoing in robotics," *Eur. J. Control*, vol. 12, no. 1, pp. 86–110, 2006.
[12] R. Findeisen and F. Allgöwer, "An introduction to nonlinear model predictive control," presented at the Benelux Meeting Syst. Control, Veldhoven, Pays Bas, The Netherlands, 2002.
[13] J. Gangloff and M. De Mathelin, "Visual servoing of a 6 dof manipulator for unknown 3-D profile following," *IEEE Trans. Robot. Autom.*, vol. 18, no. 4, pp. 511–520, Aug. 2002.
[14] R. Ginhoux, J. Gangloff, M. De Mathelin, M. Soler, and L. Sanchez, "Active filtering of physiological motion in robotized surgery using predictive control," *IEEE Trans. Robot. Autom.*, vol. 21, no. 1, pp. 67–79, Feb. 2005.
[15] K. Hashimoto and H. Kimura, "LQ optimal and nonlinear approaches to visual servoing," in *Visual Servoing*(World Scientific Series in Robotics and Intelligent Systems), K. Hashimoto, Ed, vol. 7. Singapore: World Scientific, 1993, pp. 165–198.
[16] M. Kazemi, K. Gupta, and M. Mehrandezh, "Global path planning for robust visual servoing in complex environments," in *Proc. IEEE Int. Conf. Robot. Autom.*, Kobe, Japan, May 2009, pp. 326–332.
[17] R. Mahony, P. Corke, and F. Chaumette, "Choice of image features for depth-axis control in image-based visual servo control," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Lausanne, Switzerland, Oct. 2002, pp. 390–395.
[18] Y. Mezouar and F. Chaumette, "Optimal camera trajectory with image-based control," *Int. J. Robot. Res.*, vol. 22, no. 10, pp. 781–804, 2003.
[19] T. Murao, T. Yamada, and M. Fujita, "Predictive visual feedback control with eye-in-hand system via stabilizing receding horizon approach," in *Proc. 45th IEEE CDC*, San Diego, CA, Dec. 2006, pp.1758–1763.
[20] M. Morari and E. Zafiriou, *Robust Control*. Paris, France: Dunod, 1983.
[21] N. Papanikolopoulos, P. Khosla, and T. Kanade, "Visual tracking of a moving target by a camera mounted on a robot: A combination of vision and control," *IEEE Trans. Robot. Autom.*, vol. 9, no. 1, pp. 14–35, Feb. 1993.
[22] M. Sauvée, P. Poignet, E. Dombre, and E. Courtial, "Image based visual servoing through nonlinear model predictive control," in *Proc. 45th IEEE CDC*, San Diego, CA, Dec. 2006, pp.1776–1781.
[23] F. Schramm and G. Morel, "Ensuring visibility in calibration-free path planning for image-based visual servoing," *IEEE Trans. Robot. Autom.*, vol. 22, no. 4, pp. 848–854, Aug. 2006.

A Kalman-Filter-Based Method for Pose Estimation in Visual Servoing

Farrokh Janabi-Sharifi and Mohammed Marey

***Abstract*—The problem of estimating position and orientation (pose) of an object in real time constitutes an important issue for vision-based control of robots. Many vision-based pose-estimation schemes in robot control rely on an extended Kalman filter (EKF) that requires tuning of filter parameters. To obtain satisfactory results, EKF-based techniques rely on "known" noise statistics, initial object pose, and sufficiently high sampling rates for good approximation of measurement-function linearization. Deviations from such assumptions usually lead to degraded pose estimation during visual servoing. In this paper, a new algorithm, namely iterative adaptive EKF (IAEKF), is proposed by integrating mechanisms for noise adaptation and iterative-measurement linearization. The experimental results are provided to demonstrate the superiority of IAEKF in dealing with erroneous *a priori* statistics, poor pose initialization, variations in the sampling rate, and trajectory dynamics.**

***Index Terms*—Adaptation, Kalman filter (KF), control, pose estimation, robotic manipulator, visual servoing.**

I. Introduction

In computer vision, the problem of *pose estimation* is to determine the position and orientation (pose) of a camera with respect to an object's coordinate frame using the image information. The problem is also known as extrinsic camera-calibration problem with its solution playing a crucial rule in the success of many computer-vision applications, such as object recognition [1], intelligent surveillance [2], and robotic visual servoing (RVS) [3]. Estimation of the camera displacement (CD) between the current and desired pose for RVS [4], [5] is also relevant to this problem. However, the focus of this study will be on pose estimation for RVS where the relative pose between a camera and an object is used for real-time control of a robot motion [3].

In RVS, the control error can be calculated in the image space, Cartesian space, or both (hybrid) spaces [3], [6], [7]. While partial estimation of the pose vector (e.g., depth) is required for image-based and hybrid visual-servoing schemes [8], [9], an important class of visual-servoing methods, namely the position-based visual-servoing (PBVS) scheme, requires full pose estimation to calculate Cartesian error of the relative pose between the endpoint and the object [10]. Two major difficulties with pose estimation for RVS are related to the requirements for efficiency and robustness of pose estimation [11].

The solutions to pose-estimation problem usually focus on using sets of 2-D–3-D correspondences between geometric features and their projections on the image plane. Although high-level geometric features, such as lines and conics, have been proposed, point features are typically used for pose estimation due to their ease of availability

Manuscript received September 1, 2009; revised April 17, 2010; accepted July 20, 2010. Date of publication September 2, 2010; date of current version September 27, 2010. This work was supported by the Natural Sciences and Engineering Research Council of Canada under Grant 903060-07. The work of M. Marey was supported by a grant from the Egyptian Ministry of High Education and Scientific Research.

F. Janabi-Sharifi is with the Department of Mechanical and Industrial Engineering, Ryerson University, Toronto, ON M5B 2K3, Canada (e-mail: fsharifi@ryerson.ca).

M. Marey is with IRISA/INRIA Rennes Bretagne Atlantique, Campus de Beaulieu, Universitaire de Beaulieu, 35042 Rennes Cedex, France (e-mail: mohammed.marey@irisa.fr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TRO.2010.2061290

in many objects [12]. Solutions for three points [13], and more than three points [14] have already been presented. However, exact and closed-form solutions are only available for three or four noncollinear points [15]. Such methods, although simple to implement, are often exposed to difficulty in point matching in crowded environments. Besides, point-based solutions are not robust and demonstrate high susceptibility to noise in image coordinates [16]. For three-point solutions, it has been shown that the points configuration and noise in the points coordinates can drastically affect the output errors [13]. It has also been demonstrated that when the noise level exceeds a knee level or the number of points is below a knee level, least-squares-based methods, which are commonly used for points solutions, become unstable leading to large errors [17]. Addition of more points would enhance pose-estimation robustness with the cost of increased computational expense. Nonlinear, iterative, and/or recursive methods are then recommended for more than four points as well as high-level features.

The iterative approaches formulate the problem as a nonlinear least-squares problem. Such solutions offer more accuracy and robustness, yet they are computationally more intensive than closed-form approaches, and their accuracy depends on the quality of the initial pose estimates [18], [19]. The iterative methods usually rely on nonlinear optimization techniques, such as the Gauss–Newton method [1]. To reduce the problem complexity, approximate methods have also been proposed by simplifying the perspective camera model, e.g., relaxing the orthogonality constraint on the rotation matrix [19], [20]. The survey of both exact and approximate pose-estimation methods can be found in the literature [15], [21]. In short, this class of methods exhibits convergence problems and does not effectively account for the orthonormal structure of rotation matrices [22]. Furthermore, with this class of techniques, noisy visual-servo images usually lead to poor individual pose estimates [23], thus requiring temporal filtering.

A class of recursive methods relies on temporal-filtering methods, and in particular, Kalman-filtering techniques, to address robustness and efficiency issues. Since a 3-D pose and its time rate constitute a 12-D state vector to be estimated in real time, many of these filtering methods, such as particle filters [24], can hardly model the true distribution in real time. A true 3-D pose estimation using Kalman filter (KF) for RVS has been realized in [10]. With KFs, photogrammetric equations are formed by first mapping the object features into the camera frame and then projecting them onto the image plane. A KF is then applied to provide an implicit and recursive solution of the pose parameters. Since the filter output model for RVS is nonlinear in the system states, an extended KF (EKF) is usually applied, in which the output equations are linearized about the current state estimates. The use of a KF in RVS is motivated by its several advantages, including its recursive implementation, capability to statistically combine redundant information (such as features) or sensors, temporal filtering, possibility of using lower number of features, and the possibility for changing the measurement set without disrupting the operation [3], [10]. For instance, an EKF-based platform has been proposed in [11] to integrate range sensor with vision sensor for robust pose estimation in RVS. Additionally, an EKF implementation facilitates dynamic windowing of the features of interest by providing estimation of the next time-step feature location. This allows only small window areas to be processed for image-parameter measurements and leads to a significant reductions in image-processing time. It has been shown that, in practice, an EKF provides near-optimal estimation [10].

Despite its advantages, there are a few issues with the application of EKF to pose estimation in RVS. First, a known object model is usually assumed to be available. Model-free approaches based on Euclidean reconstruction have been proposed for CD estimation [4], [5]. These approaches typically rely on fundamental, essential, and/or homography matrix estimation, e.g., in [5] and [25] and, hence, face the issue of degeneration of the epipolar geometry in some cases, thus leading to unstable estimation [4]. Despite some treatments [4], they remain susceptible to outliers. In addition, majority of them require several images for reconstruction and, hence, are more appealing for postproduction applications [26]. The assumption of known object model is not a major issue in many industrial setups since computer-aided-design (CAD) models of the objects are usually available. For uncertain environments with a poor (or unknown) model of the object, an EKF-based approach for real-time estimation of combined target model and pose has been proposed in [27] and [28]. Therefore, this issue will not be the subject of our focus. Second, while a KF provides optimal solution under the assumption of zero-mean Gaussian noise for a linear problem, the EKF formulation may not provide optimal results. In fact, linearization can generate unstable filters when the assumption of local linearity is not met [29]. In the previous work, it has been recommended to take a sufficiently high sampling rate to enforce accuracy of the linearization over the sampling period [10]. However, in practice, RVS-system bandwidth would limit the sampling rate for the filter. As it has been shown in [30], an EKF-based system might easily diverge under fast and nonlinear trajectory dynamics, even with a relatively high sampling rate. Third, statistics of the measurement and dynamic noise are assumed to be known in advance and to remain constant. Poor measurement and dynamic models or poor noise estimates would degrade the system performance and might even lead to the filter divergence. In particular, while the measurement noise-covariance matrix can be tuned through experiments, dynamic covariance matrix is difficult to tune [23]. This is because dynamics of the object motion with respect to the camera cannot be accurately predicted in a dynamic environment. Fourth, the convergence of EKF depends on the choice of initial state estimate and tuning of filter parameters. In many RVS applications, such as assembly industry, initial pose of the object with respect to the camera can be readily approximated. Yet, sufficiently good pose estimates cannot be initially available in unstructured and uncertain environments. This paper will contribute by formulating an EKF method to address the last two aforementioned issues.

Several methods have been proposed in the literature to deal with varying statistics and poor filter initialization of EKF for RVS systems. An adaptive EKF (AEKF) with a fixed set of image features has been formulated for the first time in [30] to update the dynamic-noise-covariance matrix in order to address the issue of varying and/or uncertain dynamic noise. The AEKF-based approach has later been extended in [31] to have a variable set of image features during the servoing for improving servoing robustness. Despite the adaptation capability of AEKF to unknown noise statistics, the presented AEKF methods do not provide robust and accurate pose estimation in the presence of poor filter initialization and camera calibration, particularly when tracking a fast and nonlinear trajectory is desired. This aspect will be investigated experimentally in this paper. While tuning EKF noise-covariance matrices were addressed in the aforementioned AEKF-based approaches [30], [31], tuning and initialization of other EKF parameters and mechanisms to enhance output linearization for RVS did not receive much attention. To address tuning of other filter parameters and to facilitate its initialization, an initial proposal for iterative EKF (IEKF) use in RVS has been provided in [32]. As a matter of fact, Lefebvre *et al*. [33] have studied several modifications of KFs for general nonlinear systems. They have categorized all the different versions of KFs such as the central difference filter (CDF), unscented KF (UKF), and the divided difference filter (DD1) as linear regression KFs (LRKFs) and have compared them with EKF and IEKF [34]. They have concluded that EKF and IEKF generally outperform LRKFs, yet they require a careful tuning. An interesting result of their study is that IEKF outperforms EKF, because it uses the

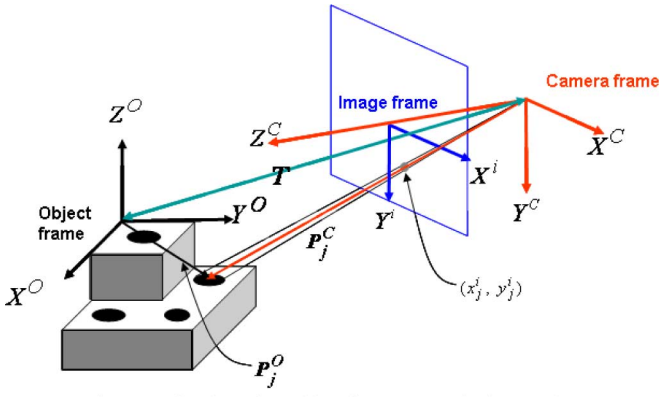

Fig. 1. Projection of an object feature onto the image plane.

measurements to linearize the measurement function, whereas in EKF and LRKFs, the measurement is not used for the same purpose. Despite its advantages, lack of adaptive noise estimation mechanism would degrade the performance of IEKF. In this paper, for the first time, an iterative AEKF (IAEKF) for RVS is proposed to overcome limitations of IEKF and AEKF. The presented work in this paper is continuation of the previous works on EKF for RVS [3], [11], [27], [28], [30], [32]. This study contributes by detailed formulation of IAEKF and experimental comparison of EKF, AEKF, IEKF, and IAEKF for RVS.

II. Feature-Point Transformation

The commonly used perspective projection model of the camera is shown in Fig. 1. Image frame is located at F (i.e., effective focal length) along the Z^C-axis with its X^i- and Y^i-axes parallel to the X^C- and Y^C-axes of the camera frame, respectively. In this study, similar to many iterative methods, point features will be used for pose estimation. Let the relative pose of the object to the camera (or end-effector) frame be $\boldsymbol{W} = (\boldsymbol{T}, \boldsymbol{\Theta})^T$, where $\boldsymbol{T} = [X, Y, Z]^T$ denotes the relative position vector of the object frame with respect to the camera frame, and $\boldsymbol{\Theta} = [\phi, \alpha, \psi]^T$ is the relative orientation vector with *roll*, *pitch*, and *yaw* parameters, respectively. Let $\boldsymbol{P}_j^C = (X_j^C, Y_j^C, Z_j^C)^T$ and $\boldsymbol{P}_j^o = (X_j^o, Y_j^o, Z_j^o)^T$ represent the coordinate vectors of the jth object feature point in the camera and object frames, respectively (see Fig. 1). The vector of $\boldsymbol{P}_j^o$ is available from the CAD model of the object or measurements and can be described in the camera frame using the following transformation:

$$\boldsymbol{P}_j^C = \boldsymbol{T} + \boldsymbol{R}(\phi, \alpha, \psi)\boldsymbol{P}_j^o \tag{1}$$

where the rotation matrix is given in [3] and [10]. For control error calculations, the Euler angles can be approximately related to the total angles in a PBVS structure using a transition matrix [10]. The coordinates of the projection of a feature point on the image plane using a pin-hole camera model will be x_j^i and y_j^i given by (see Fig. 1)

$$\begin{bmatrix} x_j^i & y_j^i \end{bmatrix}^T = \frac{F}{Z_j^C} \begin{bmatrix} \frac{X_j^C}{P_X} & \frac{Y_j^C}{P_Y} \end{bmatrix}^T \tag{2}$$

where P_X and P_Y are interpixel spacing in X^i- and Y^i-axes of the image plane, respectively. This model assumes that the origin of the image coordinates is located at the principal point, and $|Z_j^C| \gg F$. For short focal lengths, lens distortion can have a drastic effect on the feature-point locations. For details of distortion model and its relation to projection model, see [30] and [35]. The perspective projection model requires both intrinsic and extrinsic camera parameters. The camera intrinsic parameters (P_X, P_Y, F), coordinates of optical axis on the image plane (principal point) O^i, radial and tangential distortion parameters, and aspect ratio are all determined from camera-calibration tests [30]. The camera extrinsic parameters include the pose of the camera with respect to the end-effector or the robot base frame, which are calculated by inspection of camera housing and kinematic calibration [36]. Excellent solutions to the camera-calibration problem exist in the literature [37].

Substituting (1) into (2) results in two nonlinear equations with six unknown pose parameters of $\boldsymbol{W}$. Therefore, at least three noncollinear features are required for pose estimation (i.e., $p = 3$) [38]. However, to obtain a unique solution, at least four features will be needed. It has been shown that the inclusion of more than six features will not improve the performance of EKF estimation significantly [23]. In addition, the features need to be noncollinear and noncoplanar to provide good results. Therefore, in many RVS applications, $4 \leq p \leq 6$.

III. Extended Kalman Filter

For pose estimation, the state vector of dynamic model is defined to include pose and velocity parameters, i.e.,

$$\boldsymbol{x} = \left[X, \dot{X}, Y, \dot{Y}, Z, \dot{Z}, \phi, \dot{\phi}, \alpha, \dot{\alpha}, \psi, \dot{\psi}\right]^T. \tag{3}$$

The relative target velocity is usually assumed to be constant during each sample period. This is a reasonably valid assumption for sufficiently small sample periods in RVS systems. A discrete dynamic model will be then

$$\boldsymbol{x}_k = \boldsymbol{A}\boldsymbol{x}_{k-1} + \boldsymbol{\gamma}_k \tag{4}$$

with $\boldsymbol{A}$ being a block diagonal matrix with 2×2 blocks of the form $\begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix}$, T being the sample period, k being the sample step, and $\boldsymbol{\gamma}_k$ being the disturbance noise vector described by a zero-mean Gaussian distribution with covariance $\boldsymbol{Q}_k$, i.e.,

$$E[\gamma_i] = \boldsymbol{q}_i, E\left[(\boldsymbol{\gamma}_i - \boldsymbol{q}_i)(\boldsymbol{\gamma}_j - \boldsymbol{q}_j)^T\right] = \boldsymbol{Q}_i\boldsymbol{\delta}_{ij} \tag{5}$$

where $\boldsymbol{q}_i$ and $\boldsymbol{Q}_i$ are true mean and true moments about the mean of state noise sequences, respectively, and $\boldsymbol{\delta}$ is the Kronecker delta. The output model will be based on the projection model given by (1) and (2) and defines the image-feature locations in terms of the state vector $\boldsymbol{x}_k$ as follows:

$$\boldsymbol{z}_k = \boldsymbol{G}(\boldsymbol{x}_k) + \boldsymbol{\nu_k} \tag{6}$$

with measurements for p feature points

$$\boldsymbol{z}_k = [x_1^i, y_1^i, x_2^i, y_2^i, \ldots, x_p^i, y_p^i]_k^T \tag{7}$$

and

$$\boldsymbol{G}(\boldsymbol{x}_k) = F\left[\frac{X_1^C}{P_X Z_1^C}, \frac{Y_1^C}{P_Y Z_1^C}, \cdots, \frac{X_p^C}{P_X Z_p^C}, \frac{Y_p^C}{P_Y Z_p^C}\right]^T. \tag{8}$$

Here, X_j^C, Y_j^C, and Z_j^C are given by (1), and $\boldsymbol{\nu}_k$ denotes the image-parameter measurement noise that is assumed to be described by a zero-mean Gaussian distribution with covariance $\boldsymbol{R}_k$, i.e.,

$$E[\boldsymbol{\nu}_i] = \boldsymbol{r}_i, \qquad E\left[(\boldsymbol{\nu}_i - \boldsymbol{r}_i)(\boldsymbol{\nu}_i - \boldsymbol{r}_i)^T\right] = \boldsymbol{R}_i\delta_{ij} \tag{9}$$

where $\boldsymbol{r}_i$ and $\boldsymbol{R}_i$ are true mean and true moments about the mean of measurement-noise sequences, respectively. Since (6) is nonlinear, an optimal solution cannot be obtained through a KF implementation. Instead, an extension of KF (i.e., EKF) can be formulated by linearizing the output equation about the current state.

Let $\boldsymbol{x}_k$ be the state at step k, $\hat{\boldsymbol{x}}_{k,k-1}$ denote the *a priori* state estimate at step k given the knowledge of the process or measurement at the end of step $k-1$, and let $\hat{\boldsymbol{x}}_{k,k}$ be the *a posteriori* state estimate at step k given measurement $\boldsymbol{z}_k$. Then, *a priori* and *a posteriori* estimate errors, and their corresponding covariances are defined as $\boldsymbol{e}_k = \boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k,k}$, $\boldsymbol{P}_{k,k} = E[\boldsymbol{e}_k \boldsymbol{e}_k^T]$, $\boldsymbol{e}_{k,k-1} = \boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k,k-1}$, and $\boldsymbol{P}_{k,k-1} = E\left[\boldsymbol{e}_{k,k-1}\boldsymbol{e}_{k,k-1}^T\right]$, respectively. It is well known that the recursive EKF algorithm consists of two major parts of prediction and estimation as follows.

Prediction:

$$\hat{\boldsymbol{x}}_{k,k-1} = \boldsymbol{A}\hat{x}_{k-1,k-1} \tag{10}$$

$$\boldsymbol{P}_{k,k-1} = \boldsymbol{A}\boldsymbol{P}_{k-1,k-1}\boldsymbol{A}^T + \boldsymbol{Q}_{k-1}. \tag{11}$$

Linearization:

$$\boldsymbol{H}_k = \frac{\partial \boldsymbol{G}(\boldsymbol{x})}{\partial \boldsymbol{x}}\Big|_{\boldsymbol{x}=\hat{\boldsymbol{x}}_{k,k-1}}. \tag{12}$$

Kalman gain update:

$$\boldsymbol{K}_k = \boldsymbol{P}_{k,k-1}\boldsymbol{H}_k^T(\boldsymbol{R}_k + \boldsymbol{H}_k\boldsymbol{P}_{k,k-1}\boldsymbol{H}_k^T)^{-1}. \tag{13}$$

Estimation updates:

$$\hat{\boldsymbol{x}}_{k,k} = \hat{\boldsymbol{x}}_{k,k-1} + \boldsymbol{K}_k(\boldsymbol{z}_k - \boldsymbol{G}(\hat{\boldsymbol{x}}_{k,k-1}) \tag{14}$$

$$\boldsymbol{P}_{k,k} = \boldsymbol{P}_{k,k-1} - \boldsymbol{K}_k\boldsymbol{H}_k\boldsymbol{P}_{k,k-1}. \tag{15}$$

Here, $\boldsymbol{K}_k$ is the Kalman gain matrix at step k. The measurement- and process-noise covariances $\boldsymbol{Q}_k$ and $\boldsymbol{R}_k$ are usually assumed to be constant during servoing and obtained through tuning [10]. While $\boldsymbol{R}_k$ can be determined through the experiments [23], matrix $\boldsymbol{Q}_k$ is difficult to determine *a priori* due to unknown object's and/or camera's motions. In general, the aim of adaptive filtering in RVS is to estimate not only the state but the time-varying statistical parameters given by $\Upsilon_i = \{\boldsymbol{r}_i, \boldsymbol{R}_i, \boldsymbol{q}_i, \boldsymbol{Q}_i\}$ as well. An AEKF has been introduced in [30] and [31] to estimate $\boldsymbol{R}_k$ and $\boldsymbol{Q}_k$ in real time. The adaptation capability of AEKF with poor initialization of noise-covariance matrices has been demonstrated in our previous work [30]. However, the results also showed that in quicker changes of the pose, the error of AKEF will increase. This is mainly due to the time required by AEKF to react to such a sudden change. Linearization approximation in (12) cannot be treated by AEKF properly and is another source of errors, especially in tracking the trajectories with faster and higher dynamics. Besides, the linearization approximation errors would lead to high sensitivity to poor initialization and camera-calibration error. An IEKF has been proposed in our previous work [32] to alleviate this issue.

In the next section, adaptive and iterative mechanisms are combined to address the aforementioned issues simultaneously and to establish a robust framework for pose estimation in RVS.

IV. Iterative Adaptive Extended Kalman Filter

The proposed approach combines advantages of AEKF and IEKF. After initialization and prediction stages, iteration is started for m iterations by first setting $\hat{x}_k^0 = \hat{x}_{k,k-1}$, i.e., for $i = 0$, and then

$$\boldsymbol{H}_k^i = \frac{\partial \boldsymbol{G}(\boldsymbol{x})}{\partial \boldsymbol{x}}\Big|_{\boldsymbol{x}=\hat{\boldsymbol{x}}_k^i} \tag{16}$$

$$\hat{\boldsymbol{r}}_k^i \equiv \boldsymbol{z}_k - \boldsymbol{G}(\hat{\boldsymbol{x}}_k^i) \tag{17}$$

$$\boldsymbol{\Gamma}_k^i \equiv \boldsymbol{H}_k^i\boldsymbol{P}_{k,k-1}\boldsymbol{H}_k^{i^T} \tag{18}$$

$$\bar{\boldsymbol{r}}_k^i = \bar{\boldsymbol{r}}_{k-1} + \frac{1}{N}(\hat{\boldsymbol{r}}_k^i - \hat{\boldsymbol{r}}_{k-N}) \tag{19}$$

$$\begin{aligned}\boldsymbol{R}_k^i = \boldsymbol{R}_{k-1} + \frac{1}{N-1}\Big(&\left(\hat{\boldsymbol{r}}_k^i - \bar{\boldsymbol{r}}_k^i\right)\left(\hat{\boldsymbol{r}}_k^i - \bar{\boldsymbol{r}}_k^i\right)^T \\ &- \left(\hat{\boldsymbol{r}}_{k-N} - \bar{\boldsymbol{r}}_k^i\right)\left(\hat{\boldsymbol{r}}_{k-N} - \bar{\boldsymbol{r}}_k^i\right)^T \\ &+ \frac{1}{N}\left(\hat{\boldsymbol{r}}_k^i - \hat{\boldsymbol{r}}_{k-N}\right)\left(\hat{\boldsymbol{r}}_k^i - \hat{\boldsymbol{r}}_{k-N}\right)^T + \frac{N-1}{N}(\boldsymbol{\Gamma}_{k-N} - \boldsymbol{\Gamma}_k^i)\Big)\end{aligned} \tag{20}$$

$$\boldsymbol{K}_k^i = \boldsymbol{P}_{k,k-1}\boldsymbol{H}_k^{i^T}\left(\boldsymbol{R}_k^i + \boldsymbol{H}_k^i\boldsymbol{P}_{k,k-1}\boldsymbol{H}_k^{i^T}\right)^{-1} \tag{21}$$

$$\hat{\boldsymbol{x}}_k^{i+1} = \hat{\boldsymbol{x}}_k^i + \boldsymbol{K}_k^i(\boldsymbol{z}_k - \boldsymbol{G}(\hat{\boldsymbol{x}}_k^i). \tag{22}$$

At the end of iterations, the iteration output is propagated as follows:

$$\begin{aligned}\hat{\boldsymbol{x}}_{k,k} = \hat{\boldsymbol{x}}_k^m, \qquad \bar{\boldsymbol{r}}_k = \bar{\boldsymbol{r}}_k^m, \qquad \boldsymbol{R}_k = \boldsymbol{R}_k^m, \\ \boldsymbol{\Gamma}_k = \boldsymbol{\Gamma}_k^m, \qquad \boldsymbol{K}_k = \boldsymbol{K}_k^m\end{aligned} \tag{23}$$

and *a posteriori* error-covariance estimate is updated according to (15). Here, a window of past measurements of size N is selected for adaptation of $\boldsymbol{R}_k$. The observation noise sample $\hat{\boldsymbol{r}}_j$ is assumed to be representative of $\boldsymbol{\nu}_j$, and for $j = k - N \to k$ to be independent and identically distributed.

Finally, the state noise statistics are estimated adaptively as follows:

$$\hat{\boldsymbol{q}}_j = \hat{\boldsymbol{x}}_{j,j-1} - \boldsymbol{A}\hat{x}_{j-1,j-1} \tag{24}$$

and for $j = k - N \to k$, it is assumed to be independent and identically distributed. In addition, let

$$\boldsymbol{\Delta}_k \equiv \boldsymbol{A}\boldsymbol{P}_{k-1,k-1}\boldsymbol{A}^T - \boldsymbol{P}_{k,k}. \tag{25}$$

Then, the process-noise-covariance matrix will be updated according to

$$\bar{\boldsymbol{q}}_k = \bar{\boldsymbol{q}}_{k-1} + \frac{1}{N}(\hat{\boldsymbol{q}}_k - \hat{\boldsymbol{q}}_{k-N}) \tag{26}$$

$$\begin{aligned}\boldsymbol{Q}_k = \boldsymbol{Q}_{k-1} + \frac{1}{N-1}\Big(&(\hat{\boldsymbol{q}}_k - \bar{\boldsymbol{q}}_k)(\hat{\boldsymbol{q}}_k - \bar{\boldsymbol{q}}_k)^T \\ &- (\hat{\boldsymbol{q}}_{k-N} - \bar{\boldsymbol{q}}_k)(\hat{\boldsymbol{q}}_{k-N} - \bar{\boldsymbol{q}}_k)^T \\ &+ \frac{1}{N}(\hat{\boldsymbol{q}}_k - \hat{\boldsymbol{q}}_{k-N})(\hat{\boldsymbol{q}}_k - \hat{\boldsymbol{q}}_{k-N})^T + \frac{N-1}{N}(\boldsymbol{\Delta}_{k-N} - \boldsymbol{\Delta}_k)\Big)\end{aligned} \tag{27}$$

followed by predictions stage, which is represented by (10) and (11). However, it must be noted that the above algorithm is computationally intensive when compared with EKF, IEKF, and AEKF. In order to improve computing time, adaptation steps are performed outside the iterations. After initialization and prediction steps, the first limited filter algorithm [30] to estimate the measurement-noise statistics is applied to find $\boldsymbol{R}_k$ before the iterations (using (16)–(20) without index i). Next, the iteration is established for m cycles to obtain Kalman gain and estimation updates according to (16), (21), and (22). State noise statistics are estimated outside the iterations, according to (24)–(27). To ensure positive definiteness of $\boldsymbol{R}_k$ and $\boldsymbol{Q}_k$, the diagonal elements of covariance estimators are reset to their absolute values. In addition, a fading-memory approach is applied to give low weights to initial (i.e., less reliable) samples of length and growing weight to successive noise samples as follows [30]:

$$\varpi_k = (k-1)(k-2)\cdots(k-\eta)/k^\eta, \qquad \text{if} \quad k \geq \eta \tag{28}$$

with the property of $\lim_{k\to\infty}\varpi_k = 1$.

TABLE I
COMPUTATIONAL COST FOR POSE ESTIMATION WITH $p = 5, N = 10, m = 10(20, 30)$

Method	FLOPS	CPU Time (s)
EKF	29947	0.036
AEKF (LMF)	34415	0.039
IEKF	175435 (336723, 497984)	0.062 (0.1073, 0.1260)
IAEKF	179903 (341191, 502452)	0.078 (0.1115, 0.1285)

In our experiments with IEKF [32], the optimal number of iterations were found to be $m = 30$, where $m = 20$ also provided reasonably good results. It should also be noted that for computational efficiency, a fixed number of iterations is not necessary for pose-estimation tasks. The iteration can be stopped if the iterated state estimate is close to the previous value, i.e., $(\boldsymbol{K}_k^i(\boldsymbol{z}_k - \boldsymbol{G}(\hat{\boldsymbol{x}}_k^i))^T(\boldsymbol{K}_k^i(\boldsymbol{z}_k - \boldsymbol{G}(\hat{\boldsymbol{x}}_k^i)) < \tau$, where τ is a threshold that could be found from experiments.

Table I shows the computational costs for different EKF-based algorithms used for pose estimation. For flops calculation flop option of MATLAB 5.0 has been used. The CPU times were obtained using a P4 1.7-GHz PC with 256-MB RAM. Given the current technology of PCs, the increased computational costs with IEKF and IAEKF do not necessarily imply major disadvantages and a bottleneck as the total time of the filter computations with moderate number of iterations is much less than the time required for feature selection and image processing in RVS.

V. EXPERIMENTAL RESULTS

Extensive simulations and experiments were conducted to investigate and compare the performance of various Kalman-filtering approaches for pose estimation.

The default filter parameters were as follows: $\boldsymbol{R}_0$ is a diagonal matrix with diagonal elements of 0.01 (in pixels square) measured through the experiments, $\boldsymbol{P}_{0,0}$ is a block diagonal matrix with 2×2 blocks of the form diag[0.02 0.01] (in meters square, in (meters per second) square, in degrees, in (degrees per second) square), $N = 20$, $m = 30$, $\eta = 5$, and $p = 5$.

To evaluate the accuracy of estimation, the results of estimations were compared with the relative pose calculations through the robot forward kinematics using the joint encoders. Another measure of accuracy was the inspection of the Kalman-estimate output errors that are the errors between the true image-feature locations and those obtained from KF estimates, i.e., filter residues: $\boldsymbol{z}_k - \boldsymbol{G}(\hat{\boldsymbol{x}}_{k,k-1})$.

A 6-degree-of-freedom (DOF) Cartesian manipulator, i.e., AFMA-6, with an endpoint mounted AVT-MARLIN F-033C CCD camera (at IRISA-INRIA, Rennes) and a target object shown in Fig. 2 were used. The robot was calibrated and operated under Linux with visual-servoing software, i.e., VISP [39]. The camera images were sent at 50 fps (frames/s) to the host PC with Intel Core 2–2.93 GHz running under Linux on which frame grabbers had been installed. The images had the size of 128×182 pixels and had an effective focal length of $F = 12.5$ mm. The image processing and control computations were carried out on the host, and then, the control output was transmitted to the robot controller via a PCI-VME bus-adapter board. About 10 ms was required for control action. The camera parameters, namely image center and interpixel spacings, were obtained from the calibration program. Initial estimate of the pose was obtained using DeMenthon's method [19]. The sampling period was $T = 0.06325$ s.

The robot was commanded to travel through a predefined trajectory over a stationary object. The maximum velocity of the AFMA-6 endpoint was set through VISP. Therefore, for a given set of nodal points,

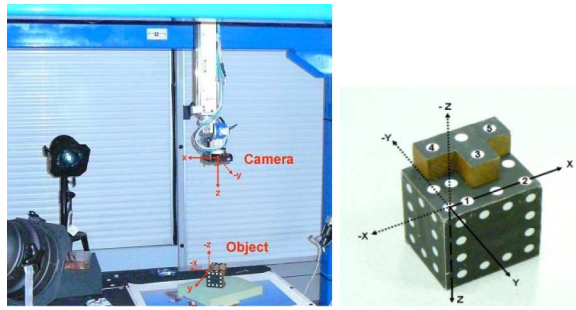

Fig. 2. (a) Experimental system for set 3 consisting of AFMA-6 manipulator and target object. (b) Image of the target object with its coordinate frame and features used.

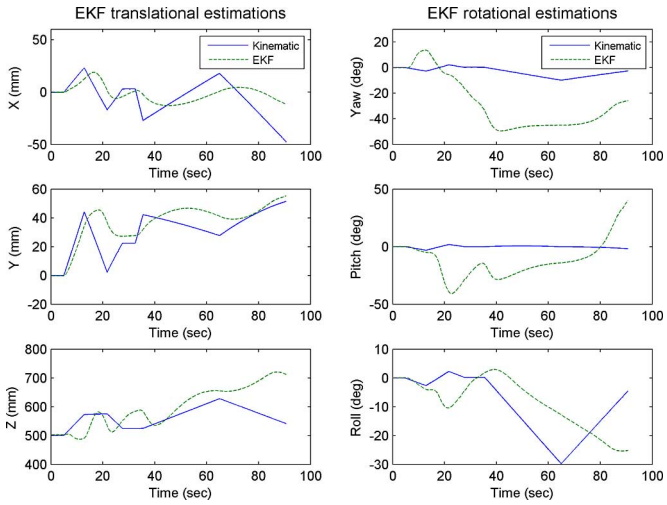

Fig. 3. Dynamic performance of pose estimation by EKF and forward kinematics estimators (experiment 1).

different trajectories with various dynamics were designed. A good estimation power of tuned EKF in relatively slow motion has already been shown [3], [10]. Therefore, EKF formed the comparison base.

The purpose of *experiment 1* was to compare the performance of various KF-based methods under an accurately calibrated robot framework. The maximum-velocity components of the endpoint trajectory were set to 50 mm/s and 5°/s, for translational and rotational coordinates, respectively, to generate a moderate motion dynamics. A null-state noise-covariance matrix was initially introduced to simulate the case of poorly tuned KF-based estimators for variety of trajectories. The endpoint relative trajectory was designed to incorporate sudden-velocity changes and significant nonlinearities. The purpose was to investigate the adaptation capability of AEKF and IAEKF to deviations from constant-velocity assumption of KF process model, and to evaluate the iterative performance of IEKF and IAEKF in approximating the output-model linearization. The inspection of the results (see Figs. 3–6 and Tables II and III) indicates that estimation accuracy of all algorithms is better in X, Y, and *roll* than in depth parameters Z, *pitch*, and *yaw*. The results also show that sudden changes in the velocity lead to divergence of EKF (see Fig. 3). This is due to the assumption of constant velocity in the state model. However, both AEKF and IAEKF were able to adapt to velocity changes (see Figs. 4 and 5). Fig. 5 shows that, although IEKF performance is superior to that of EKF, lack of a noise-adaptation mechanism in IEKF leads to significant errors and divergence toward the end of the relative pose trajectory. Table II shows pose-estimate-error statistics

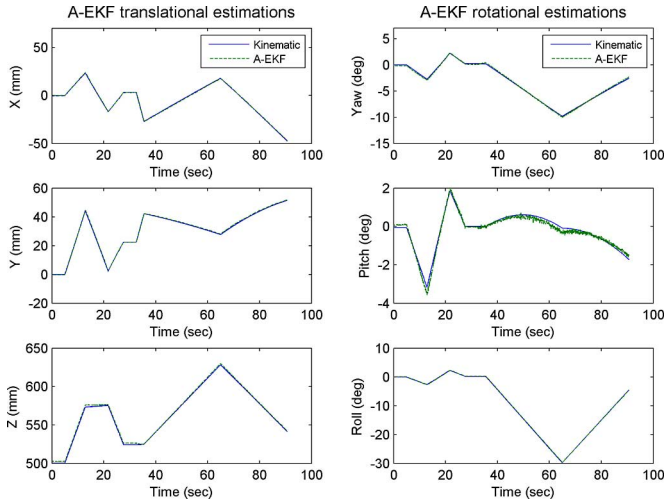

Fig. 4. Dynamic performance of pose estimation by AEKF and forward kinematics estimators (experiment 1).

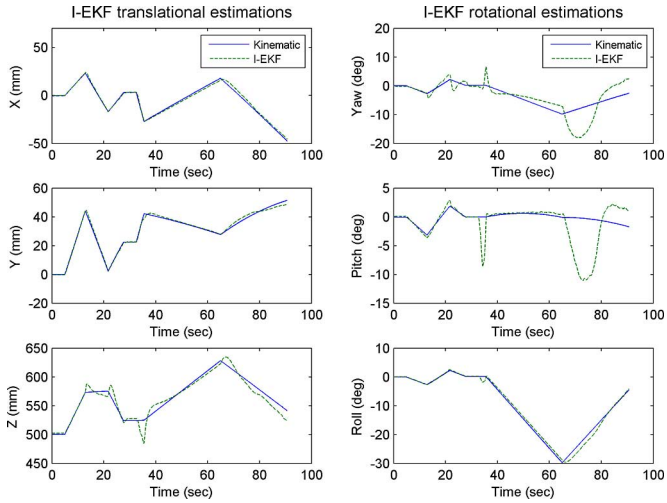

Fig. 5. Dynamic performance of pose estimation by IEKF and forward kinematics estimators (experiment 1).

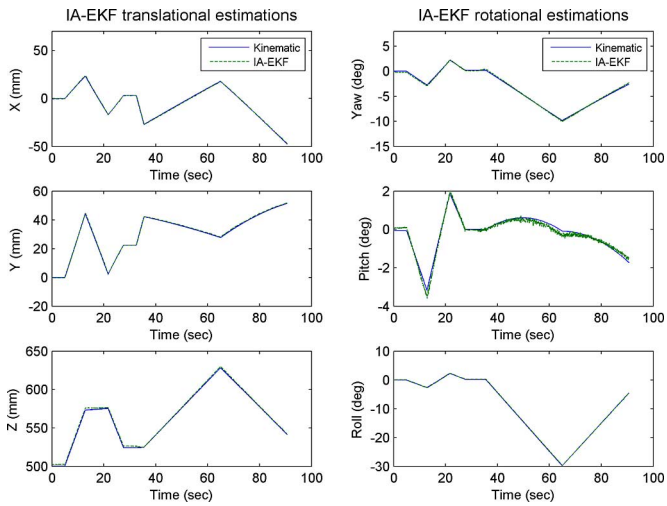

Fig. 6. Dynamic performance of pose estimation by IAEKF and forward kinematics estimators (experiment 1).

TABLE II
POSE ERROR STATISTICS FOR DIFFERENT KF-BASED ESTIMATORS WHEN COMPARED WITH KINEMATIC ESTIMATOR (EXPERIMENT 1)

Error		X (mm)	Y (mm)	Z (mm)	ϕ (deg)	α (deg)	ψ (deg)
EFK	*mean*	11.11	8.23	48.27	7.72	16.50	26.88
	std	8.47	7.53	43.55	5.63	11.42	14.75
	max	35.57	34.67	171.9	20.61	42.40	47.62
AEKF	*mean*	0.38	0.35	1.30	0.08	0.11	0.13
	std	0.42	0.31	0.70	0.08	0.13	0.14
	max	1.33	1.04	3.97	0.26	0.47	0.33
IEFK	*mean*	1.27	0.84	6.52	0.66	1.73	2.24
	std	0.92	0.69	6.06	0.65	2.93	2.57
	max	3.79	3.62	40.52	2.14	10.85	10.21
IAEKF	*mean*	0.38	0.35	1.30	0.08	0.11	0.12
	std	0.42	0.31	0.70	0.08	0.13	0.15
	max	1.32	1.04	3.97	0.26	0.47	0.33

TABLE III
IMAGE-PLANE-ERROR VARIANCE FOR DIFFERENT KF-BASED ESTIMATORS IN PIXELS SQUARE (EXPERIMENT 1)

	x_1^i	x_2^i	x_3^i	x_4^i	x_5^i
	y_1^i	y_2^i	y_3^i	y_4^i	y_5^i
EKF	1278.1	838.2	672.1	739.8	601.8
	923.90	1469.11	860.02	640.12	1205.11
AEKF	2.12	1.95	1.91	1.76	1.65
	1.31	1.16	1.32	1.35	1.33
IEKF	12.84	13.00	12.05	7.14	11.27
	4.25	5.93	11.37	3.60	8.77
IAEKF	1.35	1.26	1.21	1.13	1.07
	0.74	0.62	0.74	0.72	0.74

when different KF-based estimates are compared with kinematic estimates. Table III lists Kalman-estimate-output-error variances for five image-feature locations used in different KF-based methods. High levels of error can be observed for EKF estimates; however, both AEKF and IAEKF show good and comparable levels of accuracy, with IAEKF indicating slightly advantageous performance. Tracking accuracies of IAEKF for X and Y were approximately within ± 1.3 and ± 1 mm, respectively, and those for Z, *roll, yaw,* and *pitch* were within ± 4 mm, $\pm 0.3^\circ$, $\pm 0.5^\circ$, and $\pm 0.3^\circ$, respectively.

In *experiment 2*, the same condition, as in experiment 1, was used except that the magnitude of the maximum velocity of the robot endpoint was increased first ten times and next 27 times of the one used in the previous experiment, thereby resulting in experiments 2a and 2b, respectively (see Figs. 7 and 8). The resulting kinematic trajectories were almost the same as the trajectories in experiment 1, except that they were completed in shorter times. Consequently, faster dynamics and increased nonlinearities per sampling period would be expected. In both scenarios, EKF remained divergent with further degraded performance. The performance of AEKF was also degraded with the increased velocity. For instance, the mean errors of the AEKF estimator in experiment 2b along X- and Y-directions were approximately three and ten times more than those in experiment 1. Similarly, the standard deviation in the same directions increased 14 and 24 times in experiment 2b compared with experiment 1. This can be explained by the AEKF lag and its disability to keep up a good approximate for output linearization under faster and added nonlinear dynamics per sampling period. However, the performances of IEKF and IAEKF remained comparable with their performance in experiment 1.

In *experiment 3*, the same condition as experiment 1 was applied, but instead of null covariance matrices, tuned covariance matrices were used. The noise-covariance matrices were approximated using offline

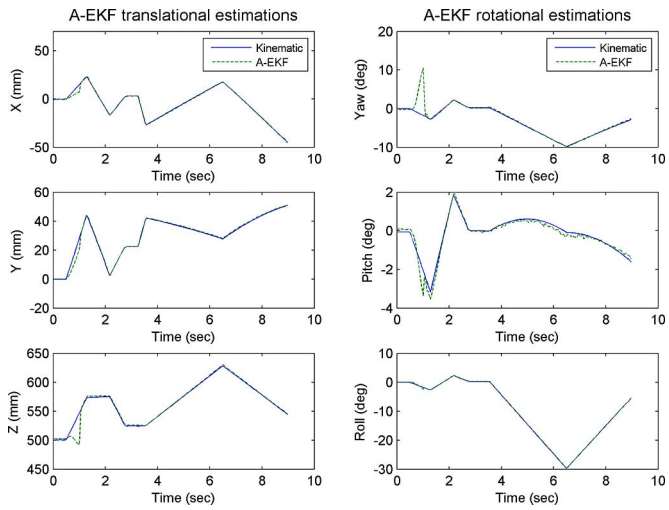

Fig. 7. Dynamic performance of pose estimation by AEKF and forward kinematics estimators (experiment 2a).

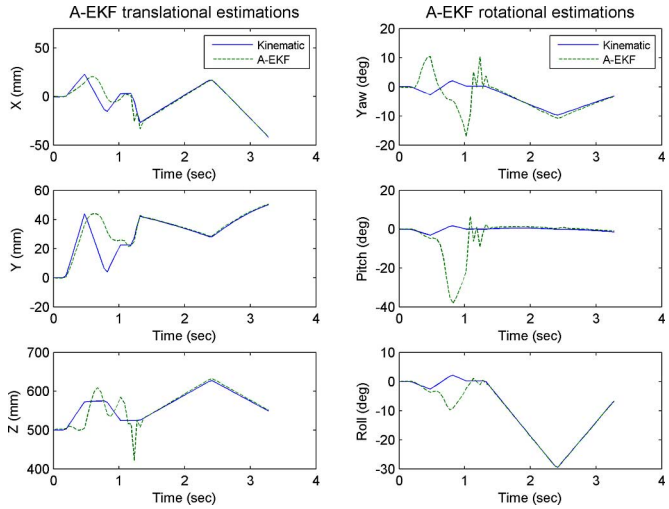

Fig. 8. Dynamic performance of pose estimation by AEKF and forward kinematics estimators (experiment 2b).

tuning with $q = 10^{-5}$ (in (meters per second) square, in (degrees per second) square) for

$$\boldsymbol{Q} = \mathrm{diag}[0, q, 0, q, 0, q, 0, q, 0, q, 0, q] \tag{29}$$

and $r = 0.01$ pixel^2 for

$$\boldsymbol{R} = \mathrm{diag}[r, r, r, r, r, r, r, r, r, r]. \tag{30}$$

The results are summarized in Tables IV and V. As it would be expected, carefully tuned covariance matrices in relatively moderate dynamic conditions enabled EKF to remain convergent. Compared with oscillating and divergent behavior of EKF in experiment 1, significant improvement was gained with tuned covariance matrices. AEKF also provides good results, which are superior to the EKF results in terms of mean error, standard deviation, and maximum error. The image-plane-error variance for EKF is not significantly different than that for AEKF (see Table V). Again, IAEKF provides the best results in terms of all comparison categories. It is also noted that with fine tuning of noise-covariance matrices, IEKF performance approaches that of IAEKF, as IEKF gives very similar results to those of IAEKF

TABLE IV
POSE-ERROR STATISTICS FOR DIFFERENT KF-BASED ESTIMATORS WHEN COMPARED WITH KINEMATIC ESTIMATOR (EXPERIMENT 3)

Error		X (mm)	Y (mm)	Z (mm)	ϕ (deg)	α (deg)	ψ (deg)
EFK	*mean*	0.46	0.46	1.64	0.22	0.30	0.32
	std	0.32	0.33	1.29	0.24	0.36	0.32
	max	2.12	1.91	6.65	1.17	1.58	1.24
AEKF	*mean*	0.39	0.38	1.36	0.12	0.19	0.19
	std	0.43	0.36	0.81	0.14	0.28	0.25
	max	1.51	1.28	3.96	0.48	1.07	0.76
IEFK	*mean*	0.38	0.35	1.20	0.08	0.11	0.12
	std	0.20	0.19	0.70	0.05	0.08	0.07
	max	1.32	1.05	3.90	0.27	0.44	0.30
IAEKF	*mean*	0.37	0.34	1.30	0.07	0.10	0.12
	std	0.42	0.31	0.70	0.08	0.12	0.14
	max	1.31	1.03	3.93	0.26	0.44	0.31

TABLE V
IMAGE-PLANE-ERROR VARIANCE FOR DIFFERENT KF-BASED ESTIMATORS IN PIXELS SQUARE (EXPERIMENT 3)

	x_1^i	x_2^i	x_3^i	x_4^i	x_5^i
	y_1^i	y_2^i	y_3^i	y_4^i	y_5^i
EKF	2.22	1.7	1.91	1.66	1.57
	1.56	0.92	0.94	1.50	1.09
AEKF	2.01	1.76	1.99	1.82	1.59
	1.46	1.10	1.22	1.38	1.25
IEKF	1.36	1.26	1.20	1.12	1.06
	0.74	0.62	0.73	0.73	0.74
IAEKF	1.35	1.25	1.20	1.12	1.06
	0.74	0.60	0.73	0.72	0.74

(see Tables IV and V). Interestingly, while the performance of EKF, AEKF, and IEKF improves with tuning of the noise-covariance matrices, the results of IAEKF remains approximately the same as those in experiment 1. This result again highlights the robustness of IAEKF to tuning errors of the measurement-noise-covariance matrices. The results were also obtained for various covariance matrices by varying q and r values according to $q \in \{10^3, 10, 10^{-1}, 10^{-3}, 10^{-5}, 10^{-20}\}$, and $r \in \{0.05, 0.1, 1, 10, 100, 1000\}$. The results again confirmed robustness of IAEKF to changes in $\boldsymbol{Q}$ and $\boldsymbol{R}$ matrices. The results of IEKF were also acceptable. However, AEKF, and particularly EKF, demonstrated high levels of sensitivity, as observed in [32]. For instance, mean-error values for IAEKF remained within $\pm 10\%$ of the values obtained with a null-process-noise-covariance matrix (see Table I).

In *experiment 4*, the sensitivity of KF-based estimators to the sampling rate was compared under dynamic conditions. The same condition as the previous experiment with the tuned measurement-noise-covariance matrices (with $q = 10^{-5}$ in (29), and $r = 10^{-2}$ in (30)) was applied, but the sampling time was changed to $T = 0.020$ s from the default value of 0.06325 s. Results consistent with the previous experiments [32] were obtained. The results for all estimators were improved with a higher sampling rate. However, IAEKF results were not significantly different than the results reported in Tables II and IV. For instance, mean-error values for IAEKF remained within $\pm 15\%$ of those reported in Table II. The results for IEKF were also relatively consistent, e.g., mean-error values remained within $\pm 20\%$ of the values reported in Table II. However, changes in EKF and AEKF results were more significant and, often, an order of magnitude different than those in Table II.

In *experiment 5*, the sensitivity of estimators to errors in initial poses was investigated by changing the initial positions 100, 200, 300, and

400 mm in all position coordinates. While IEKF and IAEKF provided acceptable results upto 200 mm deviation from the initial position, other methods failed at 100 mm deviation. The mean-error values for IAEKF and IEKF remained within $+10\%$ of those reported in Table II (i.e., almost-perfect pose initialization).

VI. Conclusion

Different KF-based methods of pose estimation have been discussed. A new pose-estimation method, namely the IAEKF algorithm, has also been introduced. All methods have been compared for their performance under different experimental conditions. It has been shown that mechanisms of noise adaptation and iterative-measurement linearization can be integrated within a novel IAEKF algorithm to obtain a superior performance in comparison with other KF-based methods. In particular, robustness of IAKEF has been established through experiments, and it has been demonstrated that IEAKF can improve pose-estimation performance in the presence of erroneous *a priori* statistics, nonlinear and fast-tracking trajectories and measurement function, slow sampling rates, and erroneous pose initialization. The improvements have been obtained at an additional computational cost, which are, in general, modest given the current PC technology and when compared with feature selection and image-processing time in RVS.

Acknowledgment

The authors would like to thank the Lagadic staff, particularly, F. Chaumette and F. Spindler, during his visit to INRIA-IRISA for useful discussions and their assistance with the experiments. The authors also acknowledge the assistance of A. Vakanski in simulations and efficiency calculations.

References

[1] D. G. Lowe, "Three-dimensional object recognition from single two-dimensional images," *Artif. Intell*, vol. 31, pp. 355–395, 1987.

[2] A. Mittal, L. Zhao, and L.S. Davis, "Human body pose estimation using silhouette shape analysis," in *Proc. IEEE Conf. Adv. Video Signal Based Surveill.*, Jul. 2003, pp. 263–270.

[3] F. Janabi-Sharifi, "Visual servoing: Theory and applications," in *Opto-Mechatronic Systems Handbook*, H. Cho, Ed. Boca Raton, FL: CRC, 2002, pp. 15-1–15-24.

[4] E. Malis and F. Chaumette, "2 1/2 D visual servoing with respect to unknown objects through a new estimation scheme of camera displacement," *Int. J. Comput. Vision*, vol. 37, no. 1, pp. 79–97, 2000.

[5] G. Chesi and K. Hashimoto, "A simple technique for improving camera displacement estimation in eye-in-hand visual servoing," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 9, pp. 1239–1242, Sep. 2004.

[6] F. Chaumette and S. Hutchinson, "Visual servo control, Part I: Basic approaches," *IEEE Robot. Autom. Mag.*, vol. 13, no. 4, pp. 82–90, Dec. 2006.

[7] F. Chaumette and S. Hutchinson, "Visual servo control, Part II: Advanced approaches," *IEEE Robot. Autom. Mag.*, vol. 14, no. 1, pp. 109–118, Mar. 2007.

[8] J. Feddema and O. R. Mitchell, "Vision-guided servoing with feature-based trajectory generation," *IEEE Trans. Robot. Autom.*, vol. 5, no. 5, pp. 691–700, Oct. 1989.

[9] E. Malis, F. Chaumette, and S. Boudet, "2-1/2 D visual servoing," *IEEE Trans. Robot. Autom.*, vol. 15, no. 2, pp. 234–246, Apr. 1999.

[10] W. J. Wilson, C. W. Hulls, and G. S. Bell, "Relative end-effector control using Cartesian position based visual servoing," *IEEE Trans. Robot. Autom.*, vol. 12, no. 5, pp. 684–696, Oct. 1996.

[11] W. J. Wilson, C. W. Hulls, and F. Janabi-Sharifi, "Robust image processing and position-based visual servoing," in *Robust Vision for Vision-Based Control of Motion*, M. Vincze and G. D. Hager, Eds. New York: IEEE Press, 2000, pp. 163–201.

[12] F. Janabi-Sharifi and W. J. Wilson, "Automatic selection of image features for visual servoing," *IEEE Trans. Robot. Autom.*, vol. 13, no. 6, pp. 890–903, Dec. 1997.

[13] R. M. Haralick, C. Lee, K. Ottenberg, and M. Nolle, "Review and analysis of solutions of the three point perspective pose estimation," *Int. J. Comput. Vis.*, vol. 12, no. 3, pp. 331–356, 1994.

[14] O. Faugeras, *Three-Dimensional Computer Vision*. Cambridge, MA: MIT Press, 1993.

[15] D. DeMenthon and L. S. Davis, "Exact and approximate solutions of the perspective-three point problem," *IEEE Trans. Pattern Analys. Mach. Intell.*, vol. 14, no. 11, pp. 1100–1105, Nov. 1992.

[16] X. Wang and G. Xu, "Camera parameters estimation and evaluation in active vision systems," *Pattern Recognit.*, vol. 29, no. 3, pp. 439–447, 1996.

[17] R. M. Haralick, H. Joo, C. Lee, X. Zhang, V. Vaidya, and M. Kim, "Pose estimation from corresponding point data," *IEEE Trans. Syst., Man, Cybern.*, vol. 19, no. 6, pp. 1426–1446, Nov./Dec. 1989.

[18] Q. Ji, M. S. Costa, R. M. Haralick, and L. G. Shapiro, "An integrated linear technique for pose estimation from different geometric features," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 13, no. 5, pp. 705–733, 1999.

[19] D. DeMenthon and L. S. Davis, "Model-based object pose in 25 lines of code," in *Proc. Eur. Conf. Comput. Vis.*, Santa Margherita Ligure, Italy, 1992, pp. 335–343.

[20] R. K. Lenz and R.Y. Tsai, "Techniques for calibration of the scale factor and image center for high accuracy 3D machine vision metrology," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 10, no. 5, pp. 713–720, Sep. 1988.

[21] R. Kumar and A. R. Hanson, "Robust methods for estimating pose and a sensitivity analysis," *CVGIP: Image Understanding*, vol. 60, pp. 313–342, 1994.

[22] C.-P. Lu, G. D. Hager, and E. Mjolsness, "Fast and globally convergent pose estimation from video images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 6, pp. 610–622, Jun. 2000.

[23] J. Wang and W. J. Wilson, "3D relative position and orientation estimation using Kalman filtering for robot control," in *Proc. IEEE Int. Conf. Robot. Autom.*, Nice, France, 1992, pp. 2638–2645.

[24] J. Carpenter, P. Clifford, and P. Fearnhead, "Improved particle filter for nonlinear problems," *Inst. Electr. Eng. Proc. Radar Sonar Navig.*, vol. 146, no. 1, pp. 2–7, 1999.

[25] Q.-T. Luong and O. Faugeras, "The fundamental matrix: Theory, algorithms, and stability analysis," *Int. J. Comput. Vis.*, vol. 17, no. 1, pp. 43–75, 1996.

[26] É. Marchand and F. Chaumette, "Virtual visual servoing: A framework for real-time augmented reality," *EUROGRAPHICS*, vol. 21, no. 3, pp. 289–298, 2002.

[27] L. Deng, W. J. Wilson, and F. Janabi-Sharifi, "Combined target model estimation and position-based visual servoing," in *Proc. IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, Sendai, Japan, Oct. 2004, pp. 1395–1400.

[28] L. Deng, W. J. Wilson, and F. Janabi-Sharifi, "Decoupled EKF for simultaneous target model and relative pose estimation using feature points," in *Proc. IEEE Int. Conf. Control Appl.*, Toronto, ON, Canada, Aug. 2005, pp. 749–754.

[29] S. J. Julier and J. K. Uhlmann, "Unscented filtering and nonlinear estimation," *Proc. IEEE*, vol. 92, no. 3, pp. 401–422, Mar. 2004.

[30] M. Ficocelli and F. Janabi-Sharifi, "Adaptive filtering for pose estimation in visual servoing," in *Proc. IEEE/RSJ Int. Conf. Intel. Robot. Syst.*, Maui, HI, 2001, pp. 19–24.

[31] V. Lippiello, B. Siciliano, and L. Villani, "Adaptive extended Kalman filtering for visual motion estimation of 3D objects," *Control Eng Pract.*, vol. 15, pp. 123–134, 2007.

[32] A. Shademan and F. Janabi-Sharifi, "Sensitivity analysis of EKF and Iterated EKF for position-based visual servoing," in *Proc. IEEE Int. Conf. Control Appl.*, Toronto, ON, Canada, Aug.2005, pp. 755–760.

[33] T. Lefebvre, H. Bruyninckx, and J. De Schutter, "Kalman filters for nonlinear systems: A comparison of performance," *Intl. J. Control*, vol. 77, no. 7, pp. 639–653, 2004.

[34] Y. Bar-Shalom and X. R. Li, *Estimation and Tracking: Principles, Techniques, and Software*. Boston, MA: Artech House, 1993.

[35] M. Ficocelli, "Camera calibration: Intrinsic parameters," Robot. Manuf. Autom. Lab., Ryerson Univ., Toronto, ON, Canada, Tech. Rep. TR-1999-12-17-01, 1999.

[36] P. I. Corke, *Visual Control of Robots: High Performance Visual Servoing*. Somerset, U.K.: Res. Studies, 1999.

[37] R. Tsai and R. Lenz, "A new technique for fully autonomous and efficient 3D robotic hand/eye calibration," *IEEE Trans. Robot. Autom.*, vol. 5, no. 3, pp. 345–358, Jun. 1989.

[38] J. S.-C. Yuan, "A general photogrammetric method for determining object position and orientation," *IEEE Trans. Robot. Autom.*, vol. 5, no. 2, pp. 129–142, Apr. 1989.
[39] É. Marchand, F. Spindler, and F. Chaumette, "VISP for visual servoing: A generic software platform with a wide class of robot control skills," *IEEE Robot. Autom. Mag.*, vol. 12, no. 4, pp. 40–52, Dec. 2005.

Autonomous Behavior-Based Switched Top-Down and Bottom-Up Visual Attention for Mobile Robots

Tingting Xu, *Student Member, IEEE*, Kolja Kühnlenz, *Member, IEEE*, and Martin Buss, *Member, IEEE*

***Abstract*—In this paper, autonomous switching between two basic attention selection mechanisms, i.e., top-down and bottom-up, is proposed. This approach fills a gap in object search using conventional top-down biased bottom-up attention selection, which fails, if a group of objects is searched whose appearances cannot be uniquely described by low-level features used in bottom-up computational models. Three internal robot states, such as observing, operating, and exploring, are included to determine the visual selection behavior. A vision-guided mobile robot equipped with an active stereo camera is used to demonstrate our strategy and evaluate the performance experimentally. This approach facilitates adaptations of visual behavior to different internal robot states and benefits further development toward cognitive visual perception in the robotics domain.**

***Index Terms*—Vision-guided robotics, visual attention control.**

I. Introduction

To achieve efficient processing of visual information about the environment, humans select their focus of attention (FOA), such that the most interesting regions will be processed first in detail. Studies about human visual perception show that visual attention selection is affected by two distinct mechanisms: top-down and bottom-up. Top-down signals are derived from the task specification or the previous knowledge and highlight the task-relevant information. It is goal-directed and essential for task accomplishment. In contrast, bottom-up attention is driven by distinct stimuli based on primary visual features. Interaction and coordination of both enable gaze-fixation-point selection and guide the visual behavior. To deal with the limited processing capability of most technical systems, especially autonomous mobile robots, a biologically plausible and technically applicable visual attention system is to be developed in order to fill the gap between the fundamental studies and the robotics research.

Fig. 1. ACE robot.

Normally, when operating in the real world, a robot has a task such as detecting and manipulating a target object. For a mobile robot, a typical task is to find a target and move toward it. In a simple scenario with unique target objects, a conventional top-down biased bottom-up strategy can help a lot in terms of efficiency [1]. However, it fails if a group of objects is searched whose appearances cannot be uniquely described by low-level features used in a primary bottom-up computation model. For example, different traffic signs are all salient in color but different in geometry and have different patterns on them. They are, therefore, not distinguishable from each other and only rely on low-level features used in bottom-up attention selection. An exhaustive search is still needed. To lower the computational cost, a search window is usually defined for exhaustive search as the robot FOA, in which the exhaustive search is conducted.

A search window based on bottom-up attention can predict image regions with higher probability to contain a target object, while a search window based on top-down attention is efficient for task accomplishment. Both bottom-up attention and top-down attention are essential for robot-attention control. On the one hand, if a task-relevant object is not located in the robot field of view (FOV), pure top-down attention selection can also use position data in the 3-D task space to direct robot attention toward the target, while bottom-up or top-down biased bottom-up attention selection only relies on the 2-D image data. On the other hand, if there is no task-relevant information in the FOV at all, pure bottom-up attention can guide the robot attention to explore the environment in a flexible way. In this paper, autonomous switching between top-down and bottom-up attention mechanisms is proposed, which enables autonomy of robots in terms of adaptations of visual behavior to different internal robot states and which fills the gap for object search not solvable using conventional combination of them. A vision-guided mobile robot, which is the Autonomous City Explorer (ACE) [2] developed at our institute (see Fig. 1), is used to demonstrate our strategy and evaluate the performance experimentally. It is equipped with an activevision system, which consists of a Bumblebee XB3 stereo camera from Point Grey Research, Inc., and a high-performance pan-tilt platform [3].

This paper is organized as follows: In Section II, related works about combination of top-down and bottom-up attention selections are introduced. In Section III, the proposed autonomous switching

Manuscript received January 25, 2010; revised May 31, 2010; accepted July 25, 2010. Date of publication August 26, 2010; date of current version September 27, 2010. This paper was recommended for publication by Associate Editor T. Kanda and Editor G. Oriolo upon evaluation of the reviewers' comments. This work was supported in part by the German Research Foundation (DFG) Excellence Initiative Research Cluster *Cognition for Technical Systems* (CoTeSys) (www.cotesys.org) and in part by the Institute for Advanced Study, Technische Universität München (www.tum-ias.de).

T. Xu and M. Buss are with the Insititute of Automatic Control Engineering, Technische Universität München, Munich 80290, Germany (e-mail: tingting.xu@ieee.org; m.buss@ieee.org).

K. Kühnlenz is with the Insititue of Automatic Control Engineering, Technische Universität München, Munich 80290, Germany, and also with the Institute for Advanced Study, Technische Universität München, Munich 80333 Germany (e-mail: k.kuehnlenz@ieee.org).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TRO.2010.2062571