Bae *et al. Robomech J (2016) 3:15*
DOI 10.1186/s40648-016-0054-2

ROBOMECH Journal

RESEARCH ARTICLE

Open Access

Vision and force/torque integration for realtime estimation of fast-moving object under intermittent contacts

Hyunki Bae, Soo Jeon* and Jan P. Huissoon

Abstract

This paper considers the fast and accurate estimation of motion variables of a rigid body object whose movement occurs from intermittent contacts with coordinating manipulators in nonprehensile manipulation tasks. The estimator operates under multiple sensory data including visual, joint torque, joint position and/or tactile measurements which are combined at the lower level to compensate for the latency and the slow sampling of the visual data. The estimator is real-time in the sense that it provides the motion data of the target object at the same fast sample rate of the servo controller without delay. The basic formulation is the multi-rate Kalman filter with the contact force vector as its process input, and the visual observation as its measurement output signal which is the down-sampled and delayed version of the configuration of the target object. Experimental tests are conducted for the case of planar object manipulation as well as the non-centroidal rotation under gravity using a robotic hand, and their results are presented to demonstrate the validity of the proposed estimation scheme.

Keywords: Coordinated manipulation, Kalman filter, Sensors and sensing systems

Background

Coordinating multiple manipulators for handling an object has potential to substantially enhance dexterity and versatility, the level of which is not possible by traditional stand-alone manipulators [1–3]. Although we expect that this will bring us numerous benefits, there are a number of technical issues to be resolved before we demonstrate the practical use of coordinated manipulation with human-like dexterity. From a hardware perspective, there has been much progress in recent years in new designs of complex manipulation systems such as anthropomorphic robotic hands that are capable of highly coordinated manipulation tasks with a large number of degrees of freedom [4–7]. On the other hand, there is a consensus (see, for example, *A Roadmap for US Robotics* [1]) that we are still far behind in some key technology areas such as perception, robust high fidelity sensing and planning and control to achieve highly dexterous object manipulation capabilities. As for the sensing in particular, it is recognized that processing sensory data with multiple modalities (e.g. vision, joint angle, tactile and force) will play a key role in synthesizing dexterous manipulation skills [1], which is clearly the case for human dexterity as we have ample evidence from neuroscience [8, 9].

This paper considers the situation where we want to keep track of the movement of the object in realtime, while the object is being manipulated by intermittent contact forces from the coordinating manipulators. Such scenarios occur when we attempt to manipulate a rigid body object through, so called nonprehensile manipulation [10] which refers to a class of dexterous manipulation tasks that make an explicit use of the dynamic forces (e.g. inertial, gravitational, or centrifugal) to allow extra motion freedoms to the object. Typical examples of nonprehensile manipulation include pushing, throwing, caging, flipping, twirling, etc. In conventional approaches to coordinated (or dexterous) manipulation [11], the object is being held stably through force-closure condition during manipulation. The force closure condition ensures

*Correspondence: soojeon@uwaterloo.ca
Department of Mechanical and Mechatronics Engineering, University of Waterloo, 200 University Ave W, Waterloo, ON N2L 3G1, Canada

that the object and manipulating bodies are kinematically related throughout the manipulation process, and thus we can easily keep track of the movement of the object using the associated kinematic relations. On the contrary, in the nonprehensile situations, the movement of the target object cannot be inferred from the kinematic relation because the object frequently loses its contact with some or all of its manipulating bodies, making the contact force applied in an intermittent way. Therefore, an additional sensor such as a vision camera is often necessary to provide the information of object motions in nonprehensile cases. The visual data, however, may not be enough to provide all the information in a timely manner. The vision may be blocked from a certain direction (occlusion) [12]. Also, the movement data obtained from visual images are often slow and delayed (latency) [13, 14], thereby restricting the speed and the accuracy of achievable object manipulation tasks, especially when the object is moving relatively fast, as is often the case with nonprehensile tasks mentioned above.

The main idea of this paper is to design a real-time rigid body estimator by augmenting the visual data with other sensory inputs available from manipulating bodies including the joint force/torque sensor or the tactile sensor mounted at the surface of the end-effector. The estimator is termed "real-time" in the sense that the visual feedback and other sensory data are updated and utilized to guide the manipulators at the same rate of their motion controller (or the servo controller), which is usually much faster than the update rate of the vision data. Therefore, the information fusion occurs at the very low level of the sensory data processing. Such a technical issue has been addressed either by some algorithmic approach (see, for example, [13] and references therein) or by sensor-based approach [14]. In this paper, we try to combine different types of sensors but make explicit use of the physical model for the rigid body dynamics requiring the determination of contact forces in detail as a necessary step to achieve a good estimation performance.

Combining the vision with force/torque or tactile data has been tried by many researchers in recent years [15–19], mainly for task-orientated (i.e. high-level) operations of robotic manipulators such as opening doors [18] or grasping a stationary object [15, 19]. Because the time scale of these tasks is much larger than that of the controller, the timing issue (i.e. the latency of the visual data) is not critical in these applications. On the other hand, the nonprehensile manipulations mentioned above often involve much faster time scale associated with the object motion. Thus the compensation of the visual latency will be important for accurate and fast estimation of the object motion, especially when we use the estimated variables for real-time control purpose. In fact, some pioneering works on nonprehensile manipulation share the motivation with our study in this respect. Specifically, the work in [20] considered a concept called the dynamics matching to close the visual feedback loop at the same rate (i.e. 1 ms) as the other sensory inputs using a specially-designed high-speed vision system.

The remainder of this paper is organized as follows. "Preliminaries" section presents the mathematical description of general rigid body motion and the determination of contact forces using the contact kinematics. The main formulation of the multirate rigid body estimator is described in "Estimator design" section and their experimental results are presented in "Experimental results" section followed by the summary of the results in "Conclusions and discussions" section.

Preliminaries

Description of rigid body motion

Figure 1 shows the schematic diagram of a rigid body in 3D space under contact forces and the gravitational force. S and B denote the spatial and the body coordinate frame, respectively. The total mass and the inertia tensor of the object, evaluated at the center of mass, are denoted by m and $\mathcal{I}$, respectively, where $\mathcal{I}$ is a diagonal matrix with moments of inertia J_1, J_2 and J_3, i.e. $\mathcal{I}_0 = \mathrm{diag}(J_1, J_2, J_3)$ when its bases lie along the principal axes. Then, the general form of rigid body dynamics can be written as [11]

$$\mathcal{M}\dot{V}^b + \mathrm{diag}([\omega^b \times], [\omega^b \times])\mathcal{M}V^b = F^b \tag{1}$$

where the superscript b indicates that the corresponding quantities are written with respect to the body coordinate frame attached to the object mass center (shown as B in Fig. 1). $\mathcal{M} = \mathrm{diag}(mI_3, \mathcal{I}_0) \in \mathbb{R}^{6\times 6}$ is the generalized inertia matrix (with $I_3 \in \mathbb{R}^{3\times 3}$ denoting the identity matrix). The notation $[\bullet\times] \in \mathbb{R}^{3\times 3}$ is the skew-symmetric matrix form of the cross product, i.e. $[a\times]b = a \times b$. The vector $V^b = \mathrm{col}\left(v^b, \omega^b\right) \in \mathbb{R}^6$ is the twist form for the generalized body velocity where $v \in \mathbb{R}^3$ is the translational velocity, and $\omega \in \mathbb{R}^3$ is the angular velocity. $F^b \in \mathbb{R}^6$ is the resultant wrench applied at the center of mass, which can further be written with respect to those of each manipulator. Namely, denoting by n_p the total number of end-effectors, $F^b = \sum_{i=1}^{n_p} F_i^b = \sum_{i=1}^{n_p} \mathrm{col}\left(f_i^b, \tau_i^b\right)$ where $f_i^b \in \mathbb{R}^3$ and $\tau_i^b \in \mathbb{R}^3$ $(i = 1, 2, \ldots n_p)$ are the force and the torque vectors, respectively, for the ith contact point (denoted by p_{c_i} in Fig. 1). Note that f_i can be either a contact force or the gravitational force. The point of application, p_{c_i}, of the gravitational force will be located at the center of mass (see Fig. 1). We consider the situation where the object is freely moving without any

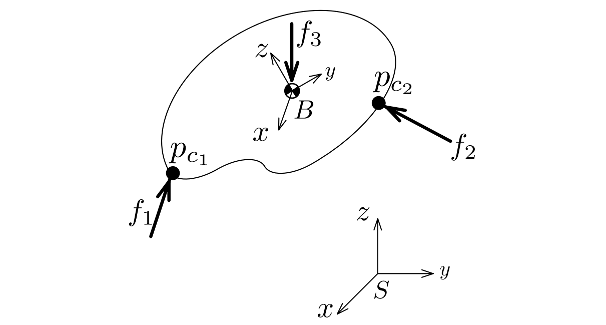


Fig. 1 Coordinate frames for rigid body motion and forces

environmental or disturbance forces (such as friction) except for the gravity.

If we describe v of Eq. (1) in the spatial coordinate frame (indicated by the superscript 's') and separate the dynamics into a translational part and a rotational one, we have

$$m\dot{v}^s = \sum f_i^s = R\sum_{i=1}^{n_p} f_i^b \tag{2a}$$

$$\mathcal{I}_0\dot{\omega}^b + \omega^b \times \mathcal{I}_0\omega^b = \sum \tau_i^b. \tag{2b}$$

where $R \in SO(3)$ denotes the rotation matrix for the configuration of the body coordinate frame. Recall that the rigid body kinematics is given by

$$\dot{p}^s = v^s = Rv^b \tag{3a}$$

$$\dot{R} = R[\omega^b \times] \tag{3b}$$

where p^s is the position of the mass center (with respect to the spatial frame).

To use the state-space approach, we convert Eq. (3b) into a vector form. Euler angles can be used to represent Eq. (3b) in the state space form but will introduce singularity. One singularity-free coordination is the unit quaternion; $q = \begin{bmatrix} q_0 & e^T \end{bmatrix}^T \in \mathbb{R}^4$ where $q_0 \in \mathbb{R}$ is the scalar part and $e = [q_1 \;\; q_2 \;\; q_3]^T \in \mathbb{R}^3$ is the vector part with $\|q\|_2 = 1$. At the cost of increased dimension, the quaternion can be used to replace Eq. (3b) with

$$\dot{q} = H_q(\omega^b)q = \frac{1}{2}\begin{bmatrix} 0 & -\omega_1^b & -\omega_2^b & -\omega_3^b \\ \omega_1^b & 0 & \omega_3^b & -\omega_2^b \\ \omega_2^b & -\omega_3^b & 0 & \omega_1^b \\ \omega_3^b & \omega_2^b & -\omega_1^b & 0 \end{bmatrix}\begin{bmatrix} q_0 \\ q_1 \\ q_2 \\ q_3 \end{bmatrix}. \tag{4}$$

The rotation matrix can be retrieved from the quaternion as $R = (q_0^2 - e^Te)I_3 + 2ee^T + 2q_0[e\times]$ [22]. Equations (2), (3a) and (4) realize the state space representation of rigid body motion in general 3D which includes 13 state variables in total for general 3D motion. For the planar motion (i.e. 2D), the formulation can be simplified such that the number of state variables is reduced to 6 (see "Experimental results" section).

In general, the dynamics of the object may behave in a little different way if the contact forces are applied in a non-smooth or impactive way [25]. However, in practice, we may not encounter such an ideal impact because of the softness (or the flexibility) of finger tips and the end-effectors.

Determination of contact force

Often times, the contact force f_i in Fig. 1 cannot be directly measured but rather needs to be estimated using proprioceptive sensory data (e.g. joint angles, torques and/or tactile sensors) available from each manipulator. To explain specific ways of doing this, we briefly review the contact kinematics and coordinate transformation of generalized forces in this subsection. More detailed treatment for mathematical formulations of grasping and contact forces can be found in well-known textbooks such as [11] and [21].

To keep our analysis general, we consider the 3D case. Figure 2 shows the schematic diagram of the i-th manipulator (finger) contacting the object at the point p_{c_i}. D_i denotes the coordinate frame for the base of the i-th manipulator that is fixed relative to S. E_i is the coordinate frame for the reference end-effector that is attached to the terminal link. We will use m_i to denote the number of joints in the ith manipulator. The joint space vector for the manipulator is denoted by $\theta_i = \left[\theta_{i,1}, \ldots \theta_{i,m_i}\right]^T$ which is in

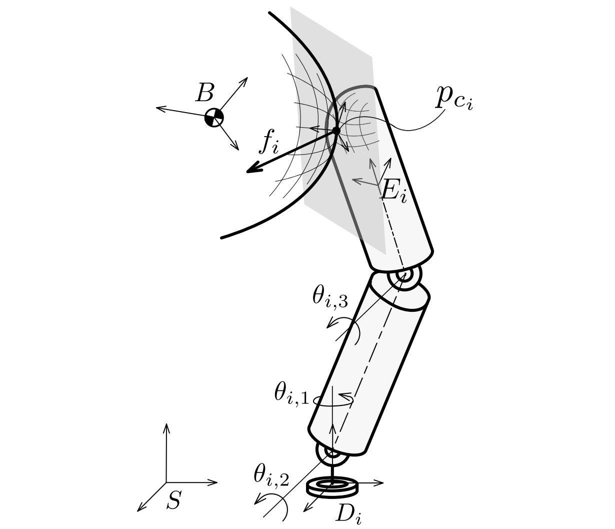


Fig. 2 Coordinate frames and variables for rolling contact kinematics

m_i-torus $\mathbb{T}^{m_i} = \mathbb{S}^1 \times \cdots \times \mathbb{S}^1$. For the sake of illustration, we depicted the robotic finger as a 3 DOF (degrees-of-freedom) elbow manipulator, i.e. $m_i = 3$, in Fig. 2.

At first, the contact point p_{c_i} can be computed from the joint angles of the manipulator and the tactile sensor. Associated with the contact point p_{c_i} are two manifolds that describe the surface geometry of the object and the end-effector, the local coordinate charts of which are denoted by c_c and c_t, respectively (see Fig. 3). The local coordinate chart (e.g. c_c) takes a point on the surface (e.g. $(u_c, v_c) \in U_c \subset \mathbb{R}^2$) and maps it to a point on the manifold (which is embedded in $\mathbb{R}^3$). For a given contact point and associated coordinate charts, the contact force can be described with respect to the contact frame which we denote by C_i here. C_i is an instantaneous coordinate system constituted by two tangent vectors ($\frac{\partial c_c}{\partial u_c}$ and $\frac{\partial c_c}{\partial v_c}$) and their outward normal (n_c) at p_{c_i} with respect to the object. See Fig. 3 for the components of C_i, which are illustrated by arrows with black triangular heads. We can also define the end-effector counterpart for the contact frame (denoted by T_i) with associated coordinates, $\frac{\partial c_t}{\partial u_t}$, $\frac{\partial c_t}{\partial v_t}$ and n_t which are illustrated by arrows with white triangular heads in Fig. 3. Note that C_i and T_i are opposite in normal direction and their tangential pairs are rotated by the contact angle ψ. Namely, $f_i^{c_i} \in \mathbb{R}^3$ (f_i written about C_i) is related to $f_i^{t_i} \in \mathbb{R}^3$ (f_i written about T_i) as

$$f_i^{c_i} = -R_{c_i t_i} f_i^{t_i}, \quad R_{c_i t_i} = \begin{bmatrix} \cos\psi & \sin\psi & 0 \\ \sin\psi & -\cos\psi & 0 \\ 0 & 0 & -1 \end{bmatrix}. \quad (5)$$

Our objective here is to represent the body wrench $F_i^b := \mathrm{col}\left(f_i^b, \tau_i^b\right)$ with respect to the joint torque vector denoted by $\Upsilon_i = \begin{bmatrix}\Upsilon_{i,1} & \cdots & \Upsilon_{i,m_i}\end{bmatrix}^T \in \mathbb{R}^{m_i}$ (e.g. $\Upsilon_{i,2}$ is the joint torque for $\theta_{i,2}$). Note that F_i^b is related to $f_i^{c_i}$ and its wrench form $F_i^{c_i}$ as

$$F_i^b = \mathrm{Ad}^T_{g^{-1}_{bc_i}} \Lambda f_i^{c_i} = \mathrm{Ad}^T_{g^{-1}_{bc_i}} F_i^{c_i} \quad (6)$$

where $\mathrm{Ad}^T_{g^{-1}_{bc_i}} \in \mathbb{R}^{6\times 6}$ is the transpose of the adjoint transformation associated with the $g^{-1}_{bc_i}$ and $\Lambda = [I_3 \quad 0_{3\times 3}]^T$ is the wrench basis which encodes the force vector for frictional point contact (which is in $\mathbb{R}^3$) into the corresponding wrench form of the contact coordinate frame C_i (which is in $\mathbb{R}^6$). The adjoint transformation denoted by $\mathrm{Ad}_\bullet \in \mathbb{R}^{6\times 6}$ (where $\bullet \in SE(3)$) is a linear mapping that operates the coordinate transformation of a wrench vector between two different coordinates (from C_i to B in this case). Also, $g^{-1}_{bc_i}$ is the inverse of the homogeneous transformation $g_{bc_i} \in \mathbb{R}^{4\times 4}$ which belongs to $SE(3)$ and maps a vector in C_i coordinate into that in B coordinate. (see [11] for more information about the homogeneous transformation and the adjoint transformation.)

Similarly, $f_i^{t_i}$ and its wrench form $F_i^{t_i} = \Lambda f_i^{t_i}$ can be written with respect to the base frame D_i as

$$F_i^{d_i} = \mathrm{Ad}^T_{g_{t_i d_i}} F_i^{t_i} = \mathrm{Ad}^T_{g^{-1}_{d_i t_i}} \Lambda f_i^{t_i} = \mathrm{Ad}^T_{g^{-1}_{d_i e_i}} \mathrm{Ad}^T_{g^{-1}_{e_i t_i}} \Lambda f_i^{t_i} \quad (7)$$

where the last equality follows from the composition rule of the adjoint transformation. Finally, from the standard result of kinematic chain manipulators [11],

$$\Upsilon_i = \left(J^s_{d_i t_i}\right)^T F_i^{d_i} \quad (8)$$

where $J^s_{d_i t_i} : \mathbb{R}^{m_i} \to \mathbb{R}^6$ is the spatial manipulator Jacobian for the ith manipulator evaluated with T_i as the terminal frame and D_i as the base frame. Combining Eqs. (7) and (8), we have

$$\begin{aligned} f_i^{t_i} &= J^\dagger_{h_i}(\theta_i, p_{c_i})\Upsilon_i \\ &= \left(\left(J^s_{d_i e_i}(\theta_i)\right)^T \mathrm{Ad}^T_{g^{-1}_{d_i e_i}}(\theta_i)\mathrm{Ad}^T_{g^{-1}_{e_i t_i}}(p_{c_i})\Lambda\right)^\dagger \Upsilon_i \end{aligned} \quad (9)$$

where $J_{h_i} \in \mathbb{R}^{m_i \times 3}$ is called the hand Jacobian for the i-th manipulator [11] and the notation $\bullet^\dagger$ stands for the pseudo-inverse of $\bullet$. The body wrench F_i^b is then related to Υ_i by Eqs. (5), (6) and (9). For $m_i \geq 3$, J_{h_i} is either a square or a tall matrix and thus its pseudo-inverse exists uniquely as long as it is full rank. If J_{h_i} is not full rank, the manipulator may be in a singular configuration and needs a special treatment, which is beyond the scope of this paper.

From Eq. (9), it is clear that we need at least three joint torque sensors to fully determine $f_i^{t_i}$ in 3D. If the tactile sensing is available to provide the contact pressure, $f_{i,3}^{t_i}$ can be directly obtained from the tactile sensor. This

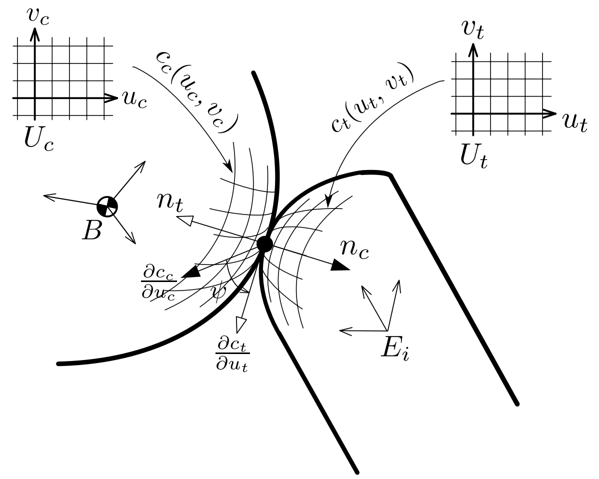


Fig. 3 Definition of coordinate frames at the contact point. (Note that $\frac{\partial c_c}{\partial v_c}$ and $\frac{\partial c_t}{\partial v_t}$ are omitted here for the brevity of illustration)

allows us to use only two joint torque sensors to figure out the remaining components for tangential direction, i.e. $f_{i,1}^{t_i}$ and $f_{i,2}^{t_i}$. If the manipulator is deficient in sensing, the contact force may not be fully determined in general. However, in some special cases, the contact force may be identified reasonably well even with fewer number of sensors. Manipulation by an impactive force is one example where the direction of the contact force is related to the pre-impact velocity. If the direction of the contact force vector can be known (for example by the vision sensor), we only need to find its magnitude which we may infer with just one joint torque sensor. An example of such cases is considered in the experimental tests.

Estimator design

In this section, we derive the rigid body estimator for the general 3D case, using the rigid body dynamics and the contact force measurement equations presented in the previous subsections. Multimodal sensory integration will also be put into perspective through the formulation of a new state estimator combining signals from the visual image with the joint torque sensor and/or the tactile data.

State space description

Denoting the state vectors for the translation and the rotation by ξ and η, respectively, (i.e. $\xi = \mathrm{col}(p^s, v^s) \in \mathbb{R}^6$, $\eta = \mathrm{col}\big(q, \omega^b\big) \in \mathbb{R}^7$), we can write Eqs. (2), (3a) and (4) in the state equation as

$$\dot{\xi} = F_t \xi + G_t\big(u(t) + w(t)\big) \tag{10a}$$

$$\dot{\eta} = F_r(t)\eta + G_r(t)\big(u(t) + w(t)\big) \tag{10b}$$

where the subscripts t and r denote the translation and rotation, respectively, $u(t) = \mathrm{col}\Big(f_1^b(t), \ldots, f_{n_p}^b(t)\Big) \in \mathbb{R}^{3n_p}$ is the force input and $w(t) \in \mathbb{R}^{3n_p}$ represents the corresponding process noise. The system matrices are given by

$$F_t = \begin{bmatrix} 0_{3\times3} & I_3 \\ 0_{3\times3} & 0_{3\times3} \end{bmatrix}, \quad G_t = \frac{1}{m}\begin{bmatrix} 0_{3\times3n_p} \\ R\begin{bmatrix} I_3 & \cdots & I_3 \end{bmatrix} \end{bmatrix}$$

$$F_r(t) = \begin{bmatrix} H_q(\omega^b(t)) & 0_{4\times3} \\ 0_{3\times4} & -\mathcal{I}_0^{-1}\big[\omega^b(t)\times\big]\mathcal{I}_0 \end{bmatrix}, \tag{11}$$

$$G_r(t) = \begin{bmatrix} 0_{4\times3n_p} \\ \mathcal{I}_0^{-1}\Big[\big[p_{c_1}^b(t)\times\big] \;\cdots\; \big[p_{c_{n_p}}^b(t)\times\big]\Big] \end{bmatrix}.$$

If the i-th end-effector is not in contact with the object, both $p_{c_i}^b(t)$ and f_i^b can simply be set to zero.

Note that the translational dynamics depends on the rotation matrix $R(t)$ to be estimated using the attitude dynamics. Similarly, the system matrix for the attitude dynamics, $F_r(t)$, is a function of the state variable ω^b and the attitude input matrix G_r is a function of $\big[p_c^b(t)\times\big]$ which needs to be estimated from the translational motion. The state dependency of these system matrices can be handled by the pseudo-linear formulation [23, 24] which simply replaces the actual state variables (i.e. ω^b, R, and $\big[\omega^b(t)\times\big]$) in the system matrices by the estimated ones.

Representing the time instant $t = t_k := kT_s$ by the integer k, the discretetized form of Eq. (10) is

$$\xi(k+1) = \Phi_t \xi(k) + \Gamma_t(k)\big(u(k) + w(k)\big) \tag{12a}$$

$$\eta(k+1) = \Phi_r(k)\eta(k) + \Gamma_r(k)\big(u(k) + w(k)\big) \tag{12b}$$

where the system matrices $\Phi_\bullet$ and $\Gamma_\bullet$ are computed from $F_\bullet$ and $G_\bullet$, respectively, by the zero-order-hold (ZOH) assuming that signal values are almost constant within the base sample time, T_s.

The input signal $u(t)$ is available from joint torque sensors and/or tactile sensors with a fast sample rate T_s. On the other hand, the position and orientation of the rigid body can be measured through a stereo vision system at much slower sample rate and with latency due to the vision processing time. The corresponding output equations can be described as

$$y_t(k) = p^s(k - N_v) = C_t \xi(k - N_v) + \nu_t(k - N_v) \tag{13a}$$

$$= \begin{bmatrix} I_3 & 0 \end{bmatrix} \xi(k - N_v) + \nu_t(k - N_v) \tag{13b}$$

$$y_r(k) = q(k - N_v) = C_r \eta(k - N_v) + \nu_r(k - N_v) \tag{13c}$$

$$= \begin{bmatrix} I_4 & 0_{4\times3} \end{bmatrix} \eta(k - N_v) + \nu_r(k - N_v) \tag{13d}$$

for $k = N_v \ell$ ($\ell = 0, 1, 2, \ldots$) and zero otherwise, where N_v is the integer multiple of T_s that relates the vision time step T_v with the base sample time such that $T_v = N_v T_s$. Equations (12) and (13) complete the formulation of the system dynamics and the associated sensory data in the state space form.

Due to the multi-rate nature of the system equations given in Eqs. (12) and (13), the estimation process consists of two parts: the prediction step at every fast sample time T_s, and the correction step at every vision sample time T_v. The key point is that the correction is operated in a *retrospective way* to fully recover the measurement delay coming from the delayed vision data, thereby ensuring that the estimates reflect the current (real-time) state information. In the remaining part of the paper, the bar notation, $\bar{\bullet}$ will be used to denote the values measured by the sensors while the hat notation, $\hat{\bullet}$, will be used to denote the estimated values.

Multirate estimation combining vision and contact force

As will be shown later, the measurement delay can be handled in the last stage. So we first derive the estimator without considering the output measurement delay here. Assume that we have the best estimates of ξ and η at the discrete-time step k based on the vision data at the same time step, which we denote as $\hat{\xi}(k|k)$ and $\hat{\eta}(k|k)$, respectively. Then, until the next vision data become available at the time step $k+N_v$, the state variables are predicted by propagating Eq. (12) using the contact force vector as an input signal

$$\hat{\xi}(k+i+1|k) = \Phi_t\hat{\xi}(k+i|k) + \Gamma_t(k)\bar{u}(k+i) \tag{14a}$$

$$\begin{aligned}\hat{\eta}(k+i+1|k) = {} & \Phi_r(k)\hat{\eta}(k+i|k)\\ & + \Gamma_r(k)\bar{u}(k+i)\end{aligned} \tag{14b}$$

where $i = 0, 1, ..., N_v - 1$. k is the discrete-time index for the moment at which the last correction was made by the vision data. $\bar{u}(k) = u(k) + w(k)$ is the forces applied at the contact point(s) which can be measured by the force-torque and/or tactile sensors on fingers using the kinematic relation provided in the previous section. The sensor noise $w(k) \in \mathbb{R}^{3n_p}$ is assumed independent, Gaussian distributed, and stationary with its covariance denoted by $\mathbb{E}\left[w(k)w(k)^T\right] = W \in \mathbb{R}^{3n_p \times 3n_p}$. Note that the state-dependent system matrices Γ_t, Φ_r and Γ_r are updated by the vision sample rate T_v in Eq. (14) (i.e. they are updated at the time step k and kept constant for $i = 0, 1, ..., N_v - 1$). However, it is also possible to update them by the base sample rate T_s using the predicted state variables.

During the prediction step given in Eq. (14), the estimation error will be accumulated. Denote $Z(k|k)$ as the estimation error covariance at time k right after updating by the kth image. Then, $Z(k|k)$ propagates through the prediction process in Eq. (14) as

$$\begin{aligned}Z_\bullet(k+i+1|k) = {} & \Phi_\bullet(k+i)Z_\bullet(k+i|k)\Phi_\bullet(k+i)^T\\ & + \Gamma_\bullet(k+i)W\Gamma_\bullet(k+i)^T\end{aligned} \tag{15}$$

where the subscript $\bullet$ indicates either t or r in the remainder of the paper. When the new image is obtained at time $k+N_v$, the Kalman filter gains can be computed for the translational and rotational motions as

$$\begin{aligned}L_\bullet(k+N_v) = {} & Z_\bullet(k+N_v|k)C_\bullet^T\\ & \times \left[C_\bullet Z_\bullet(k+N_v|k)C_\bullet^T + V_\bullet\right]^{-1}\end{aligned} \tag{16}$$

where $V_\bullet$ is the error covariance matrix for the output noise signal $v_\bullet$ (v_t or v_r), i.e. $\mathbb{E}\left[v_\bullet(k)v_\bullet(k)^T\right] = V_\bullet$. The measurement noise $v_\bullet$ is assumed to be stationary, independent and Gaussian distributed. Then the update equations can be written as

$$\begin{aligned}&\hat{\xi}(k+N_v|k+N_v)\\ &\quad = \hat{\xi}(k+N_v|k) + L_t(k+N_v)\\ &\quad \times \left(y_t(k+N_v) - C_t\hat{\xi}(k+N_v|k)\right)\end{aligned} \tag{17a}$$

$$\begin{aligned}&\hat{\eta}(k+N_v|k+N_v)\\ &\quad = \hat{\eta}(k+N_v|k) + L_r(k+N_v)\\ &\quad \times \left(y_r(k+N_v) - C_r\hat{\eta}(k+N_v|k)\right).\end{aligned} \tag{17b}$$

Note that the estimated quaternion variables in $\hat{\eta}(k|k)$ needs an additional step for normalization (see "Description of rigid body motion" section). Finally, the covariance of the estimation error $Z(k+N_v|k)$ is corrected for the next recursive step:

$$\begin{aligned}Z_\bullet(k+N_v|k+N_v) = {} & Z_\bullet(k+N_v|k)\\ & - L_\bullet(k+N_v)C_\bullet Z_\bullet(k+N_v|k).\end{aligned} \tag{18}$$

Until now, we assumed the output signal is measured without any delay. However, the measurement of the vision sensor at the time step k is, in fact, obtained from the camera image taken at the time step $k - N_v$. Therefore, the prediction step Eq. (14) needs to be changed to

$$\begin{aligned}\hat{\xi}(k+i+1|k-N_v) = {} & \Phi_t\hat{\xi}(k+i|k-N_v)\\ & + \Gamma_t(k-N_v)\bar{u}(k+i)\end{aligned} \tag{19a}$$

$$\begin{aligned}\hat{\eta}(k+i+1|k-N_v) = {} & \Phi_r\hat{\eta}(k+i|k-N_v)\\ & + \Gamma_r(k-N_v)\bar{u}(k+i).\end{aligned} \tag{19b}$$

Likewise, the propagation of the estimation error covariance given by Eq. (15) changes to

$$\begin{aligned}&Z_\bullet(k+i+1|k-N_v)\\ &\quad = \Phi_\bullet(k+i)Z_\bullet(k+i|k-N_v)\Phi_\bullet(k+i)^T\\ &\quad + \Gamma_\bullet(k-N_v)W\Gamma_\bullet(k-N_v)^T\end{aligned} \tag{20}$$

and the filter gain in Eq. (16) also changes to

$$\begin{aligned}L_\bullet(k+N_v) = {} & Z_\bullet(k+N_v|k-N_v)C_\bullet^T\\ & \times \left[C_\bullet Z_\bullet(k+N_v|k-N_v)C_\bullet^T + V_\bullet\right]^{-1}.\end{aligned} \tag{21}$$

By the time we arrive at the time step $k+N_v$ and compute Eq. (21), we now know what values of $p^s(k)$ $(= y_t(k+N_v))$, and $q(k)$ $(= y_r(k+N_v))$ should have been. This means that we can make the correction for the N_v time step prior to the current time step and compute $\hat{\xi}(k|k)$ and $\hat{\eta}(k|k)$ *in retrospect*. As soon as we compute $\hat{\xi}(k|k)$ and $\hat{\eta}(k|k)$, we also know what values of $\hat{\xi}(k+i+1|k)$ and $\hat{\eta}(k+i+1|k)$ should have been for $i=0, 1, ..., N_v-1$, and can compute them all at once. All these estimates are for the past and thus useless except for the last pair, i.e. $\hat{\xi}(k+N_v|k)$ and $\hat{\eta}(k+N_v|k)$. So we should take them as the estimate for the current time step, $k+N_v$. This process repeats for the next batch of the prediction steps. In summary, the correction step shown in the previous subsection should be modified into the following set of procedures. It applies to both translation and rotation in the same way so we explain here with the translational motion only.

1. As soon as we arrive at the time step $k+N_v$, compute the Kalman filter gain for the time step k, i.e. $L_t(k)$ using Eq. (21). Then, using the delayed vision data $y_t(k+N_v)=p^s(k)$, make a correction for the time step k, i.e. $\hat{\xi}(k|k)$, by Eq. (17a). At the same time, make a correction to the estimation error covariance in a similar way as Eq. (18) but with the twice of the vision step as the update interval, i.e.

$$Z_t(k|k) = Z_t(k|k-2N_v) - L_t(k)C_tZ_t(k|k-2N_v). \tag{22}$$

   The reason for doubling the update interval will become clear at the end of step 3).
2. Compute $\hat{\xi}(k+1|k)$, ..., $\hat{\xi}(k+N_v|k)$ using the prediction equation Eq. (14a), after which we take only $\hat{\xi}(k+N_v|k)$ as the estimate for the current time step $k+N_v$. Similarly, compute the propagation of the estimation error covariance $Z_t(k+1|k)$, ..., $Z_t(k+N_v|k)$ in the same way as Eq. (15). This completes the *retrospective* correction for the time step $k+N_v$.
3. Next, we progress for the subsequent time steps $k+N_v+1$, ..., $k+2N_v$ in a similar way as Eq. (19a). At the same time, we continue computing the propagation of the estimation error covariance using Eq. (20), after which we have $Z_t(k+N_v|k-N_v)$. As a result, the update process operates by the estimation error accumulated during $2N_v$ time steps.
4. When we arrive at the next vision time step $k+2N_v$, now we have $y_t(k+2N_v)=p^s(k+N_v)$ and can compute the Kalman filter gain for the time step $k+N_v$, i.e. $L_t(k+N_v)$ using Eq. (21). Now, we are back to step 1) so can repeat the process.

One important point to be made here is that the integrity of the proposed estimator strongly relies on the accuracy of the contact force vector as well as the contact location. As discussed in "Determination of contact force" section, the accuracy of the contact force can be improved by redundant sensing. Similarly, the information on the contact point may be refined by the use of high-speed, high-fidelity tactile sensor arrays.

Experimental results

Testbed configuration

We conducted two experimental tests for the proof-of-concept of the rigid body estimation explained in the previous section. In the first experiment, the setup is configured as shown in Fig. 4 considering the planar 2-D motion of a target object. A square object is made of aluminum blocks and is placed on the air table such that it moves on the horizontal plane without any gravity and friction. A robot hand (BHand), manufactured by *Barrett Technologies Inc.*, is mounted to the side of the table so that the robotic fingers can manipulate the object for its planar motion. A vision camera is mounted above to take images of the object while it is moving. Figure 5 shows the sample view from the camera, which indicates, with added labels, the index number for each finger and the locations of some reference frames that we defined in "Determination of contact force" section. The body coordinate frame B is located at the center of the square object. The spatial (inertial) frame is attached to the center of the palm of the BHand and D_1 (the base frame for finger 1) is placed at the inner joint (first joint) of the finger 1.

The specifications of vision sensor is shown in Table 1. The acquisition timing of the image is precisely

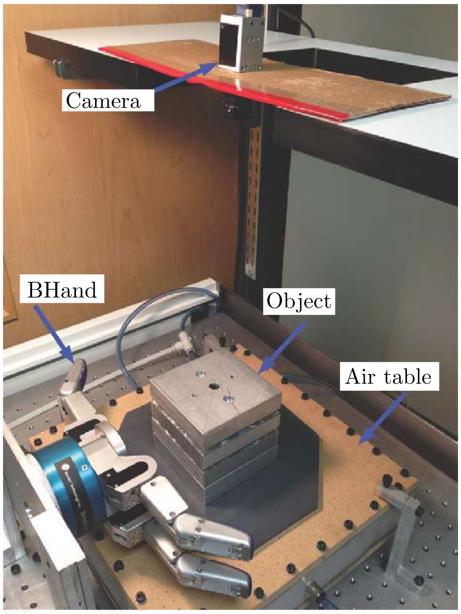


Fig. 4 Configuration of testbed for the case with the object moving on the horizontal plane

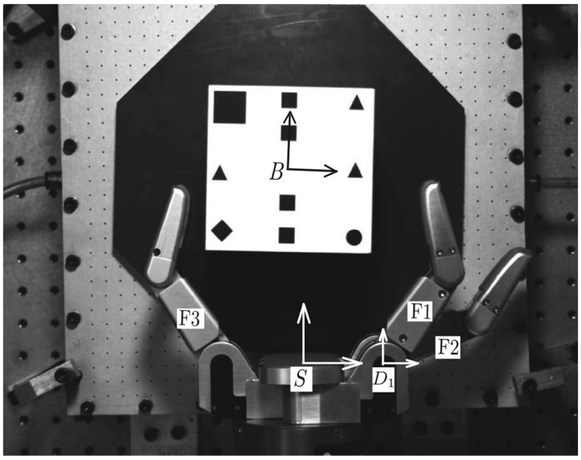


Fig. 5 A view from the vision camera for the case with the object moving on the horizontal plane free from gravity

Table 1 Specifications of vision

Specification	Values	Units
Communication	CameraLink™	
Exposure time	10	[ms]
Bit depth	8	Bits
Pixel size	2048 × 1088	Pixels
Resolution	284	μm/pixel

controlled by the external trigger signal coming from the real-time machine. The image processing is based on the simple pattern matching which provides the location of the object center about the spatial frame (i.e. p^s) as well as the rotation angle (denoted by ϕ in this section) of the body frame with respect to the spatial one.

During the experiment, the BHand is manipulated by an operator who controls a set of potentiometers each of which corresponds to the desired position of the first joint of each finger. Each finger takes this desired position as a reference command and runs the PID controller. The square object is then manipulated through contacts with the fingers. As shown in Fig. 5, each finger of the BHand has its configuration of a two link manipulator. However, the second (or the outer) joint is kinematically linked to the first one so only one reference signal is used to control the whole finger. In the remaining part of the paper, F1, F2 and F3 will denote the finger 1, the finger 2 and the finger 3, respectively.

In order to demonstrate the performance of the proposed estimator under the effect of gravitational force, we conducted another experiment as shown in Fig. 6 where the gravitational force is denoted by the downward arrow labeled with *mg*. The BHand (and the camera) is arranged in 90^o rotation from to the previous configuration. So the back plane of the wrist of the BHand is mounted on the horizontal table. Initially, the object is stably supported by the F3 and one of its corners against gravity as shown in Fig. 6. Then, the object is moved by the action of F3 making a non-centroidal rotation about its corner at the bottom until it gets caught by the F1 (and F2).

Estimator for 2D motion

The movement of the object is constrained to the 2D plane in both configurations of the experimental tests mentioned above. In such 2-D motions, the rotation has only one degree of freedom with no distinction between the spatial frame and the body frame (i.e. $\omega^b = \omega^s = \omega \in \mathbb{R}$) and thus Eq. (2b) is simplified to a linear form:

$$\mathcal{I}_0 \dot{\omega} = \sum \tau_i.$$

The dimension of the state space equation in Eq. (10) also reduces down to 6 from 13.

The surface of the finger has a geometric shape as shown in Fig. 7a. We took F1 as an example here but other fingers have the exact same geometry. The allowed contact area is indicated by the grey line on the surface of the finger. The grey-colored section of the finger consists of two arcs which have different radii and center points. The first section has a radius 105 mm with its centre point located at $(0, -95.5)$ written with respect to the end-effector frame E_1 shown in Fig. 7a. On the other hand, the second section has a radius 8 mm with its centre point located at $\left(97 \sin \frac{\pi}{18}, 0\right)$ written about E_1. (Note that $97 \sin \frac{\pi}{18} \approx 95.5 \tan \frac{\pi}{18} \approx 16.84$). Therefore, the coordinate chart $c_t : U_t \in \mathbb{R} \to \mathbb{R}^2$ for the grey-colored section of the finger can be written as

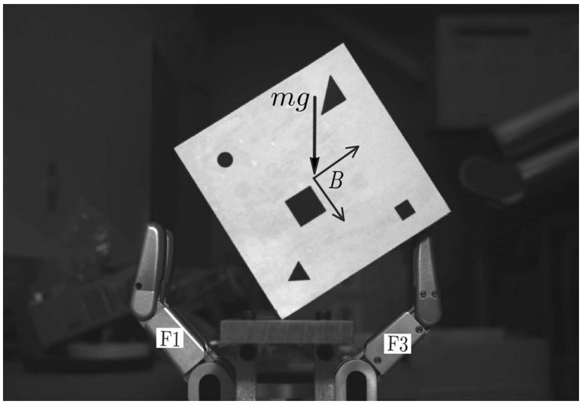


Fig. 6 A view from the vision camera for the case with the object moving under gravity

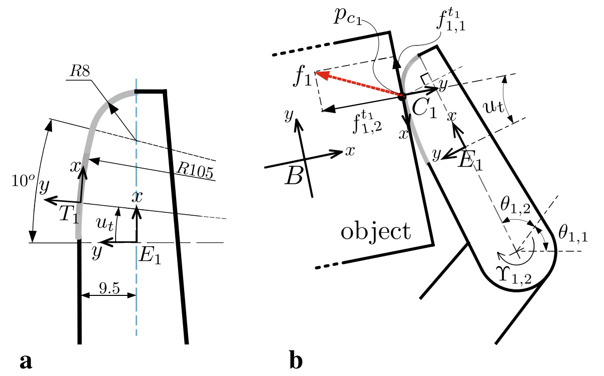


Fig. 7 Contact geometry and coordinate frames for contact force. **a** Geometry of contact area, **b** contact force configuration

$$c_t = \begin{cases} (105\sin u_t,\ 105\cos u_t - 95.5), & 0 \le u_t < \frac{\pi}{18} \\ \left(8\sin u_t + 97\sin\frac{\pi}{18},\ 8\cos u_t\right), & \frac{\pi}{18} \le u_t < \frac{\pi}{2} \end{cases} \tag{23}$$

where the $(x,\ y)$ coordinate values are written in [mm] with respect to the end effector frame E_1. Note that c_t parameterizes the regular surface geometry including the point $u_t = 10^o = \frac{\pi}{18}$. Defining c_t as in Eq. (23) also implies that T_1, i.e. the contact frame for the finger can be configured as shown in Fig. 7a.

Since the surface of the object aligns with x and y axes of the body coordinate frame, the coordinate chart for the object c_c can simply be taken as a mapping from one of these axes to the contact point represented by the body coordinate frame. Similarly, the contact frame C_1 for the object is configured such that its y-axis is the same as the x-axis of B for the situation shown in Fig. 7b. Figure 7b also illustrates how the contact force f_1 is represented with respect to the contact frame T_1 for the finger. More specifically, $f^{t_1}_{1,1}$ denotes the component of f_1 along the x-axis (or, equivalently, the direction of $\frac{\partial c_t}{\partial u_t}$) of T_1 and $f^{t_1}_{1,2}$ corresponds to its normal component.

Using the coordinate frames shown in Fig. 7, we can compute the spatial manipulator Jacobian $J^s_{d_1e_1}$ as

$$J^s_{d_1e_1}(\theta_1) = \begin{bmatrix} 0 & 0 & 0\ 0\ 0\ 1 \\ \ell_1 s_1 & -\ell_1 c_1 & 0\ 0\ 0\ 1 \end{bmatrix}^T \tag{24}$$

where ℓ_1 denotes the length of the inner link, $c_1 := \cos\theta_{1,1}$ and $s_1 := \sin\theta_{1,1}$. After going through some tedious but straightforward computation of associated adjoint matrices, we can compute the hand Jacobian for F1 as

$$\begin{aligned} J_{h_1} &= \left(J^s_{d_1e_1}(\theta_1)\right)^T \cdot \mathrm{Ad}^T_{g^{-1}_{e_1t_1}}(u_t) \cdot \mathrm{Ad}^T_{g^{-1}_{d_1e_1}}(\theta_1)\Lambda \\ &= \left[\begin{array}{c|c} \ell_1 s_{2-u_t} - c_{u_t}c_{t,2} & \ell_1 c_{2-u_t} - s_{u_t}c_{t,2} \\ \quad - s_{u_t}\left(\ell_2 + c_{t,1}\right) & \quad + c_{u_t}\left(\ell_2 + c_{t,1}\right) \\ \hline -s_{u_t}\left(\ell_2 + c_{t,1}\right) & c_{u_t}\left(\ell_2 + c_{t,1}\right) \\ \quad - c_{u_t}c_{t,2} & \quad - s_{u_t}c_{t,2} \end{array}\right] \end{aligned} \tag{25}$$

where ℓ_2 is the length from the outer link joint axis to the centre of E_1. $c_{t,1}$ and $c_{t,2}$ denote the first and the second element of c_t in Eq. (23), respectively. Similarly to the notation we used in Eq. (24), "c" and "s" denote the cosine and the sine functions, respectively, for the angles indicated by their subscripts, e.g. $s_{u_t} := \sin u_t$.

In order to use Eqs. (25) and (9) for the contact force determination, we need at least two joint torque sensors, one at the inner link and the other at the outer link. However, each finger of the BHand is equipped with a strain gauge to measure the joint torque only at the outer link, i.e. $\Upsilon_{1,2}$ (see Fig. 7b). To get around this issue, we conducted the experiment by applying the contact force in an impactive way rather than by a continuous rolling contact. If the force is applied as an impact and if the object motion is mostly translational, then the force direction can be approximated by the relative velocity between the object centre and the finger contact point (due to the principle of impulse and momentum [25]). In fact, the contact force is likely to be an impact in our experimental setup since the object block can move freely without friction on the horizontal plane and the finger is controlled with a position mode PID servo with high gain for fast response. Following our notational convention, let us denote $v^{t_1}_{t_1}$ the velocity of the (imminent) contact point p_{c_1} at the finger 1 surface right before the impact, written with respect to the contact frame T_1. Likewise, denote v^{t_1} as the velocity of the mass centre of the object written with respect to T_1 frame. Also denote $\Delta v^{t_1}_1 = v^{t_1}_{t_1} - v^{t_1}$. Then the contact force for finger 1 can be approximated as $f^{t_1}_1 \approx \rho\Delta v^{t_1}_1 = \rho\left(v^{t_1}_{t_1} - v^{t_1}\right) \in \mathbb{R}^2$ with some unknown scale factor $\rho \in \mathbb{R}$. From Eq. (9), we have

$$\begin{bmatrix} \Upsilon_{1,1} \\ \Upsilon_{1,2} \end{bmatrix} = \rho J_{h_1}\Delta v^{t_1}_1 \Longrightarrow \rho = \frac{\Upsilon_{1,2}}{e_2^T J_{h_1}\Delta v^{t_1}_1} \tag{26}$$

where $e_2 = [0\ \ 1]^T$. Consequently, we can obtain $f^{t_1}_1$ as

$$f^{t_1}_1 = \frac{\Upsilon_{1,2}}{e_2^T J_{h_1}\Delta v^{t_1}_1} \cdot \Delta v^{t_1}_1. \tag{27}$$

Experimental results and evaluation

The data acquisition and the real-time control have been implemented using a real-time desktop PC running under a real-time operating system provided by *National Instruments Inc.* The real-time loop consists of two dual-rate parallel processes; one for the fast loop (T_s) and the other for the vision loop (T_v). The sensor signals from the BHand can only be communicated through high speed CAN protocol, which made us to choose $T_s = 2.5$ ms as the fastest possible sample rate for acquiring the joint encoder position values and the strain gauge values. The reference command for the joint PID controller is also updated by the T_s. Through the timing test of the vision camera, it has been identified that the total latency of the vision data is around 25 ms. To evaluate the performance of the estimated state variables, we need a monitoring signal (i.e. the true position and orientation data for the target object) for comparison. For this, the original vision data has been acquired every 25 ms and processed offline to remove the latency and to be interpolated into 2.5 ms data which is then stored as the monitoring signal. The original vision data sampled at 25 ms is further down-sampled to create $T_v = 50$ ms of the observation data with latency (i.e. $y_t(k)$ and $y_r(k)$ in Eq. (13)). In the experiments, we used the original vision data (i.e. delayed and down-sampled position measurement of the object) to identify the contact point, which worked reasonably well despite some errors due to the visual latency. It may be attributed to the fact that position variable does not change abruptly in contrast to the force which can be applied in a discontinuous (or an impactive) way.

Table 2 lists the parameters used for the experimental test. The first experimental results, i.e. the 2D motion free from gravity, are presented with time plots in Figs. 8, 9 and 10. During the time period shown in these plots, the object is first struck by the Finger 1 and then subsequently by the Finger 3. Figure 8 shows the joint torques measured by the strain gauges at the outer link joints during this period. Figure 8 reveals that the time duration of force application is less than 100 ms for both fingers, which conforms to our earlier discussion on the use of impactive forces for experimental tests. During the time period shown in Fig. 8, the fingers are held stationary, i.e. $v_{t_1}^{t_1} = v_{t_3}^{t_3} = 0$ and the object was simply bouncing between two fingers. In this case, $\Delta v_1^{t_1} = -v^{t_1}$ and $\Delta v_3^{t_1} = -v^{t_3}$. The resulting directions of $\Delta v_1^{t_1}$ and $\Delta v_3^{t_1}$ (right before the impact) are drawn in Fig. 8 with corresponding directional angles. The contact force for each finger is then determined using Eq. (27).

Figure 9 shows the experimental results for the position variables; the position of the mass center $p^s = [x\ y]^T$ and the orientation angle ϕ. The thick solid line is the

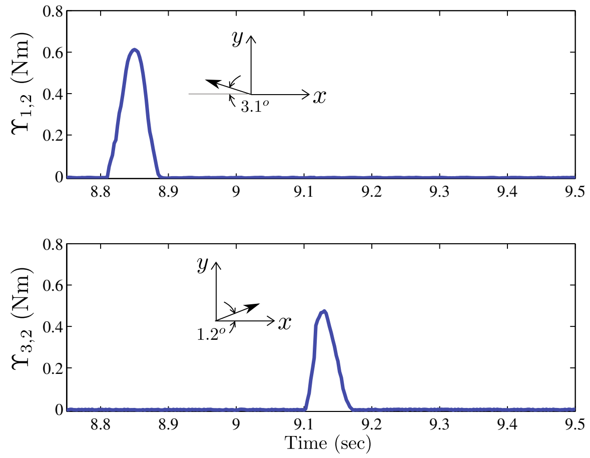


Fig. 8 Outer link joint torques measured by the strain gauges for the case with the object moving on the horizontal plane free from gravity

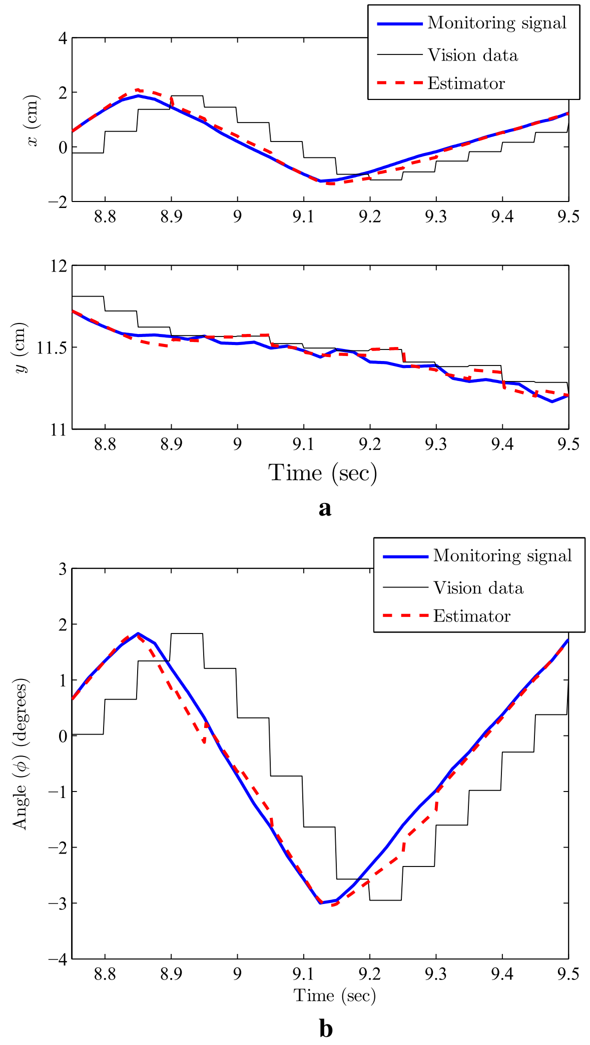


Fig. 9 Comparison of rigid body motion data—position. **a** Translation, **b** orientation

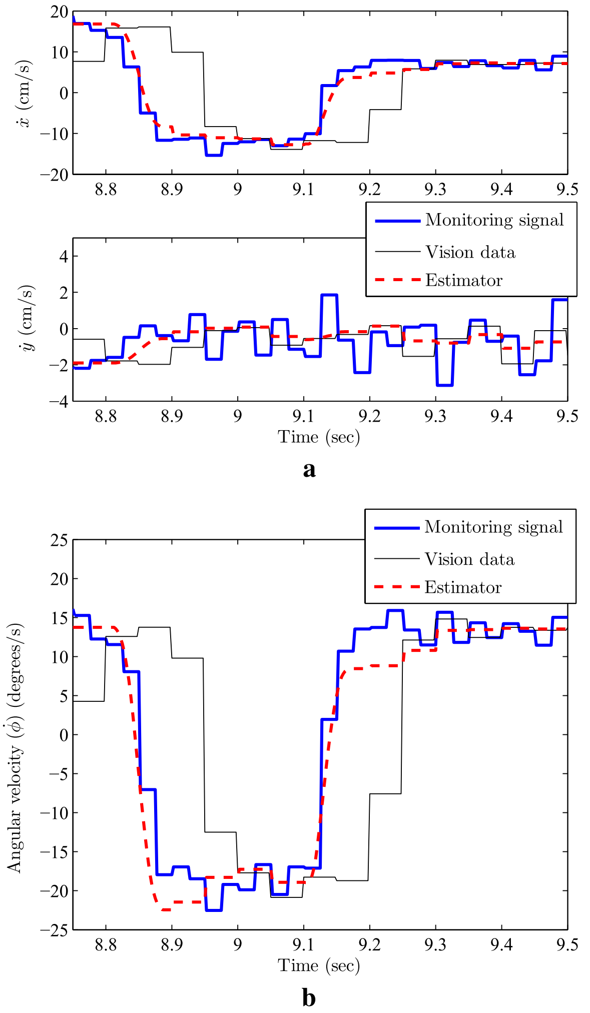
Table 2 Parameter values of the experiment

Parameter	Value	Unit
Object size	10 × 10 cm	
Object mass (m)	3.165	kg
Object inertia ($\mathcal{I}_0$)	$9.72 \cdot 10^{-3}$	kg m^2
Base sample rate (T_s)	2.5	ms
Vision sample rate (T_v)	50	ms
Force noise covariance (W)	$0.04I_2$	N^2
Vision noise covariance (V)	0.01	mm^2

monitoring signal obtained by offline processing of the original vision data sampled at 25 ms rate and is regarded as the true position of the object. The thin solid line is the real-time vision data with $T_v = 50$ ms latency, which we used as the real-time observation signal for the estimator. The thin dotted line is the motion data obtained from the real-time rigid body estimator. It shows that the latency and inter-sample values are compensated so that the estimated motion variables are close to the monitoring signal.

Figure 10 compares the velocities from the rigid body estimator with those of other signals. Again, the thick solid line is the reference (or the true) velocity which is obtained by numerical differential of the monitoring signal. The thin solid line is the velocity computed using the delayed vision data. Remember that we need two consecutive sample points to compute the velocity by numerical differentiation. Therefore, the velocity obtained from the observed vision data (i.e. the one with $T_v = 50$ ms sampling) has additional 50 ms of latency resulting in 100 ms of total latency as shown in Fig. 10. In contrast, as we can see in the thick dotted line, the rigid body estimator can effectively compensate for such latency providing more accurate real-time velocity signal.

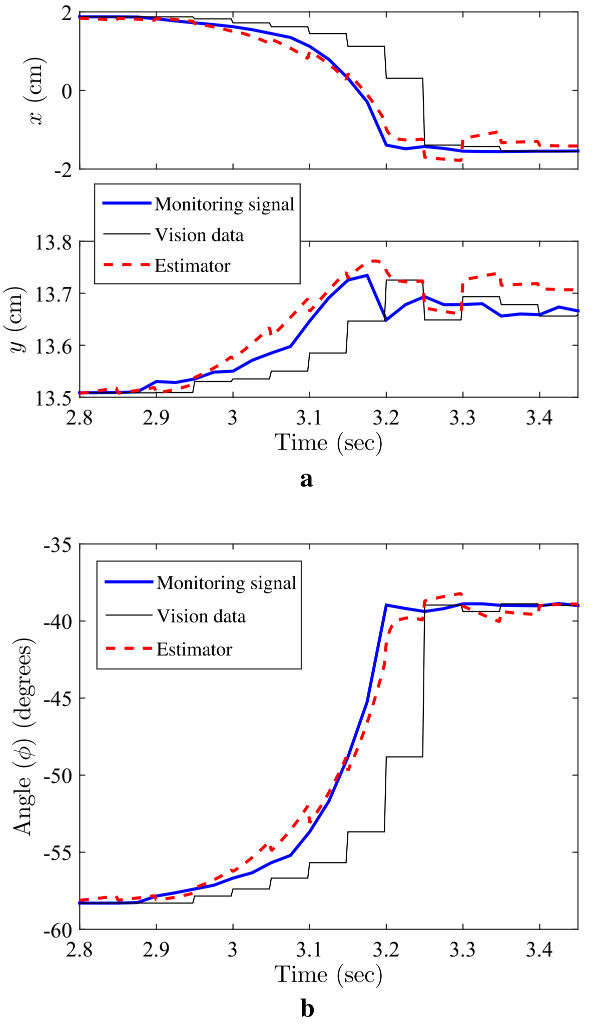
The experimental results for the second experiment, i.e. the case with the object moving under the effect of gravity are shown in Figs. 11 and 12 which are presented in the same format as those for the planar motion are presented in Figs. 9 and 10. The catching action by F1 and F2 occurred a little before 3.2 s of time and generated an impulse-like force to the object, which can be seen from the sharp change in the velocity data of Fig. 10. We can clearly see from Fig. 9 that the estimated motion variables



a

b

Fig. 10 Comparison of rigid body motion data—velocity. **a** Translation, **b** orientation



a

b

Fig. 11 Comparison of rigid body motion data when the object is under gravitational force—position. **a** Translation, **b** orientation

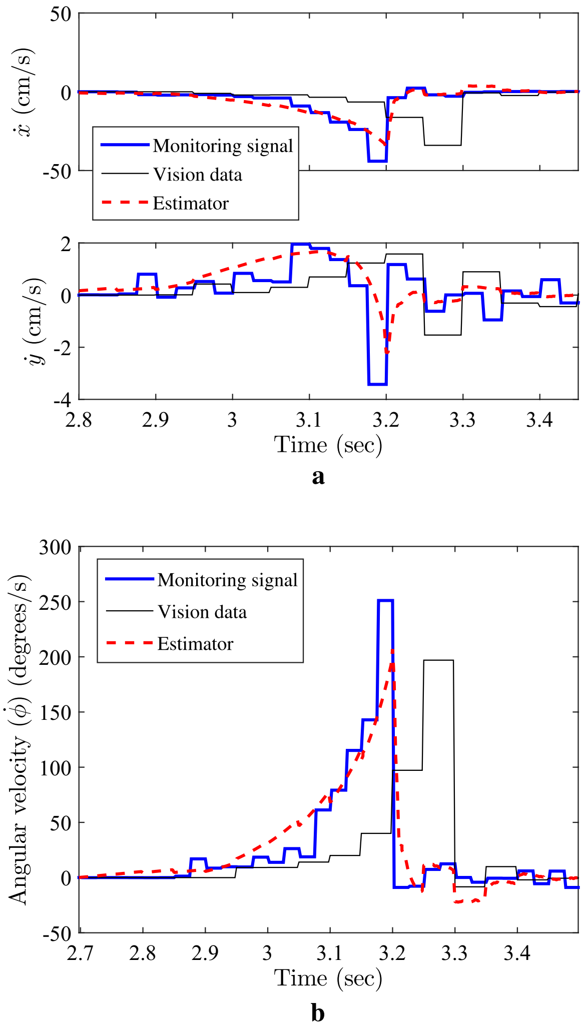


Fig. 12 Comparison of rigid body motion data when the object is under gravitational force—velocity. **a** Translation, **b** orientation

have compensated very well for the latency and missing intersample values of the vision, except for small errors during the catching period. The velocity profiles shown in Fig. 12 also confirm that the proposed rigid body estimator effectively compensates for the vision latency and can be used to provide more accurate real-time velocity data.

Conclusions and discussions

Motivated by the dynamic dexterous manipulation tasks where the fast and accurate estimation of the target object motion is often required, this paper presented a new rigid body estimation scheme combining the visual data with other sensory inputs for contact force measurement. The main point is the complementary use of the fast-sampled force measurement in compensating for the slow-sampled visual data with latency. Mathematically, the rigid body estimator takes the multi-rate extended Kalman filter form with retrospective correction step for the delay compensation.

As demonstrated by the experimental results, the estimator is expected to perform very well if the contact force can be correctly determined. This will be especially so for large values of vision latency since the estimator relies on the numerical integration (or the prediction) between two consecutive vision samples. Use of high fidelity tactile and joint torque sensors is expected to enhance the performance as well as the robustness of the proposed rigid body estimator.

In this paper, we assumed that we have a prior knowledge on the inertial properties of the object such as the mass or the inertia matrix. Such inertial properties may also be identified through preliminary manipulation trials and the application of parameter identification techniques. In fact, when we (humans) manipulate an object with our hands, in the beginning, we have only a limited sense of inertial properties but, through a variety of manipulation tasks, we gradually get to learn inertial properties more accurately and eventually become more skillful. Also, in order for the proposed estimator to be useful in a wide range of applications, it should be tested with more complex dynamic manipulation tasks such as throwing and catching an object in 3D space. We plan to pursue these research ideas as our future works.

Authors' contributions

HB worked on the formulation of the problem under the supervision of the second and third authors. He also designed the experimental setup and carried out the experimental tests. SJ initiated the research idea, supervised the first author for analysis and experiment, and helped him with drafting the initial version of the manuscript. JPH contributed to research ideas and helped drafting the paper by providing feedback. All authors read and approved the final manuscript.

Acknowledgements

Authors acknowledge the financial support from NSERC (Natural Sciences and Engineering Research Council of Cananda) under the Discovery Grant Program (RGPIN-2015-05273).

Competing interests

The authors declare that they have no competing interests.

Received: 22 January 2016 Accepted: 1 July 2016
Published online: 16 July 2016

References

1. Robotics-VO (2013) A roadmap for U.S. robotics: from internet to robotics—2013 edition. http://www.robotics-vo.us/sites/default/files/2013%20Robotics%20Roadmap-rs.pdf. Accessed 10 July 2016
2. Falco J, Marvel J, Elena M (2013) Dexterous manipulation for manufacturing applications workshop. http://www.nist.gov/el/isd/upload/NIST-IR-7940.pdf. Accessed 10 July 2016
3. Okamura AM, Smaby N, Cutkosky MR (2000) An overview of dexterous manipulation. IEEE Int Conf Robot Autom 1:255–262
4. Grebenstein M (2014) Approaching human performance. Springer, Cham

5. Martin J, Grossard M (2014) Design of a fully modular and backdrivable dexterous hand. Int J Robot Res 33(5):783–798
6. Quigley M, Salisbury C, Ng AY, Salisbury JK (2014) Mechatronic design of an integrated robotic hand. Int J Robot Res 33(5):706–720
7. Deshpande AD, Xu Z, Weghe MV, Brown BH, Ko J, Chang LY, Wilkinson DD, Bidic SM, Matsuoka Y (2013) Mechanisms of the anatomically correct testbed hand. IEEE/ASME Trans Mechatron 18(1):238–250
8. Kandel ER, Schwartz JH, Jessell TM et al (2000) Principles of neural science, vol 4. McGraw-Hill, New York
9. Franklin DW, Wolpert DM (2011) Computational mechanisms of sensorimotor control. Neuron 72(3):425–442
10. LaValle SM (2006) Planning algorithms. Cambridge University Press, New York
11. Murray RM, Sastry SS (1994) A mathematical introduction to robotic manipulation. CRC Press, Boca Raton
12. Rizzi AA, Koditschek DE (1996) An active visual estimator for dexterous manipulation. IEEE Trans Robot Autom 12(5):697–713
13. Wang C, Lin CY, Tomizuka M (2014) Statistical learning algorithms to compensate slow visual feedback for industrial robots. ASME J Dyn Syst Meas Control 137(3):031011
14. Jeon S, Tomizuka M, Katou T (2009) Kinematic kalman filter (kkf) for robot end-effector sensing. ASME J Dyn Syst Meas Control 131(2):021010
15. Hebert P, Hudson N, Ma J, Burdick J (2011) Fusion of stereo vision, force-torque, and joint sensors for estimation of in-hand object location. IEEE Int Conf Robot Autom 1:5935–5941
16. Allen PK, Miller AT, Oh PY, Leibowitz BS (1997) Using tactile and visual sensing with a robotic hand. IEEE Int Conf Robot Autom 1:676–681
17. Kamiyama K, Kajimoto H, Kawakami N, Tachi S (2004) Evaluation of a vision-based tactile sensor. IEEE Int Conf Robot Autom 2:1542–1547
18. Prats M, Sanz PJ, Del Pobil AP (2009) Vision-tactile-force integration and robot physical interaction. IEEE Int Conf Robot Autom 1:3975–3980
19. Kragic D, Crinier S, Brunn D, Christensen HI (2003) Vision and tactile sensing for real world tasks. IEEE Int Conf Robot Autom 1:1545–1550
20. Namiki A, Komuro T, Ishikawa M (2002) High-speed sensory-motor fusion based on dynamics matching. Proc IEEE 90(7):1545–1550
21. Mason M (2001) Mechanics of robotic manipulation. MIT Press, Cambridge
22. Titterton D, Weston J (2004) Strapdown inertial navigation technology, vol 17, 2nd edn. Radar, Sonar, Navigations and Avionics-IET, Stevenage
23. Bar-Itzhack IY, Harman RR, Choukroun D (2002) State-dependent pseudo-linear filters for spacecraft attitude and rate estimation. In: AIAA Guidance, Navigation, and Control Conference and Exhibit
24. Leung W, Damaren C (2004) A comparison of the pseudo-linear and extended kalman filters for spacecraft attitude estimation. In: AIAA Guidance, Navigation, and Control Conference and Exhibit
25. Brogliato B (1999) Nonsmooth mechanics: models. Dynamics and control. Springer, London