

Visual Motion Estimation of 3D Objects: An Adaptive Extended Kalman Filter Approach

Vincenzo Lippiello, Bruno Siciliano, Luigi Villani
 Dipartimento di Informatica e Sistemistica
 Università degli Studi di Napoli Federico II
 via Claudio 21, 80125 Napoli, Italy
 {vincenzo.lippiello, siciliano, lvillani}@unina.it

Abstract—An algorithm for the visual estimation of the pose of a moving object is presented in this paper. The algorithm exploits the prediction capability of the Extended Kalman Filter to realize in real time a dynamic optimal selection of the object image features used for pose estimation. The robustness of the system with respect to the measurement noise and modelling errors is enhanced by using an adaptive scheme. Experimental case studies are presented to prove the effectiveness of the proposed approach.

I. INTRODUCTION

The real-time estimation of the position and orientation of a moving target object from image measurements plays a central role in position-based visual servoing, where this information is used to perform closed-loop position control of robots [1], [2]. This problem has been largely investigated in the computer vision literature (see, e.g., [3], [4]) as well as in the robotics literature (see, e.g., [5], [6]).

To enhance the estimation accuracy in spite of noise and disturbances, the Extended Kalman Filter (EKF) is a common adopted solution [3]–[6]. If the quality of the camera sensors is good, the illumination of the scene is stable, and the velocity of the target object is almost constant, then it is reasonable to assume that the statistics of the state noise and of the observation noise are known a priori and remain constant, as required by the EKF. On the other hand, if one or more of the above conditions are not verified, it may be convenient to adopt an Adaptive Extended Kalman Filter (AEKF). This can be designed using different approaches, e.g., introducing additional variables to be estimated by the filter, as the noise statistics [7], using interacting multiple models [8], adjusting the transition matrix of the filter [9], or updating the covariance matrices of the state and observation noise [10].

The EKF is the core of a visual motion estimation scheme presented in our previous works [11], [12]. This algorithm can be effectively adopted for polyhedral objects, typical of industrial applications, and is based on the use of point features (the object corners). To reduce the processing time and improve the estimation accuracy, a computationally efficient technique is used for the selection of an optimal subset of feature points, among all the visible points, at each sample time. Moreover, the adoption of a Binary Space Partition tree for representing the object geometry allows recognizing and discarding all the feature points that are occluded with respect to the camera.

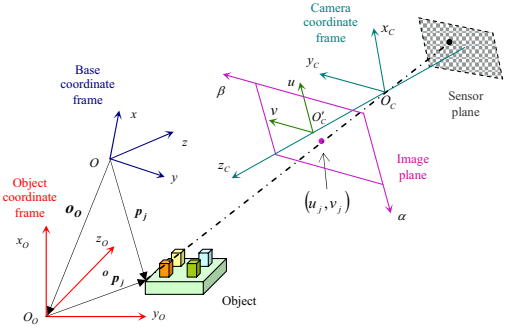


Fig. 1. Pin-hole model of the camera and reference frames.

In this work, the visual motion estimation algorithm presented in [11], [12] has been enhanced by using an adaptive formulation of the EKF, to cope with uncertain or varying noise statistics. The algorithm is based on the approach proposed in [10] and successfully tested in different applications, e.g., power systems [13], mobile robots [14], visual pose estimation [15]. The adaptive algorithm computes the statistics of both the state noise and the observation noise. With respect to the work [10], the contribution of this paper mainly concerns with the adaptive law for the observation noise statistics, which has been suitably designed for a visual motion estimation problem based on the use of a variable set of image features. This aspect is not considered in previous works on AEKF applied to visual motion estimation (see, e.g., [15]), where only an adaptive law for the state noise covariance matrix is used.

The performance of the EKF and of the AEKF are compared in a number of experimental case studies developed on a test bed consisting of one fixed camera and a robot manipulator carrying the target object. The robot allows moving the object according to a known position and orientation trajectory, so that the estimation error can be evaluated.

II. MODELING

Consider the pin-hole model of a video camera fixed with respect to a base coordinate frame $O-xyz$ represented in Fig. 1. Let $O_c-x_c y_c z_c$ be a frame attached to the camera (camera frame), with the z_c -axis aligned to the optical axis and the origin in the optical center. In the following, a

superscript will be used to denote the reference frame of a variable, when different from the base frame.

The sensor plane is parallel to the $x_c y_c$ -plane at a distance $-f_e$ along the z_c -axis, where f_e is the effective focal length of the camera lens. The image plane is parallel to the $x_c y_c$ -plane at a distance f_e along the z_c -axis. The intersection of the optical axis with the image plane defines the principal optic point O' , which is the origin of the image frame $O'-uv$.

A point P with coordinates ${}^c\mathbf{p} = [{}^cx \ {}^cy \ {}^cz]^T$ in the camera frame is projected onto the point of the image plane whose coordinates can be computed with the equation

$$\begin{bmatrix} u \\ v \end{bmatrix} = \frac{f_e}{c_z} \begin{bmatrix} {}^cx \\ {}^cy \end{bmatrix}. \quad (1)$$

The position and orientation of an object frame $O_o-x_o y_o z_o$ with respect to the base frame can be expressed in terms of the coordinate vector of the origin $\mathbf{o}_o = [x_o \ y_o \ z_o]^T$ and of the rotation matrix $\mathbf{R}_o(\phi_o)$, where $\phi_o = [\varphi_o \ \vartheta_o \ \psi_o]^T$ is the vector of the Roll, Pitch and Yaw angles.

Consider m feature points of the object. The coordinate vector ${}^c\mathbf{p}_j$ of the feature point P_j ($j = 1, \dots, m$) can be expressed in the camera frame as

$${}^c\mathbf{p}_j = \mathbf{R}_c^T (\mathbf{o}_o - \mathbf{o}_c + \mathbf{R}_o(\phi_o) {}^o\mathbf{p}_j) \quad (2)$$

where $\mathbf{o}_c$ and $\mathbf{R}_c$ are, respectively, the position vector and the rotation matrix of the camera frame referred to the base frame, whereas ${}^o\mathbf{p}_j$ is the coordinate vector of P_j expressed in the object frame. Notice that ${}^o\mathbf{p}_j$ is a constant vector that is assumed to be known, since it can be computed from a CAD model of the object. Moreover, the quantities $\mathbf{o}_c$ and $\mathbf{R}_c$ are constant, because the camera is assumed to be fixed to the workspace, and can be computed through a suitable calibration procedure [16].

By folding the $3m$ equations (2) into the equation (1), a system of $2m$ nonlinear equations is achieved. The equations depend on the measurements of the m feature points in the image plane of the camera, while the six components of the vectors $\mathbf{o}_o$ and ϕ_o are the unknown quantities to be estimated. To solve these equations at least three non-aligned points (six independent equations) are required.

The computation of the solution is nontrivial and for visual tracking applications it has to be repeated at a high sampling rate. The Kalman filter provides a computationally tractable solution, which can also incorporate and exploit redundant measurement information.

III. EXTENDED KALMAN FILTER

In order to estimate the pose of the object, a discrete-time state space model of the object motion has to be considered, based on the (12×1) state vector

$$\mathbf{w} = [x_o \ \dot{x}_o \ y_o \ \dot{y}_o \ z_o \ \dot{z}_o \ \varphi_o \ \dot{\varphi}_o \ \vartheta_o \ \dot{\vartheta}_o \ \psi_o \ \dot{\psi}_o]^T. \quad (3)$$

For simplicity, the object velocity is assumed to be constant over one sample period T . This approximation is reasonable in the hypothesis that T is sufficiently small. The

corresponding dynamic modelling error can be considered as an input disturbance γ_k . The discrete-time dynamic model can be written as

$$\mathbf{w}_k = \mathbf{A}\mathbf{w}_{k-1} + \gamma_k \quad (4)$$

where the state transition matrix $\mathbf{A}$ is a constant (12×12) block diagonal matrix of the form

$$\mathbf{A} = \text{diag} \left\{ \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix}, \dots, \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \right\}.$$

The outputs of the Kalman filter are chosen as the vectors of the normalized coordinates of the m feature points in the image plane of the camera

$$\zeta_k^u = \begin{bmatrix} \frac{u_1}{f_e} & \dots & \frac{u_m}{f_e} \end{bmatrix}_k^T \quad (5a)$$

$$\zeta_k^v = \begin{bmatrix} \frac{v_1}{f_e} & \dots & \frac{v_m}{f_e} \end{bmatrix}_k^T. \quad (5b)$$

In view of (1), the corresponding output model can be written in the form

$$\zeta_k^u = \mathbf{g}^u(\mathbf{w}_k) + \boldsymbol{\nu}_k^u \quad (6a)$$

$$\zeta_k^v = \mathbf{g}^v(\mathbf{w}_k) + \boldsymbol{\nu}_k^v \quad (6b)$$

where $\boldsymbol{\nu}_k^u$ and $\boldsymbol{\nu}_k^v$ are the observation noise vectors for the u and v components of the normalized image plane, whereas the vector functions $\mathbf{g}^u(\mathbf{w}_k)$ and $\mathbf{g}^v(\mathbf{w}_k)$ are defined as

$$\mathbf{g}^u(\mathbf{w}_k) = \begin{bmatrix} \frac{{}^cx_1}{c_{z_1}} & \dots & \frac{{}^cx_m}{c_{z_m}} \end{bmatrix}_k^T \quad (7a)$$

$$\mathbf{g}^v(\mathbf{w}_k) = \begin{bmatrix} \frac{{}^cy_1}{c_{z_1}} & \dots & \frac{{}^cy_m}{c_{z_m}} \end{bmatrix}_k^T. \quad (7b)$$

The coordinates of the feature points ${}^c\mathbf{p}_j$ in equations (7) are computed from the state vector $\mathbf{w}_k$ via equation (2).

The components of the disturbance quantities γ_k , $\boldsymbol{\nu}_k^u$ and $\boldsymbol{\nu}_k^v$ are considered as independent, non-stationary, Gaussian, white noise sequences with the statistical properties: $\mathbb{E}[\gamma_k] = \mathbf{q}_k$, $\mathbb{E}[\boldsymbol{\nu}_k^u] = \mathbf{r}_k^u$, $\mathbb{E}[\boldsymbol{\nu}_k^v] = \mathbf{r}_k^v$, $\mathbb{E}[(\mathbf{w}_k - \mathbf{q}_k)(\mathbf{w}_l - \mathbf{q}_l)^T] = \mathbf{Q}_k \delta_{kl}$, $\mathbb{E}[(\boldsymbol{\nu}_k^u - \mathbf{r}_k^u)(\boldsymbol{\nu}_l^u - \mathbf{r}_l^u)^T] = \mathbf{R}_k^u \delta_{kl}$, $\mathbb{E}[(\boldsymbol{\nu}_k^v - \mathbf{r}_k^v)(\boldsymbol{\nu}_l^v - \mathbf{r}_l^v)^T] = \mathbf{R}_k^v \delta_{kl}$, where $\mathbb{E}[\cdot]$ indicates the statistical mean operator applied to the components of a vector or matrix, and δ is the Kronecker symbol.

Notice that, since the output model is nonlinear in the system state, the Extended Kalman Filter must be adopted.

The equations of the EKF with fixed statistical parameters are given by:

$$\mathbf{w}_{k,k} = \mathbf{w}_{k,k-1} + \begin{bmatrix} \mathbf{K}_k^u & \mathbf{K}_k^v \end{bmatrix} \begin{bmatrix} \zeta_k^u - \mathbf{g}^u(\mathbf{w}_{k,k-1}) - \mathbf{r}_k^u \\ \zeta_k^v - \mathbf{g}^v(\mathbf{w}_{k,k-1}) - \mathbf{r}_k^v \end{bmatrix} \quad (8a)$$

$$\mathbf{P}_{k,k} = \mathbf{P}_{k,k-1} - \begin{bmatrix} \mathbf{K}_k^u & \mathbf{K}_k^v \end{bmatrix} \begin{bmatrix} \mathbf{H}_k^u \\ \mathbf{H}_k^v \end{bmatrix} \mathbf{P}_{k,k-1}, \quad (8b)$$

where $\mathbf{w}_{k,k-1}$ is the propagated state vector and $\mathbf{P}_{k,k-1}$ is the (12×12) covariance matrix conditioned on observations prior to time k , computed as:

$$\mathbf{w}_{k,k-1} = \mathbf{A}\mathbf{w}_{k-1,k-1} + \mathbf{q}_{k-1} \quad (9a)$$

$$\mathbf{P}_{k,k-1} = \mathbf{A}\mathbf{P}_{k-1,k-1}\mathbf{A}^T + \mathbf{Q}_{k-1}. \quad (9b)$$

In (8), $\mathbf{K}_k^u$ and $\mathbf{K}_k^v$ are the $(12 \times m)$ Kalman matrix gains

$$\mathbf{K}_k^u = \mathbf{P}_{k,k-1} \mathbf{H}_k^{u\top} (\mathbf{R}_k^u + \mathbf{\Gamma}_k^u)^{-1} \quad (10a)$$

$$\mathbf{K}_k^v = \mathbf{P}_{k,k-1} \mathbf{H}_k^{v\top} (\mathbf{R}_k^v + \mathbf{\Gamma}_k^v)^{-1}, \quad (10b)$$

being $\mathbf{H}_k^u$ and $\mathbf{H}_k^v$ the $(m \times 12)$ Jacobian matrices of the output vector functions

$$\mathbf{H}_k^u = \left. \frac{\partial \mathbf{g}^u(\mathbf{w})}{\partial \mathbf{w}} \right|_{\mathbf{w}=\mathbf{w}_{k,k-1}} \quad (11a)$$

$$\mathbf{H}_k^v = \left. \frac{\partial \mathbf{g}^v(\mathbf{w})}{\partial \mathbf{w}} \right|_{\mathbf{w}=\mathbf{w}_{k,k-1}}, \quad (11b)$$

and $\mathbf{\Gamma}_k^u$ and $\mathbf{\Gamma}_k^v$ defined as

$$\mathbf{\Gamma}_k^u = \mathbf{H}_k^u \mathbf{P}_{k,k-1} \mathbf{H}_k^{u\top} \quad (12a)$$

$$\mathbf{\Gamma}_k^v = \mathbf{H}_k^v \mathbf{P}_{k,k-1} \mathbf{H}_k^{v\top}. \quad (12b)$$

Notice that the implementation of the filter requires a prior estimate of $\mathbf{w}_{0,0}$, $\mathbf{P}_{0,0}$, and of the statistical parameters of the state and observation noise.

IV. ADAPTIVE EXTENDED KALMAN FILTER

If a high-quality camera sensor is used, the illumination of the scene is stable, and the velocity of the tracked object is nearly constant, then it is possible to use constant statistical parameters with optimal results. On the other hand, if these conditions are not satisfied, it may be convenient to update in real time the statistical parameters $\{\mathbf{q}_k, \mathbf{Q}_k, \mathbf{r}_k^u, \mathbf{r}_k^v, \mathbf{R}_k^u, \mathbf{R}_k^v\}$. This leads to the Adaptive Extended Kalman Filter (AEKF).

Following the adaptive approach proposed in [10], the statistical parameters are assumed to be constant over N sample times. Moreover, since not all the visual features are always available during the motion and their location into the scene is strongly variable, it may be reasonable to assume the statistics of the observation noise to be equal for all the measurements of the feature points in the scene at time k . Hence the quantities $\{\mathbf{r}_k^u, \mathbf{r}_k^v, \mathbf{R}_k^u, \mathbf{R}_k^v\}$ are replaced by the quantities $\{r_k^u \mathbf{v}_m, r_k^v \mathbf{v}_m, \sigma_k^{u2} \mathbf{I}_m, \sigma_k^{v2} \mathbf{I}_m\}$, where $\mathbf{v}_m$ indicates a $(m \times 1)$ vector of components equal to 1 and $\mathbf{I}_m$ indicates the $(m \times m)$ identity matrix. Moreover, the samples of the observation noise sequences $\mathbf{v}_i^u$ ($\mathbf{v}_i^v$) are independent for $i = 1, \dots, N$ and have a gaussian distribution with mean $r^u \mathbf{v}_m$ ($r^v \mathbf{v}_m$) and variance $\sigma^{u2} \mathbf{I}_m$ ($\sigma^{v2} \mathbf{I}_m$), where the parameters r^u , r^v , σ^u and σ^v are constant over N sample times.

In view of the nonlinear relation (6), an intuitive approximation of the observation noise sample vectors at time k is given by the quantities

$$\mathbf{r}_k^u = \zeta_k^u - \mathbf{g}^u(\mathbf{w}_{k,k-1}) \quad (13a)$$

$$\mathbf{r}_k^v = \zeta_k^v - \mathbf{g}^v(\mathbf{w}_{k,k-1}) \quad (13b)$$

which can be considered as independent and identically distributed over N samples. It can be shown (see, e.g.,

[10]) that an unbiased estimator for r^u and r^v can be taken as

$$\hat{r}^u = \frac{1}{N} \sum_{i=1}^N \bar{\rho}_i^u \quad (14a)$$

$$\hat{r}^v = \frac{1}{N} \sum_{i=1}^N \bar{\rho}_i^v, \quad (14b)$$

where $\bar{\rho}_i^u$ and $\bar{\rho}_i^v$ are scalar quantities equal to the mean values of the components of the vectors $\boldsymbol{\rho}_i^u$ and $\boldsymbol{\rho}_i^v$ respectively. Moreover, an unbiased estimator for σ^{u2} and σ^{v2} may be obtained as

$$\hat{\sigma}^{u2} = \frac{1}{m(N-1)} \sum_{i=1}^N \left\{ \|\boldsymbol{\rho}_i^u - \hat{r}^u \mathbf{v}_m\|^2 - \frac{N-1}{N} \text{tr}(\mathbf{\Gamma}_i^u) \right\} \quad (15a)$$

$$\hat{\sigma}^{v2} = \frac{1}{m(N-1)} \sum_{i=1}^N \left\{ \|\boldsymbol{\rho}_i^v - \hat{r}^v \mathbf{v}_m\|^2 - \frac{N-1}{N} \text{tr}(\mathbf{\Gamma}_i^v) \right\}. \quad (15b)$$

where $\text{tr}(\cdot)$ is the trace of the input matrix.

For the state noise statistics, in view of the linear dynamic state relation at time k given by (4), an intuitive approximation for the state noise vector at time k is

$$\boldsymbol{\varrho}_k = \mathbf{w}_k - \mathbf{A} \mathbf{w}_{k,k-1}, \quad (16)$$

which may be considered independent and identically distributed over N samples. It can be shown that an unbiased estimator for the mean value $\mathbf{q}$ of the state noise may be obtained as

$$\hat{\mathbf{q}} = \frac{1}{N} \sum_{i=1}^N \mathbf{e}_i, \quad (17)$$

while an unbiased estimator for the covariance matrix $\mathbf{Q}$ is given by

$$\hat{\mathbf{Q}} = \frac{1}{N-1} \sum_{i=1}^N \left\{ (\mathbf{e}_i - \hat{\mathbf{q}})(\mathbf{e}_i - \hat{\mathbf{q}})^\top - \frac{N-1}{N} \mathbf{\Delta}_i \right\}, \quad (18)$$

where $\mathbf{\Delta}_i = \mathbf{A} \mathbf{P}_{i,i-1} \mathbf{A}^\top - \mathbf{P}_{i,i}$.

In sum, the equations (13)–(18) provide a heuristic unbiased estimator for the statistical parameters of a EKF used for visual motion estimation, on the assumption that the last N samples are statistically independent and identically distributed.

V. ITERATIVE AEKF

Using the previous results, an iterative limited memory formulation of the AEKF may be designed. The required prior knowledge information is represented by an initial estimate of the quantities $\mathbf{w}_{1,0}$, $\mathbf{P}_{1,0}$, $\hat{\mathbf{q}}_0$, $\hat{\mathbf{Q}}_0$, $\hat{r}_0^u$, $\hat{r}_0^v$, $\hat{\sigma}_0^u$ and $\hat{\sigma}_0^v$.

For generality, it is assumed that the observation noise statistical parameters are constant over N_r time samples while the state noise statistical parameters are constant over N_q time samples.

The first step of the iterative algorithm is the linearization of the output Kalman model about the last predicted

value of the state $\mathbf{w}_{k,k-1}$, according to (11), and the computation of the matrices $\mathbf{\Gamma}_k^u$ and $\mathbf{\Gamma}_k^v$ in (12).

Starting from the time N_r , the second step is the evaluation of the current noise vector and the computation of the estimated observation noise statistics as follows

$$\boldsymbol{\rho}_k^u = \boldsymbol{\zeta}_k^u - \mathbf{g}^u(\mathbf{w}_{k,k-1}) \quad (19a)$$

$$\boldsymbol{\rho}_k^v = \boldsymbol{\zeta}_k^v - \mathbf{g}^v(\mathbf{w}_{k,k-1}) \quad (19b)$$

$$\hat{\mathbf{r}}_k^u = \hat{\mathbf{r}}_{k-1}^u + \frac{1}{N_r}(\bar{\boldsymbol{\rho}}_k^u - \bar{\boldsymbol{\rho}}_{k-N_r}^u) \quad (19c)$$

$$\hat{\mathbf{r}}_k^v = \hat{\mathbf{r}}_{k-1}^v + \frac{1}{N_r}(\bar{\boldsymbol{\rho}}_k^v - \bar{\boldsymbol{\rho}}_{k-N_r}^v) \quad (19d)$$

$$\begin{aligned} \hat{\sigma}_k^{u2} = & \hat{\sigma}_{k-1}^{u2} + \frac{1}{m(N_r-1)} \left\{ \|\boldsymbol{\rho}_k^u - \hat{\mathbf{r}}_k^u \mathbf{I}_m\|^2 \right. \\ & - \|\boldsymbol{\rho}_{k-N_r}^u - \hat{\mathbf{r}}_{k-N_r}^u \mathbf{I}_m\|^2 + \\ & \left. + \frac{1}{N_r} \|\boldsymbol{\rho}_k^u - \boldsymbol{\rho}_{k-N_r}^u\|^2 + \frac{N_r-1}{N_r} \text{tr}(\mathbf{\Gamma}_{k-N_r}^u - \mathbf{\Gamma}_k^u) \right\} \end{aligned} \quad (19e)$$

$$\begin{aligned} \hat{\sigma}_k^{v2} = & \hat{\sigma}_{k-1}^{v2} + \frac{1}{m(N_r-1)} \left\{ \|\boldsymbol{\rho}_k^v - \hat{\mathbf{r}}_k^v \mathbf{I}_m\|^2 \right. \\ & - \|\boldsymbol{\rho}_{k-N_r}^v - \hat{\mathbf{r}}_{k-N_r}^v \mathbf{I}_m\|^2 + \\ & \left. + \frac{1}{N_r} \|\boldsymbol{\rho}_k^v - \boldsymbol{\rho}_{k-N_r}^v\|^2 + \frac{N_r-1}{N_r} \text{tr}(\mathbf{\Gamma}_{k-N_r}^v - \mathbf{\Gamma}_k^v) \right\}. \end{aligned} \quad (19f)$$

The third step consists in the evaluation of the Kalman gains

$$\mathbf{K}_k^u = \mathbf{P}_{k,k-1} \mathbf{H}_k^{uT} (\mathbf{\Gamma}_k^u + \hat{\sigma}_k^{u2} \mathbf{I}_m)^{-1} \quad (20a)$$

$$\mathbf{K}_k^v = \mathbf{P}_{k,k-1} \mathbf{H}_k^{vT} (\mathbf{\Gamma}_k^v + \hat{\sigma}_k^{v2} \mathbf{I}_m)^{-1}, \quad (20b)$$

while the fourth step is the state correction on the basis of the current measurements

$$\mathbf{w}_{k,k} = \mathbf{w}_{k,k-1} + [\mathbf{K}_k^u \quad \mathbf{K}_k^v] \begin{bmatrix} \boldsymbol{\rho}_k^u - \hat{\mathbf{r}}_k^u \mathbf{I}_m \\ \boldsymbol{\rho}_k^v - \hat{\mathbf{r}}_k^v \mathbf{I}_m \end{bmatrix} \quad (21a)$$

$$\mathbf{P}_{k,k} = \mathbf{P}_{k,k-1} - [\mathbf{K}_k^u \quad \mathbf{K}_k^v] \begin{bmatrix} \mathbf{H}_k^u \\ \mathbf{H}_k^v \end{bmatrix} \mathbf{P}_{k,k-1}. \quad (21b)$$

Starting from the time N_q , the fifth step is the evaluation of the current state noise vector and the computation of the estimated state noise statistics as follows

$$\boldsymbol{\varrho}_k = \mathbf{w}_{k,k} - \mathbf{A} \mathbf{w}_{k,k-1} \quad (22a)$$

$$\boldsymbol{\Delta}_k = \mathbf{A} \mathbf{P}_{k,k-1} \mathbf{A}^T - \mathbf{P}_{k,k} \quad (22b)$$

$$\hat{\mathbf{q}}_k = \hat{\mathbf{q}}_{k-1} + \frac{1}{N_q}(\boldsymbol{\varrho}_k - \boldsymbol{\varrho}_{k-N_q}) \quad (22c)$$

$$\begin{aligned} \hat{\mathbf{Q}}_k = & \hat{\mathbf{Q}}_{k-1} + \frac{1}{N_q-1} \left\{ (\boldsymbol{\varrho}_k - \hat{\mathbf{q}}_k)(\boldsymbol{\varrho}_k - \hat{\mathbf{q}}_k)^T \right. \\ & - (\boldsymbol{\varrho}_{k-N_q} - \hat{\mathbf{q}}_k)(\boldsymbol{\varrho}_{k-N_q} - \hat{\mathbf{q}}_k)^T \\ & + \frac{1}{N_q} (\boldsymbol{\varrho}_k - \boldsymbol{\varrho}_{k-N_q})(\boldsymbol{\varrho}_k - \boldsymbol{\varrho}_{k-N_q})^T \\ & \left. + \frac{N_q-1}{N_q} (\boldsymbol{\Delta}_{k-N_q}^u - \boldsymbol{\Delta}_k) \right\}. \end{aligned} \quad (22d)$$

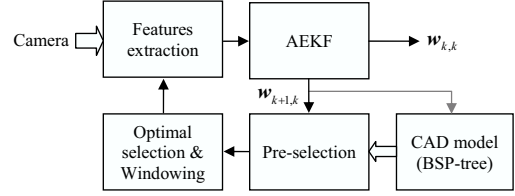


Fig. 2. Block scheme of the visual motion estimation algorithm.

The sixth and last step consists in the evaluation of the predicted state for the next sample time

$$\mathbf{w}_{k+1,k} = \mathbf{A} \mathbf{w}_{k,k} + \hat{\mathbf{q}}_k \quad (23a)$$

$$\mathbf{P}_{k+1,k} = \mathbf{A} \mathbf{P}_{k,k} \mathbf{A}^T + \hat{\mathbf{Q}}_k. \quad (23b)$$

Notice that the update of the noise statistics starts from time N_r for the observation noise and from time N_q for the state noise. Before such times, those quantities are constant and equal to the initial values.

VI. VISUAL MOTION ESTIMATION ALGORITHM

The accuracy of the estimate provided by the Kalman filter depends on the number of the available feature points. Inclusion of extra points may improve the estimation accuracy but increases the computational cost. It has been shown that a number of feature points between four and six, if properly chosen, may represent a good trade-off [6]. Automatic selection algorithms have been developed to find the optimal feature points (see e.g., [17], [18]). In the following, an efficient selection technique proposed in [11] is adopted, which exploits the prediction of the object pose provided by the Kalman filter to perform a pre-selection of the points that are visible to the camera at the next sample time. A detailed presentation of the algorithm, which is based on the Binary Space Partitioning (BSP) tree structure to represent the object geometry, can be found in [11], [12].

The algorithm may be represented as the closed-loop-like scheme of Fig. 2. It can be seen that the prediction of object pose at next sample time ($\mathbf{w}_{k+1,k}$) provided by the Kalman filter is fed back to the pre-selection algorithm which, on the basis of a BSP representation of the object, is capable to find and locate all the feature points which are expected to be visible from the camera. Then, an optimal selection algorithm is used to choose an optimal subset of feature points and compute the size and location of the corresponding windows on the image plane to be grabbed for image processing. Finally, the feature extraction algorithm provides the effective position of the feature points to be input to the Kalman filter which computes the actual object pose $\mathbf{w}_{k,k}$.

The above algorithm allows the nice features of Kalman filtering to be fully exploited. In fact, the AEKF provides robustness with respect to visual measurement noise and changing in the lighting conditions. Moreover, the closed loop structure allows achieving robustness with respect to the loss of feature points caused by occlusion or exit from



Fig. 3. COMAU Smart3-S robot carrying the moving object and SONY 8500CE cameras.

the visible space of the camera. Finally, the use of point-type features combined with the optimal selection and windowing algorithms allows minimizing the computational time required for image processing.

VII. EXPERIMENTS

The effectiveness and the performance of the proposed visual motion estimation algorithm have been experimentally tested using a single-camera visual system.

The experimental set-up is composed by a PC with Pentium IV 1.7GHz processor equipped with a MATROX Genesis board, a SONY 8500CE B/W camera, and a COMAU Smart3-S robot. The MATROX board is used as frame grabber and for a partial image processing (e.g., windows extraction from the image). The PC host is also used to realize the whole BSP structure management, the pre-selection algorithm, the selection algorithm, the dynamic windowing and the Kalman filtering. Some steps of image processing have been parallelized on the MATROX board and on the PC, so as to reduce computational time.

The robot is used to move an object in the visual space of the camera; thus the object position and orientation with respect to the base frame of the robot can be computed from joint position measurements via the direct kinematic equation. In order to test the accuracy of the estimation provided by the Kalman filter, the camera was calibrated with respect to the base frame of the robot using a suitable calibration procedure presented in [16], where the robot is exploited to place a calibration pattern in some known pose of the visible space of the camera.

The camera resolution is 576×763 pixels and the nominal focal length of the lenses is 16 mm. The camera is disposed at a distance of about 130 cm from the object. The sampling time used for estimation is limited by the camera frame rate, which is about 26 fps. A simple neon illumination has been used, in order to test the robustness of the setup in the case of noisy visual measurements. In fact, during the object motion, the local illumination conditions of the windows of the image plane selected for feature extraction are quite variable due to reflections or shadows.

The image features are the corners of the object, which can be extracted with high robustness in various envi-

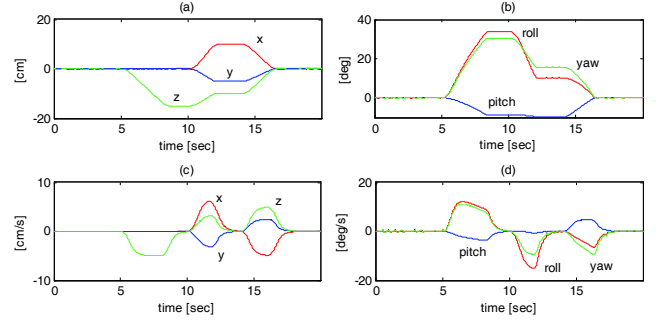


Fig. 4. Time history of the object trajectory **TrajPO**. Position (a), RPY angles (b), linear velocity (c), time derivative of RPY angles (d).

ronmental conditions. The feature extraction algorithm is based on Canny's method for edge detection [19] and on a simple custom implementation of a corner detector. The object used in the experiment has 40 corners, which are all candidate for feature extraction. Fig. 3 shows the stereo vision system and the robot carrying the object. Notice that only the left camera is used in the experiments.

For both EKF and AEKF, the initial value of the matrix $P_{1,0}$ has been chosen as the null matrix; moreover, the initial value of the state vector $w_{1,0}$ has been set null for the velocity components, while the pose components has been roughly estimated through direct measurements.

The covariance matrix Q has been chosen as a diagonal matrix, both in the non-adaptive and in the adaptive case; moreover, to prevent some typical implementation problems of Kalman filters, some of the modifications used in [10] have been adopted.

The values of the statistical parameters used for the EKF are set as initial values for the AEKF; they are:

$$\begin{aligned}\hat{r}_0^u &= \hat{r}_0^v = 0, \\ \hat{\sigma}_0^{u2} &= \hat{\sigma}_0^{v2} = 9.0, \\ \hat{q}_0 &= \mathbf{0}, \\ \hat{Q}_0 &= \text{diag} \{0, 5, 0, 5, 0, 5, 0, 20, 0, 20, 0, 20\} \cdot 10^{-6}.\end{aligned}$$

The physical dimensions are: pixel and pixel², respectively, for the mean and variances of the observation noise; mm, mm/s, rad and rad/s for the components of the mean of the state noise; mm², (mm/s)², rad² and (rad/s)² for the corresponding covariances.

The initial values of the observation noise covariances have been evaluated during the camera calibration procedure while the initial values of the state noise covariances have been set on the basis of the velocity range of the object trajectories. These values have been further tuned on the basis of a set of experiments carried out using the EKF, to achieve satisfactory tracking performance.

Notice that all the elements of the covariance matrix $\hat{Q}_0$ corresponding to the position components of the state have been considered initially zero for the AEKF and constantly zero for the EKF. Moreover, the values $N_q = N_r = 30$ have been chosen.

The case study is aimed at comparing the performance of the EKF to that of the AEKF considering different object

		EKF		AEKF		Improvement	
		Mean	StD	Mean	StD	Mean	StD
TrajP	e_P (mm)	13.71	11.16	12.59	10.89	8.2%	2.4%
	e_O (deg)	3.50	3.20	2.29	1.58	34.6%	50.6%
TrajO	e_P (mm)	11.23	8.66	9.15	6.13	18.5%	29.2%
	e_O (deg)	5.95	5.26	4.78	3.33	19.7%	36.7%
TrajPO	e_P (mm)	13.58	8.70	11.92	8.06	12.2%	7.4%
	e_O (deg)	7.19	6.86	3.65	3.43	49.2%	50.0%

TABLE I

COMPARISON OF THE ESTIMATION ERRORS FOR EKF AND AEKF.

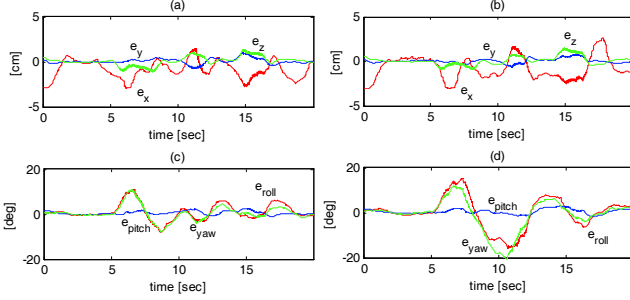


Fig. 5. Time history of the estimation errors with AEKF and EKF for the trajectory **TrajPO**. AEKF: position (a) and orientation (c) errors; EKF: position (b) and orientation (d) errors.

trajectories. It should be remarked that the adoption of the AEKF in lieu of the EKF causes only a modest increase of the computational cost that, in terms of overall processor time, is about 16 percent. This result is in accordance to those reported in [10] and [15].

Three different object trajectories are considered, both for the EKF and the AEKF:

- **TrajP**: the object position varies according to the time history reported on the left of Fig. 4 while the object orientation is left constant. The norm of the maximum linear velocity is about 10 cm/s.
- **TrajO**: the object position is left constant while the object orientation varies according to the time history reported on the right of Fig. 4. The norm of the maximum velocity for the RPY angles is about 20 deg/s.
- **TrajPO**: the object position and orientation varie according to the time history of Fig. 4.

Notice the trajectory **TrajPO** is the composition of **TrajP** and **TrajO**, but the resulting trajectories of the feature points on the image plane are different in the three cases. Hence the considered trajectories represent a significant test base to make a comparison.

The results of the experiments are summarized in Table I where the mean value and the standard deviation of the norm of the position error e_P and orientation error e_O are reported, together with the relative improvement achieved using the AEKF with respect to the EKF. It can be seen that the use of the AEKF allows a general improvements of the tracking performance, especially for the mean value and the standard deviation of the orientation error components.

For the trajectory **TrajPO**, the time history of the

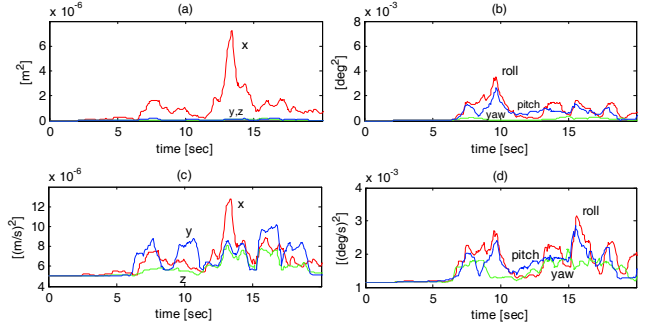


Fig. 6. Time history of the elements of the state noise covariance matrix for the trajectory **TrajPO**. Position (a), orientation (b), linear velocity (c), rotational velocity (d) components.

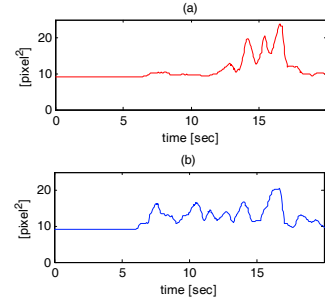


Fig. 7. Time history of the observation noise variance for the trajectory **TrajPO**. $\hat{\sigma}^{u^2}$ (a), $\hat{\sigma}^{v^2}$ (b).

estimation errors are shown in Fig. 5. It can be seen that the initial values of the errors are the same in the case of the EKF and the AEKF. In particular, there is an initial offset for the x -component of the position error, due to the initial misalignment between the real initial position and the initial value of the Kalman filter. This error is initially recovered by the Kalman filter during the first 5 s, in the absence of motion. During the motion, the position errors keep limited for all the components, but is higher for the x -component. In fact, the x -axis is aligned to the optical axis of the camera, thus the evaluation of the x -component of the object position, for a single-camera system, is more sensitive to measurements and modelling errors with respect to the position components lying on the image plane [6]. In general, the picks on the errors happen when the velocity is higher, due to the modelling error for the EKF. These errors are partially recovered by the AEKF, in reason of the adaptive law of the state noise covariance matrix Q .

The time histories of some of the statistical parameters which are updated on-line in the AEKF are also reported. In particular, the time histories of the elements of the (diagonal) covariance matrix of the state noise are shown in Fig. 6 while the time history of the observation noise for the u and v components are shown in Fig. 7. It can be observed that all the updated parameters keep limited values; moreover, it can be recognized that there exists a correlation between the values of the elements of the Q matrix and the object trajectory. In particular,

the pick values of the elements of $\mathbf{Q}$ corresponding to the position and orientation components can be related to the pick values of the linear and angular velocity of the object. Analogously, the pick values of the elements of $\mathbf{Q}$ corresponding to the linear and angular velocity can be related to the object accelerations.

VIII. CONCLUSION

In this paper an algorithm for the visual motion estimation of the position and orientation of a moving object of known geometry has been proposed. The algorithm fully exploits the prediction capability of the Extended Kalman Filter for the pre-selection of the object features to be extracted from the image. An adaptive version of the EKF has been designed, which is capable of automatically tuning the statistics of the observation noise and of the state noise. The experimental results on a single-camera visual system confirm the effectiveness of the AEKF. In fact, the effects on the estimation error of the modelling error and of the measurement noise are reduced with respect to the non-adaptive formulation, at the expense of a small increase of computational load. Further experiments aimed at evaluating the effects on the tracking error of the update laws for the matrix $\mathbf{Q}$ and $\mathbf{R}$ separately can be found in [20].

Acknowledgments

This work was supported by MIUR and ASI.

REFERENCES

- [1] W.J. Wilson, C.W. Hulls, and G. Bell, "Relative end-effector control using cartesian position based visual servoing," *IEEE Trans. on Robotics and Automation*, vol. 12, pp. 684–696, 1996.
- [2] B. Thuilot, P. Martinet, L. Cordesses, and J. Gallice, "Position based visual servoing: keeping the object on the field of vision," *Proceedings of 2002 IEEE International Conference on Robotics and Automation*, Washington D.C., pp. 1624–1629, 2002.
- [3] T. Broida and R. Chellappa, "Estimation of object motion parameters from noisy images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 1, pp. 90–99, 1986.
- [4] A. Chiuso, P. Favaro, H. Jin, and S. Soatto, "Structure from motion casually integrated over time," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 1–13, 2002.
- [5] S. Lee and Y. Kay, "An accurate estimation of 3-D position and orientation of a moving object for robot stereo vision: Kalman filter approach," *Proceedings of 1990 IEEE International Conference on Robotics and Automation*, Cincinnati, OH, pp. 414–419, 1990.
- [6] J. Wang and W.J. Wilson, "3D relative position and orientation estimation using Kalman filter for robot control," *Proceedings of 1992 IEEE International Conference on Robotics and Automation*, Nice, France, pp. 2638–2645, 1992.
- [7] M. Bai, D.H. Zhou, and H. Schwarz, "Adaptive augmented state feedback control for an experimental planar two-link flexible manipulator," *IEEE Transactions on Robotics and Automation*, vol. 14, pp. 894–901, 1998.
- [8] K.J. Bradshaw, I.D. Reid, and D.W. Murray, "The active recovery of 3D motion trajectories and their use in prediction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 3, pp. 219–233, 1997.
- [9] P. Wira, J.P. Urban, "A new adaptive Kalman filter applied to visual servoing tasks," *Proceedings of 4th International Conference on Knowledge-Based Intelligent Engineering Systems and Allied Technologies*, Brighton, UK, pp. 267–270, 2000.
- [10] K.A. Myers and B.D. Tapley, "Adaptive sequential estimation with unknown noise statistics," *IEEE Transactions on Automatic Control*, vol. 21, pp. 520–523, 1976.
- [11] V. Lippiello, B. Siciliano, and L. Villani, "A new method of image features pre-selection for real-time pose estimation based on Kalman filter," *Proceedings of 2002 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Lousanne, Switzerland, pp. 372–377, 2002.
- [12] V. Lippiello and L. Villani, "Managing redundant visual measurements for accurate pose tracking," *Robotica*, vol. 21, pp. 511–519, 2003.
- [13] A.A. Girgis and W.L. Peterson, "Adaptive estimation of power system frequency deviation and its rate of change for calculating sudden power system overloads," *IEEE Transactions on Power Delivery*, vol. 5, pp. 585–594, 1990.
- [14] L. Jetto, S. Longhi, and G. Venturini, "Development and experimental validation of an adaptive extended Kalman filter for the localization of a mobile robot," *IEEE Transactions on Robotics and Automation*, vol. 15, pp. 219–229, 1999.
- [15] M. Ficocelli and F. Janabi-Sharifi, "Adaptive filtering for pose estimation in visual servoing," *Proceedings of the 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Maui, HI, pp. 19–24, 2001.
- [16] J. Weng, P. Cohen, and M. Herniou, "Camera calibration with distortion models and accuracy evaluation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, pp. 965–980, 1992.
- [17] J.T. Feddema, C.S.G. Lee, and O.R. Mitchell, "Weighted selection of image features for resolved rate visual feedback control," *IEEE Transactions on Robotics and Automation*, vol. 7, pp. 31–47, 1991.
- [18] F. Janabi-Sharifi and W.J. Wilson, "Automatic selection of image features for visual servoing," *IEEE Transactions on Robotics and Automation*, vol. 13, pp. 890–903, 1997.
- [19] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, pp. 679–698, 1986.
- [20] V. Lippiello, B. Siciliano and L. Villani, "Adaptive Extended Kalman Filtering for Visual Motion Estimation of 3D Objects," submitted to *IEEE Transactions on Control Systems Technology*, 2004.