Bounded Recursive Optimization Approach for Pose Estimation in Robotic Visual Servoing *

Yuchen Zhang[1], Bo Chen[1,2], Li Yu[1] and Haiyu Song[3]

[1] Department of Automation, Zhejiang University of Technology, Hangzhou 310023, P.R.China.
[2] Institute of Cyberspace Security, Zhejiang University of Technology, Hangzhou 310023, P.R.China.
[3] College of Information, Zhejiang University of Finance & Economics, Hangzhou 310018, P.R.China.
Email: bchen@zjut.edu.cn; lyu@zjut.edu.cn

Abstract. Pose estimation problem is concerned with determining position and orientation of an object in real time using the image information, and has found applications in many fields such as object recognition and robotic visual servoing. Most of vision-based pose estimation schemes are derived from extended Kalman filter, which requires that the noises obey the Gaussian distribution under known covariance. However, the statistical information in robot control may not be accurately obtained or satisfied. In this paper, a novel bounded recursive optimization approach is proposed to solve the pose estimation problem in visual servoing, where the addressed noises do not provide any statistical information, and the bounds of noises are also unknown. Finally, the pose estimation simulation is conducted to show the advantages and effectiveness of the proposed approach.

Keywords: Pose Estimation, Visual Servoing, Bounded Recursive Optimization Approach

1 Introduction

As one of the important issues in robotic visual servoing (RVS), pose estimation has attracted considerable research interest during the past few decades. The major concern of pose estimation in RVS is to determine the position and orientation of an object for real-time control of robot motion. Most solutions to pose estimation problem rely on sets of 2-D-3-D correspondences between geometric features and their projections on the image plane. Particularly, point features are typically used for pose estimation due to their ease of availability in many objects [1].

Recently, there are three major methods for pose estimation have been investigated extensively, i.e., camera-calibration based methods, iterative methods and

* The work was supported by the National Natural Science Foundation of China under Grant 61673351 and 61603331.

extended Kalman filter (EKF) based methods. The camera-calibration based methods rely on the geometric relationship between noncollinear feature points and the corresponding feature points on grid body plane. Solutions for three points and more than three points have already been presented [2, 3]. However, the related results will be drastically affected by the points configuration and noise in the points coordinates [2]. Iterative methods formulate the pose estimation problem as a nonlinear least-squares problem, and their solutions rely on nonlinear optimization techniques, such as Gauss-Newton method [4]. The major problem for this class of methods is their convergence difficulty to guarantee [5]. EKF is a kind of recursive method and can provide near-optimal estimation for pose parameters. EKF-base platform has been proposed in reference [6] to integrate range sensor with vision sensor for robust pose estimation in RVS. However, when the assumption of local linearity is not satisfied, the linearization of measurement equations in EKF can generate unstable filters. Moreover, the statistical information of dynamic and measurement noises in EKF are always assumed to be known in advance and to remain constant, which is not possible for pose estimation problem in practice.

To overcome the drawback of EKF on noise assumptions, several methods have been proposed in the literature. Combined with the idea of adaptive filter, an adaptive EKF was proposed in reference [7] to update the dynamic noise covariance in pose estimation. This approach was then extended to iterative adaptive EKF [8] by integrating mechanisms for noise adaptation and iterative-measurement linearization. However, both adaptive EKF and iterative adaptive EKF require the assumption of Gaussian noises and the linearization error remains a problem.

Without Gaussian assumption of noise, the energy-bounded noises [9] and bounded noises [10,11] have been considered in filter designing. It should be pointed out that the bounded noises, which do not need any statistical information and bounds of noises, can be easily satisfied in practical RVS systems. Inspired by the idea of bounded recursive optimization (BRO) in reference [10, 11], the pose estimation problem under bounded noises is converted into a recursive convex optimization problem in this paper that can be easily solved by standard software packages.

2 Problem Formulation

Consider the problem of estimating position and orientation of an object in robotic visual servoing, the projection model [8] of feature points is shown in Fig.1. $O^C-(X^C,Y^C,Z^C)$, $O^I-(X^I,Y^I,Z^I)$ and $O^O-(X^O,Y^O,Z^O)$ are coordinate systems of camera frame, image frame and object frame, respectively. The X^I-axis and Y^I-axis of image frame are parallel to that of camera frame, and O^I is located F (i.e., effective focal length) from Z^C-axis. It is considered that the camera frame is fixed, while position parameters $T \triangleq [X,Y,Z]^T$ and orientation parameters $\Theta \triangleq [\alpha,\beta,\gamma]^T$ of an object in camera frame need to be

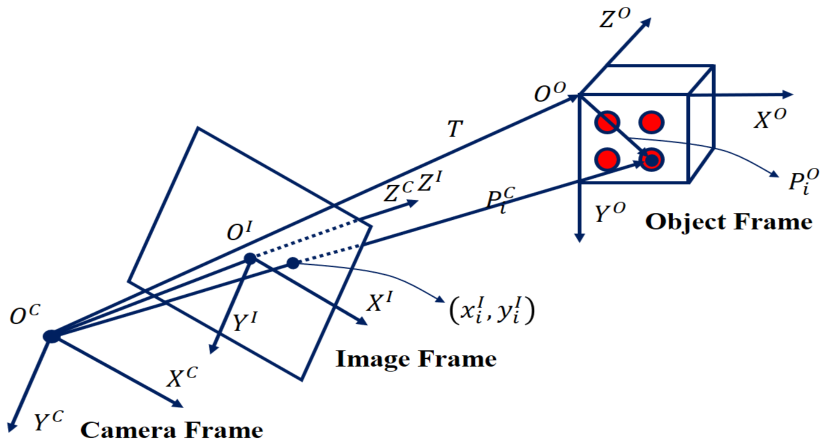

Fig. 1. Perspective projection model of camera.

estimated. For pose estimation, both pose and velocity parameters are defined to be state vector, i.e., $x = [X, Y, Z, \dot{X}, \dot{Y}, \dot{Z}, \alpha, \beta, \gamma, \dot{\alpha}, \dot{\beta}, \dot{\gamma}]^T$. The pose estimation system model is given by

$$x_k = Ax_{k-1} + w_k \tag{1}$$

$$z_k = G(x_k) + v_k \tag{2}$$

where k is the sample step, w_k and v_k are bounded noises, i.e.,

$$w_k^T w_k \leq \delta_w, \ v_k^T v_k \leq \delta_v \tag{3}$$

where δ_w and δ_v are unknown.

It is a reasonable assumption that the relative target velocity is usually to be constant during each sample period, thus the state transition matrix is given by

$$A = \text{mdiag}\left\{F_A, F_A\right\}, F_A = \begin{bmatrix} I_3 & tI_3 \\ 0 & I_3 \end{bmatrix} \tag{4}$$

where t is the sample period. The measurement $z_k \triangleq [x_{1,k}^I, y_{1,k}^I, ..., x_{p,k}^I, y_{p,k}^I]^T$ is a nonlinear function of state x_k, where p is the number of feature points and $[x_{i,k}^I, y_{i,k}^I]^T$ represents the coordinate of feature point in image frame at time k. To describe the relationship between z_k and x_k, the following coordinate transformation [12] and pin-hole camera model [13] will be introduced by

$$P_i^C = T + R(\Theta)P_i^O \tag{5}$$

$$[x_{i,k}^I, y_{i,k}^I]^T = \frac{F}{Z_i^C}\left[\frac{X_i^C}{P_X} \quad \frac{Y_i^C}{P_Y}\right]^T \tag{6}$$

where $P_i^O \triangleq [X_i^O, Y_i^O, Z_i^O]^T$ and $P_i^C \triangleq [X_i^C, Y_i^C, Z_i^C]^T$ are the coordinate vectors of the i-th feature point in object frame and camera frame, respectively. P_i^O is modeled or measured in advance. The rotation matrix $R(\Theta)$ is given by

$$R(\Theta) = \begin{bmatrix} \cos\beta\cos\gamma & \cos\beta\sin\gamma & -\sin\beta \\ \sin\alpha\sin\beta\cos\gamma & \sin\alpha\sin\beta\sin\gamma & \sin\alpha\cos\beta \\ \cos\alpha\sin\beta\cos\gamma & \cos\alpha\sin\beta\sin\gamma & \cos\alpha\cos\beta \end{bmatrix} \tag{7}$$

P_X and P_Y are interpixel spacing in X^I-axes and Y^I-axes of the image plane, respectively. The parameters (P_X, P_Y, F) of pin-hole camera model are all determined from camera-calibration tests [13].

Consequently, based on nonlinear measurement z_k, the aim of this paper is to design a BRO approach for pose estimation such that the statistical information of noise are not required.

Remark 1. In terms of pose estimation problem, the noise statistics for EKF [6] is not possible to be known in practice since the exact statistical information of noises will vary with time. In this sense, the proposed BRO approach, which does not require any noise statistics or bounds of noises, is more reasonable in practical applications.

3 Main Results

The estimation of pose at time k is calculated by the following recursive form:

$$\hat{x}_k^- = A\hat{x}_{k-1} \tag{8}$$

$$\hat{x}_k = \hat{x}_k^- + K_k(z_k - G(\hat{x}_k^-)) \tag{9}$$

where an optimal time-varying estimator gain K_k will be presented in Theorem 1 by minimizing an upper bound of estimation error square.

Theorem 1 For a given $\eta > 0$, an optimal estimator gain K_k for pose estimation can be obtained by solving the following convex optimization problem:

$$\begin{aligned} &\min_{\vartheta_k>0, \Phi_k>0, P_k>0} \mathrm{Tr}\{P_k + \Phi_k\} \\ &\text{s.t.}: \begin{cases} \begin{bmatrix} -I & G_{L,k}A & B_{L,k} \\ * & -P_k & 0 \\ * & * & -\Phi_k \end{bmatrix} < 0 \\ P_k - \vartheta_k I < 0 \\ \vartheta_k < \eta \end{cases} \end{aligned} \tag{10}$$

where $G_{L,k} \triangleq I - K_k H_k$, $B_{L,k} \triangleq [G_{L,k} \ \ -K_k]$ and the Jacobian matrix H_k is given by

$$H_k \triangleq \frac{\partial G(x)}{\partial x}|_{x=\hat{x}_k^-} \tag{11}$$

Proof. Define $\tilde{x}_k \triangleq x_k - \hat{x}_k$ and $\tilde{x}_k^- \triangleq x_k - \hat{x}_k^-$, one has that

$$\tilde{x}_k = \tilde{x}_k^- - K_k[G(x_k) - G(\hat{x}_k^-) + v_k] \tag{12}$$

Notice that $G(x_k)$ can be represented as $G(x_k) = G(\hat{x}_k^-) + H_k\tilde{x}_k^- + \Delta([\tilde{x}_k^-]^2)$ by Taylor series expanding about "$\hat{x}_k^-$", where $\Delta([\tilde{x}_k^-]^2)$ represents the high-order terms of the Taylor series expansion. $\Delta([\tilde{x}_k^-]^2)$ is an unknown noise, and the term $\tilde{v}_k$ is introduced to model the affection factors caused by this unknown noise [10, 11]. Thus, by defining $\xi_k \triangleq \mathrm{col}\{w_{k-1}, \tilde{v}_k\}$, the nonlinear error system is equivalent to:

$$\tilde{x}_k = G_{L,k}A\tilde{x}_{k-1} + B_{L,k}\xi_k \tag{13}$$

To construct an upper bound of estimation error square, the following performance index is introduced:

$$J_k \triangleq \tilde{x}_k^T\tilde{x}_k - \tilde{x}_{k-1}^T P_k \tilde{x}_{k-1} - \xi_k^T \Phi_k \xi_k \tag{14}$$

where $P_k > 0$ and $\Phi_k > 0$, then it follows from (13) that

$$J_k = \begin{bmatrix} \tilde{x}_{k-1} \\ \xi_k \end{bmatrix}^T \begin{bmatrix} A^T G_{L,k}^T G_{L,k} A - P_k & A^T G_{L,k}^T B_{L,k} \\ * & B_{L,k}^T B_{L,k} - \Phi_k \xi_k \end{bmatrix} \begin{bmatrix} \tilde{x}_{k-1} \\ \xi_k \end{bmatrix} \tag{15}$$

The condition $J_k < 0$ must be satisfied to make the term $\tilde{x}_{k-1}^T P_k \tilde{x}_{k-1} + \xi_k^T \Phi_k \xi_k$ an upper bound of estimation error square. According to Schur complement lemma, $J_k < 0$ is equivalent to the first inequality in (10). Moreover, the optimization objective "$\mathrm{Tr}\{P_k + \Phi_k\}$" is selected when minimizing this upper bound at time k.

Based on Theorem 1, the computation procedure of BRO approach for pose estimation is summarized as Algorithm 1.

Algorithm 1

1: Initialization: $\hat{x}_0$;
2: **for** $k := 1, 2, ...$ **do**
3: Calculate state estimation time update $\hat{x}_k^-$ by eq. (8);
4: Determine the Jacobian matrix "H_k" by eq. (11);
5: Calculate the optimal gain K_k by solving the optimization problem (10);
6: Determine $\hat{x}_k$ by eq. (9)
7: **end for**

Remark 2. The convex optimization problem (10) is established in terms of linear matrix inequalities (LMIs), and thus it can be directly solved by the function *"mincx"* of MATLAB LMI Toolbox [14]. On the other hand, the proposed BRO approach does not need the initial estimation error covariance matrix, and only depends on the initial estimated value.

4 Simulation Results

In this section, the simulation is conducted to study the performance of proposed BRO approach for pose estimation and the results are compared with EKF. The camera parameters are given by $\frac{F}{P_X} = 816.96$ and $\frac{F}{P_Y} = 811.69$, and the sample time is taken as $t = 0.05s$. There are four feature points used for pose estimation, and their coordinates in object frame are $(-12.5; -46; 0)$, $(12.5; -46; 0)$, $(12.5; -21; 0)$ and $(-12.5; -21; 0)$.

The moving object is assumed to travel through a predefined trajectory, thus the relative target velocity is changed once in a while. When the object in a relative slow motion, a well tuned EKF has shown good estimation performance [12]. In this case, the results of EKF in pose estimation is used for comparison.

Case 1: The dynamic and measurement noises are set as zero-mean Gaussian white noises with their covariance $Q_w = \text{diag}[0, q_1, 0, q_1, 0, q_1, 0, q_2, 0, q_2, 0, q_2]$ and $Q_v = \text{diag}[r_1, r_1, r_1, r_1, r_1, r_1, r_1, r_1, r_1, r_1, r_1, r_1, r_1, r_1, r_1, r_1]$, where $q_1 = 0.01$, $q_2 = 0.0001$, $r_1 = 0.01$. The purpose of the simulation in Case 1 is to show that the BRO approach can achieve good estimation performance without any statistical information of noises. It is assumed that EKF knows the covariance of noises and the well tunned results are plotted in Fig.2, while the BRO approach does not use any statistical information of noises and the pose estimation results are shown in Fig.3. Compared with EKF, the estimated position and orientation by BRO can better track the object as far as the relative velocity changed. The absolute value of estimation error by EKF and BRO under Gaussian noises is plotted in Fig.4, which shows the effectiveness of BRO approach without any statistical information of noises. Moreover, it is seen from Fig.4 that the pose estimation performance of BRO approach is better than that of EKF method in most cases.

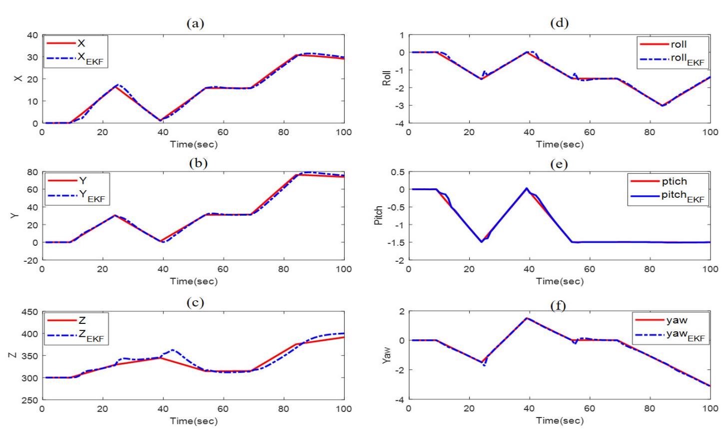

Fig. 2. Dynamic performance of pose estimation by EKF with Gaussian noises.

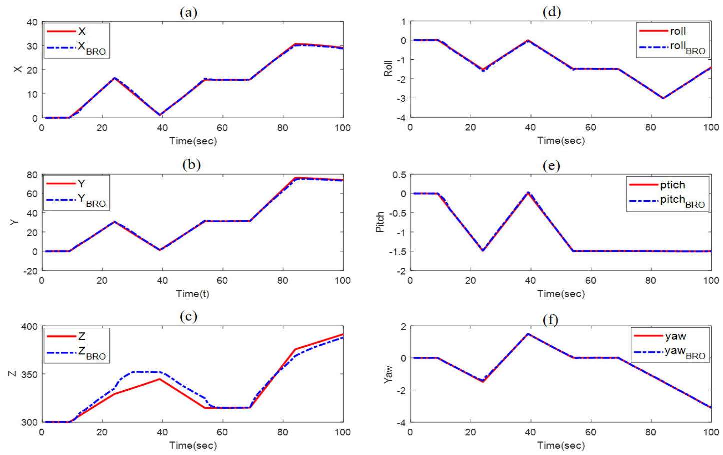

Fig. 3. Dynamic performance of pose estimation by BRO with Gaussian noises.

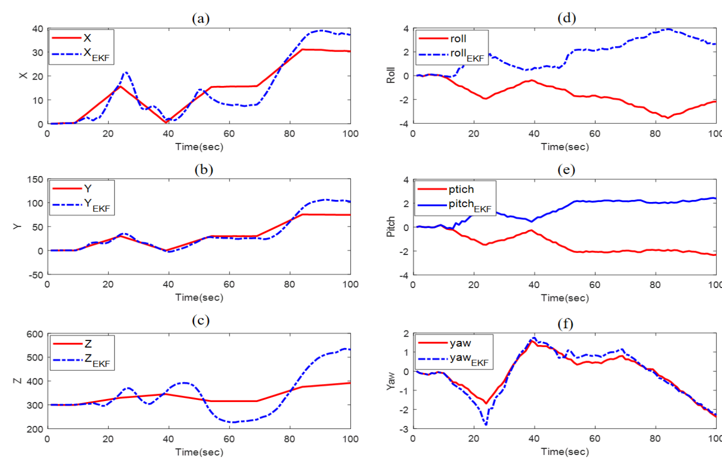

Fig. 4. Abosulte value of estimation error under Gaussian noises.

Case 2: The dynamic and measurement noises are set as bounded noises, i.e., $w(t) = 0.2\phi_w(t) - 0.1$ and $v(t) = 0.2\phi_v(t) - 0.1$, where $\phi_w(t)$ and $\phi_v(t)$ are random variables that can be generated by the function "rand" of MATLAB. It is obvious that $\phi_w(t)$ and $\phi_v(t)$ are not Gaussian white noises. The pose estimation results of EKF and BRO under bounded noises are plotted in Fig.5 and Fig.6, respectively. It is seen from Fig.5 that EKF suffers from significant pose estimation performance degradation under bounded noises, while the estimation results by BRO approach is within a tolerable range, this is as expected for the bounded noises based methods. The absolute value of estimation errors by EKF and BRO under bounded noises are plotted in Fig.7, which shows the superiority of the proposed BRO approach under bounded noises as compared with the EKF.

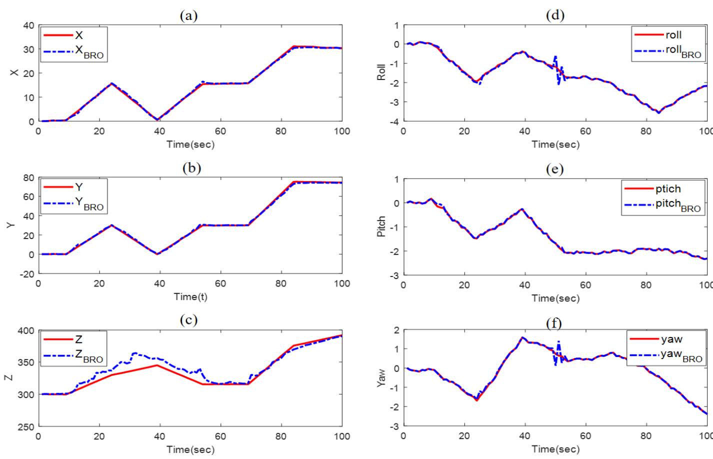

Fig. 5. Dynamic performance of pose estimation by EKF with bounded noises.

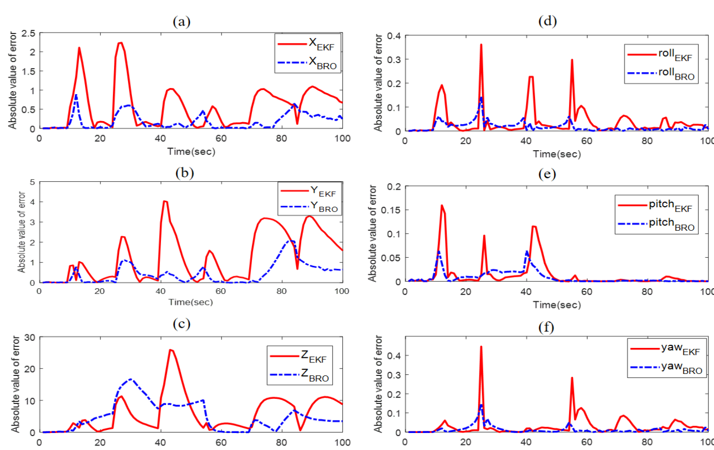

Fig. 6. Dynamic performance of pose estimation by BRO with bounded noises.

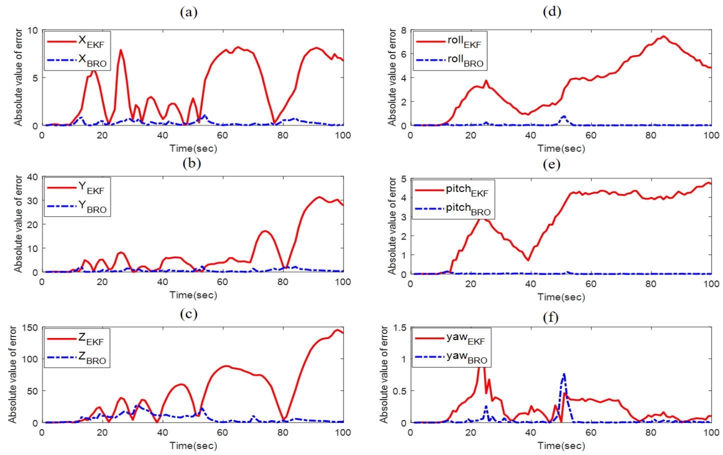

Fig. 7. Absolute value of estimation error under bounded noises.

5 Conclusion

In this paper, a BRO approach is developed for pose estimation in RVS, where the pose estimation problem was converted into a convex optimization problem that can be directly solved by the LMI Toolbox in MATLAB. Notice that the BRO approach does not require any statistical information of measurement and dynamic noises. Finally, a pose estimation simulation was exploited to show the advantages and effectiveness of the proposed methods.

References

1. Janabi-Sharifi, F., & Wilson, W. J. (1997). Automatic selection of image features for visual servoing. IEEE Transactions on Robotics and Automation, 13(6), 890-903.
2. Haralick, B. M., Lee, C. N., Ottenberg, K., & Nlle, M. (1994). Review and analysis of solutions of the three point perspective pose estimation problem. International journal of computer vision, 13(3), 331-356.
3. Faugeras, O., & FAUGERAS, O. A. (1993). Three-dimensional computer vision: a geometric viewpoint. MIT press.
4. Lowe, D. G. (1987). Three-dimensional object recognition from single two-dimensional images. Artificial intelligence, 31(3), 355-395.
5. Lu, C. P., Hager, G. D., & Mjolsness, E. (2000). Fast and globally convergent pose estimation from video images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(6), 610-622.
6. Wilson, W. J., Hulls, C. W., & Janabi-Sharifi, F. (2000). Robust image processing and position-based visual servoing (pp. 163-201). IEEE Press, New York.

7. Ficocelli, M., & Janabi-Sharifi, F. (2001). Adaptive filtering for pose estimation in visual servoing. In Proceedings 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems. Expanding the Societal Role of Robotics in the the Next Millennium (Cat. No. 01CH37180) (Vol. 1, pp. 19-24). IEEE.
8. Janabi-Sharifi, F., & Marey, M. (2010). A kalman-filter-based method for pose estimation in visual servoing. IEEE transactions on Robotics, 26(5), 939-947.
9. Chen, B., Hu, G., Ho, D. W., Zhang, W. A., & Yu, L. (2017). Distributed robust fusion estimation with application to state monitoring systems. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 47(11), 2994-3005.
10. Chen, B., Hu, G., Ho, D. W., & Yu, L. (2019). A new approach to linear/nonlinear distributed fusion estimation problem. IEEE Transactions on Automatic Control, 64(3), 1301-1308.
11. Chen, B., & Hu, G. (2018). Nonlinear state estimation under bounded noises. Automatica, 98, 159-168.
12. Wilson, W. J., Hulls, C. W., & Bell, G. S. (1996). Relative end-effector control using cartesian position based visual servoing. IEEE Transactions on Robotics and Automation, 12(5), 684-696.
13. Ficocelli, M., & Janabi-Sharifi, F. (2001). Adaptive filtering for pose estimation in visual servoing. In Proceedings 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems. Expanding the Societal Role of Robotics in the the Next Millennium (Cat. No. 01CH37180) (Vol. 1, pp. 19-24). IEEE.
14. Boyd, S., El Ghaoui, L., Feron, E., & Balakrishnan, V. (1994). Studies in applied mathematics: Vol. 15. Linear matrix inequalities in system and control theory. Philadelphia, PA: SIAM.