A Method for Sustaining Consistent Sensory-Motor Coordination Under Body Property Changes Including Tool Grasp/Release

Cota Nabeshima[†‡], Yasuo Kuniyoshi[‡]

† *CYBERDYNE Inc., Gakuen Minami D25-1, Tsukuba-shi, Ibaraki, Japan.*

‡ *The University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo, Japan.*

{*nabesima, kuniyosh*}*@isi.imi.i.u-tokyo.ac.jp*

Abstract

Body physical and spatial properties change because of unexpected accidents or objects it wears (e.g. tool). To keep stable and consisntent behaviors in acting time, it is necessary to adapt body representation to the unexpected changes. This adaptation is achieved when a robot sustains its coordination between sensation and body motion; i.e. observation and identification of the discrepancy between its knowledge and actually obtained sensory feedback. For the sustainability, the robot should autonomously judge the reliability and the tolerance of the identification. Standing on basic mathematics, we propose a method with an indirect but clear criterion of the reliability for the sustainable coordinations in this paper. We also implement the method on visual-motor coordination and on kinesthetic-motor coordination. Remarkably, our method achieves marker-free, easily convergent and enough accurate (i.e. easily applicable) hand-eye calibration method with irrelevant objects in view. The evaluation of the method is provided with experiments in a real robot. Our experimental results show the novelty of the concept of sustainable coordination and the availableness of our method for the concept. We hope this paper be a powerful approach for building autonomous robots.

Keywords: Adaptive body schema, Tool-body assimilation, Marker-free hand/head-eye calibration, Inertia identification, Convergence criterion

1 Introduction

Our daily behaviors deeply depend on coordinations among informations which have different modalities: motion, vision, kinesthesia, or so. Without those coordinations, we could neither reach a visible cup (visual-motor coordination), nor touch and grope an invisible place (kinesthetic-motor coordination). These abilities are obviously primary for us; i.e. they are also essential for the robots who autonomously act in the real world.

As for almost all past robots, their designers built *a-priori* in those sensory-motor coordinations. Although, of course, this design strategy works quite fine in designed and regulated situation, the de-

signers should preliminarily know "everything" about the situation. However, there are situations where preliminary built-in coordination is inappropriate to use; i.e. a robot has to obtain the coordinations by itself. For instance, we can assume situations where physical reaction unexpectedly causes the spatial alignment of robot's equipments to change, or it purposely picks up and holds unexpected objects (e.g. tools). That is, robots should sustainably observe and identify the necessary coordination in their working period (**sustainable coordination**); otherwise, they can not keep stability and consistency of their behaviors.

In our past research [1, 2], we clarified adaptation of kinematic hand-eye coodination and kinesthetic inertia identificaiton are essential for robotic tool-use. Pursuing our past research, we mainly focus on and discuss those coordinations in this paper: hand-eye (visual-motor) coordination and inertia (kinesthetic-motor) identification. We know those coordinations are primary and inadvertently performed by animals. Unfortunately we also know robotic techniques for the sustainable coordination are yet below the level of animate life. In Sec. 2, we review the existing techniques and their deficiency. To find a clue to complement the techniques, we also introduce and discuss physiological, biological, and brain-scientific literatures.

Our discussion points out that the robotic system should have two autonomous abilities for the sustainability: detection of self-derived information, and estimation of allowable error in identification. The former is necessary to filter irrelevant information and accelerate the identification; and the latter is to determine the beginning/end of the identification. Corresponding the two abilities, we introduce two methods in Sec. 3. One method is to detect self-body in camera view, which works as a filter of unnecessary visual information. This detection method is marker-free, convergence-free, and enough accurate because of a "motion coincidence" criterion. The other method is to estimate the allowable error, which can be a threshold to switch between observation and identification. Because this estimation method uses singular values, it can share the calculation with the identification process; i.e. it works simultaneously with identification. We combine the two methods into a sustainable coordination method. For the applications of our method, we also show its implementations for the hand-eye coordination and the inertia identification in Sec. 3.

In Sec. 4, we experiment our method in a real robot equipped with an arm, a stereo camera, force-torque sensor and computers. The robot switches observation and identification in response to the disturbance provided by the experimenter. Our results indicate its sustainability for the coordination. Based on the experiments, we discuss the future of our method in Sec. 5, and conclude this paper in Sec. 6.

2 Sensory-motor coordination in robots and animals

2.1 Hand-eye coordination

If the kinematics of the held object (tool) is given, a robot can use it and learn more higher-level tasks such as the tennis-playing robot of Miyamoto and Kawato [3]. However, the kinematics is easily changeable and unknown even when the robot purposely picks up a novel object. Body, as well as a held object, is also easily changeable and unknown in essentials. Unless spatial information of its body is *a-priori* given, a robot cannot know even its body area in the camera view. Yoshikawa [4] proposed a method to extract the body area exploiting the fact that the body is stationary to the camera during the body waggling. However, it does not include the binding with the joint motion which causes the visible motion; i.e. it cannot identify the body kinematics. Without the body kinematics, the robot cannot use the extracted body area to generate motions.

To address the binding problem, researchers have proposed the methods using "motion contingency" cues; "motion contingency" means nature where incoming information changes in response of body motion. A robot of Michel et al. [5] estimated time lag between camera and joint angle sensors using the timing of their motions co-change; clustered the areas which have the similar time lag; and define the clustered areas as the body area. Unfortunately because it used a difference images to estimate visual motion, the body area obtained by their method included non-body background, and was hard to exclude moving objects but body parts.

There are methods using optical flow to obtain visual motion. A method of Kemp and Edsinger [6] divides camera image to subimages, calculates the mutual information between the body motion and the optical flow within each subimage, and determines the subimage which is most effected by the body motion. To calculate the mutual information, it first clusters largely changing subimages and joint angles. One variance (vision) is spatially calculated within each subimage cluster, and the other variance (body motion) is temporally done within each joint angle cluster. By obtaining the mutual information from these variances, the body area is estimated as the area of the highest mutual information at the joint angle cluster including the angle. Their method enables a robot to learn the body area in case of other irrelevant moving objects in camera view due to the law of large numbers. However, it requires a large amount of well-balanced data samples to obtain the mutual information on all subimages. This fact inevitably makes their method offline.

A online method was proposed by Fitzpatrick and Metta [7] divides camera image to subimages, calculates cross-correlation between each joint angle and the optical flows within each subimage, and determines the body area. Because of linearity of cross-correlation, their method can detect the plausible body area at a short time in case of other irrelevant objects in view. However, the linearity of cross-correlation is weak to non-linear transformations which are innate in optical system: refraction, reflection or perspective. Furthermore, it is not clear how to deal with fake or negative correlation. To manage the non-linearity in optical system, Natale [8, 9] proposed to use periodic body motion to extract the

visible body area which moves at the same period as the body motion. His method restricts robot's motion to periodic one; i.e. the robot is hard to determine the body area at shorter time than the time which is taken to estimate the period.

The methods reviewed above aspire for the learning of kinematics with the joint angles as input. However, a lot of data samples are necessary to learn the kinematics, which is generally a non-linear function. On the other hand, the kinematics of an object (tool) held by the robot hand is obtained as a linearly transformed kinematics of the hand [1, 2, 10].

Kemp and Edsinger assumed only the tips of daily tools are usable, an they proposed a stable method to estimate the 3D coordinate of the tip of the held object with a joint angle input [11]. Their significant assumption is that the tip of a tool always has the largest optical flow in the camera view. 3D coordinates of the tip are measured by a stereo camera for the learning samples. The robot offline learns probability relation between the obtained 3D coordinate and 2D image coordinate, and estimates the 3D coordinate which has the maximum likelihood with both the joint angle and 2D image coordinate as input Because of their assumption, the conditional probabilities which are calculated in their method ends up inaccurate when other moving objects are in the camera view, and the calculation requires much more well-balanced data samples. This requirement is ill-fitted to the case where the held object (tool) is frequently changed. Moreover, the velocity of the tip of a serial-link manipulator is not always maximal in view. If we actually adopt their method, it is necessary to control the velocity of the tip to be maximal in the camera view, or to keep other body parts out of the camera view.

In our previous work, we focused on the linear time-invariant relationship between the coordinates of a body part and a object fixed to the body part, and we proposed a method using the canonical correlation to identify a held object or body parts [2]. Our robot instantly and stably obtained kinematics of a new held object, exploiting known kinematics of the body. Of course, our method have a similar problem to the method of Fitzpatrick and Metta [7] because of linearity of the canonical correlation.

In Sec. 3, we propose a novel method to rapidly identify body parts in camera view in order to compensate the conventional works. Our novel method uses optical flow and a simpler but robust criterion than motion period; we no longer need motion restriction such as periodical one.

2.2 Inertia identification

A robot has to know the inertia parameters of its held or adhered object to detect a touch on it [2]. However, the inertia parameters are liable to change due to situation. The robot should extemporaneously identify them, and autonomously determine whether the identification is enough or not.

Inertia parameter identification is a field of robotics [12]. Many methods in the early period identified and provided mass, length or COG of the links as inertia parameters, which exploit different cues: e.g. static torque by gravity [13], frequency response to input [14], or joint angle information to estimate [15].

For identification of inertia parameters including link length and COG, we know methods using pseudo-inverse matrix [16, 15] or Kalman filter [17]. The method of Sujan and Dubowsky [18] uses mutual

information, and deduce body motion to accelerate the convergence of the Kalman filter. However, those methods do not provide us a relevance criterion of the identification nor an autonomous way to determine the beginning/end of the identification.

We point out their problem for the sustainable identification is the usage of the least-square method, which is directly adopted to the identification equation. In Sec. 3, we propose another simple way which adopts the least-square method to calculate the dimension of the null-space of the identification equation. It takes small computation cost and provides easy detection of convergence, which is better suited for a sustainable online method.

2.3 Sustainable coordination in animals

2.3.1 Brain's adaptation to foreign objects

Many animal species hold a object on the body using their hands or mouthes [19]. This holding leads to the change of sensory-motor coordination. The adaptation to the change is often seen (i.e. sustainable) in various animal species. Berlucchi and Aglioti [20] said a human adapts its multi-sensory body representation in the brain to a foreign object. Here, focusing on "touch," we review what behaviors are allowed by the sustainable coordination, and how the multi-sensory coordination is represented in the brain.

When we ape touch an object by the held tool, we feel and deal with the held one as if it is our own hand [21]. We also imagine the invisible edge of a worn hat and go under a gate without touch [22]. We human [23] and horses [24] adapt to prosthetic limbs alternative to amputated limbs. We can say many animals (at least, ape) are able to manipulate and feel a foreign object as their own body after adaptation. By this ability, for instance, we can grope in invisible space using an extemporaneously obtained object [2]; or wild gorillas observed by Breuer et al. [25] could stab a picked-up branch into water hole and measure its depth.

In macaque's intraparietal cortex, there are bimodal neurons responding to visual and tactile stimuli on its hand. Iriki et al. [26] showed the visual receptive field of the bimodal neurons extend to its held tool [26]. This phenomenon occurs under conditions where a tool is hidden behind a blinder [27], or even whre a tool is reflected on a TV monitor [28].

Yamamoto et al. reported the phenomenon where a reversal of temporal judgement of subjects when the experimenter gives tactile stimuli with several hundred millisecond interval on the subject's crossed hands respectively [29]. They also showed a occurrence of the phenomenon when subjects hold sticks by each hand, and cross only the sticks or only the hands [30].

The knowledge provided by Iriki et al. and Yamamoto et al. seems a proof that the brain deals with a foreign object (tool) as a part of the body. That is, it suggests the brain internally as well as behaviorally adapts its body representation to extemporaneously attached objects on the body, even when it is obtained temporarily.

2.3.2 Kinesthetic adaptation: cognitive scientific view

To perceive a contact location on the surface of a held object without vision [25, 30] is a estimation problem of the contact from kinesthetic response coming through the object. To physically address this problem, inertia parameters of the object are necessary to know explicitly or implicitly [2]. Only if those parameters are known, an agent (robot/human) can estimate the force and torque generated by a contact; it derives a possible line of the contact location in 3D space from the estimated force and torque; it estimates the contact location as the intersection between the line and the surface of the object. Here, the surface other than the contact point never effect the estimation in principle. This notion is experimentally supported; the reversal of temporal judgement of subjects [30] occurs even with various shaped tool [31].

Do animals "know" inertia parameters of a foreign object to perceive a contact on it? Turvey had his subjects swing an object behind a blinder and sensuously represent its length [32]; accordingly, they guessed the right length (dynamic touch). By his experiments where inertia parameters of the held object were changed as its tactile texture remains, it was revealed that our perception of length depends only on the inertia parameters. Turvey's experiments imply us that other animals can identify and know inertia parameters of attached object.

2.3.3 Kinematic adaptation: brain scientific view

We can imagine and control the invisible edge of a worn hat [22]. Kinematic body representation in brain is often called "body schema" or "body image" [33, 34]. They represent position and attitude, or occupied area of each body part, which are updated according to body motion [35]. These representation adapt to body changes caused by growth or injury.

It is known the kinematic body representation quickly adapts to the body change. Amputees sometimes feel the existence of their amputated limbs by itch or ache. This phenomenon called "phantom limbs" seems because the body representation corresponding with amputated part still remains after the body changes [36, 37]. This phenomenon is not only reported by patients but suggested to be a problem in brain; Ramachandran and Rogers-Ramachandran [38] treated phantom limbs through several minutes trial using a mirror box; and they also showed that the treatment changed the body representation in brain. We note the body representation can adapt quickly because of sensory feedback.

2.3.4 Kinematic adaptation: cognitive scientific view

How do animals quickly adapt their body representation? If an agent knows and uses spatial information in the environment as reference information, the agent can obtain spatial information of its body—sensory-motor coordination. Conversely, if an agent knows and uses spatial information of its body, as reference information, the agent can obtain spatial information in the environment. That is, spatial cognition of body and that of environment are mutually dependent and cooperated.

The body schema is a representation which binds visual, tactile, and proprioceptive senses [39].

Because this binding is based on somatic (tactile and proprioceptive) sense, Maravita et al. [40, 41] said it is organized with in "peripersonal space" [42]. Grounding on experiments of adaptation of spatial recognition as we wear a prism or glasses on our eyes, Sekiyama et al. [43] said that an agent can stably recognize space around its body owing to the body schema which binds spatial sensory informations.

Now, we can say calculation during the kinematic adaptation of the body representation is a problem associated with identification of spatial information from time-series multi-sensory information. For the adaptation, reference information is necessary, whose identity ensured among multi-sensory information. Here, we assume an agent actively invokes an event; if the agent observes the event in each sensory input, it can ensure the identity of the input information, and use it as the reference. We believe that animals also use self-derived information as the reference, and they adapt their sensory-motor coordination.

Mirror-image recognition is an example where an agent extracts and exploits self-derived information in vision. Although chimpanzees initially get frightened at mirrors, they gradually understand the functionality of mirror [44]. This mirror-image recognition occurs not only in apes [45] but also dolphins [46] or elephants [47]. We can say that the treatment of phantom limbs [38] exploited the mirror-image recognition, too.

How do animals extract self-derived information? Rochat's experiments in human infants [48] are suggestive for the solution. In his experiments, subject's body was captured by video camera, and the captured moving images were projected onto a screen in front of the subject. As the result, his subjects predominately gazed to the projected images; moreover, they did to the tesseral images which shuffled the original images. Rochat also showed this gaze did not occur when he operated temporal delay on the moving images. Rochat's results indicate infants extract self-derived information using temporal information as a clue even from distorted visual images which have no physical connection with their body. Recently, there are researches tackling how temporal delay on sensory information effects our brain (e.g. [49]).

It is natural to use a temporal clue under uncertainty of spatial information. We often call a property "motion contingency", where information changes time-adjacently to active motion/behavior. We suppose that animals exploit the motion contingency, detect a reference information from multi-sensory input, and adapt a sensory-motor coordination—kinematic body representation. To deliver the sustainable coordination method, it seems better way to use contingency and detect a reference information; it can work as a filter which weeds out irrelevant information for sensory-motor coordination.

3 Proposed method for sustainable coordination

If sensory-motor coordination is stationary for a long time, an agent should identify it only once. Conventional robotic system has been assumed not to change in its working time, and have used pseudo-inverse matrix, robust estimation, or Kalman filters to identify and initialize parameters which limit and control the coordination. However, these are basically unsustainable methods if adopting them in direct way.

Methods based on the pseudo-inverse matrix and the robust estimation take more memories and calculation time corresponding to the number of input data because they require appropriate sampling of the data; i.e. they are not real-time. Methods based on Kalman filters have difficulty to design appropriate covariance matrix of noise, and tend to make its convergence worse when many parameters should be identified. We can say these methods are not sustainable because they can not automatically calculate reliability of the identification nor determine allowable limit of the error of identified coordination; reluctantly, the designers choose to decide the end of identification by themselves.

If changes of body propoerty often occur because of tool holding, object attachment, or injury (failure), an agent should sustainably adapt sensory-motor coordination in its working time. A method we show in Sec. 3.1 identifies parameters from steadily input data, calculates reliability of the identification in an indirect way, and continues and finishes data sampling for identification according to the reliability. Its computation is easy and optimal as far as the least square error, which takes constant space and time complexities and derives allowable error simultaneously with the identification of parameters.

3.1 Identification of sensory-motor coordination and its allowable error

Here we assume spatial sensory-motor coordination can be linearized using parameters. Using data input at a given time $\mathbf{A}_t$ ($\mathrm{M}\times\mathrm{N}$ matrix) and the parameters $\boldsymbol{x}$ (N-dimensional vector), we define the equation of a sensory-motor coordination as,

$$\mathbf{A}_t \boldsymbol{x} = 0. \tag{1}$$

If $\boldsymbol{x}$ is derived by identification, its squared error s_t is,

$$s_t = || \mathbf{A}_t \boldsymbol{x} ||^2. \tag{2}$$

An agent can decide to renew $\boldsymbol{x}$ by setting a threshold for s_t as allowable error; i.e. with the threshold σ, $\boldsymbol{x}$ is to be renewed if $s_t > \sigma$. In this subsection, we show how to automatically identify $\boldsymbol{x}$ and σ. Our method is lightweight and takes constant complexities; therefore, it also suitable for observation of the error, which should be done at all times.

Mean of square error S_t among n samples of data which is already input at a given time can be represented as,

$$\begin{aligned} S_t &= \frac{1}{n}\sum_t s_t \\ &= \frac{1}{n}\sum_t \boldsymbol{x}^{\mathrm{T}}\mathbf{A}_t^{\mathrm{T}}\mathbf{A}_t\boldsymbol{x} \\ &= \boldsymbol{x}^{\mathrm{T}}\left(\frac{1}{n}\sum_t \mathbf{A}_t^{\mathrm{T}}\mathbf{A}_t\right)\boldsymbol{x} \\ &= \boldsymbol{x}^{\mathrm{T}}\,\mathbf{C}_t\,\boldsymbol{x}. \end{aligned} \tag{3}$$

Note that $\mathbf{C}_t$ is the variance-covariance matrix for the input data with the mean 0. We formulate

parameter identification as a derivation problem of $\boldsymbol{x}$ which minimize S_t, where $\boldsymbol{x}$ is defined as,

$$\boldsymbol{x} \quad = \quad \mathrm{argmin}(\, S_t \,). \tag{4}$$

If we see S_t is a function of $\boldsymbol{x}$, S_t is extremal with $\boldsymbol{x}$ which minimizes mean square error S_t. That is, partial differential of S_t by $\boldsymbol{x}$,

$$\frac{\partial S_t}{\partial \boldsymbol{x}} \quad = \quad 2\,\mathbf{C}_t\,\boldsymbol{x} \tag{5}$$

should be 0. It means equation,

$$\mathbf{C}_t\,\boldsymbol{x} = 0 \tag{6}$$

should be solved for identification.

Because of Eq. (6), $\boldsymbol{x}$ is a base vector of null-space of $\mathbf{C}_t$. Now we define $n(\,\mathbf{C}_t\,)$ as the dimension of nullspace of $\mathbf{C}_t$. If and only if $n(\,\mathbf{C}_t\,) = 1$, identified $\boldsymbol{x}$ is the stable solution. Note that eventual $\boldsymbol{x}$ is derived by normalization with other constraint condition because $\boldsymbol{x}$ has arbitrariness of multiplication by scalars.

However, it is a hard problem to determine whether $n(\,\mathbf{C}_t\,) = 1$ from real data input. We show how to obtain a unit vector minimizing S_t. Here we assume the singular value decomposition of $\mathbf{C}_t$ is given as $\mathbf{C}_t = \mathbf{U}_t\,\mathbf{S}_t\mathbf{V}_t^{\mathrm{T}}$. $\mathbf{S}_t$ is a diagonal matrix called singular matrix whose elements are λ_1, λ_2, ..., λ_{N}, ($\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_{\mathrm{N}} \geq 0$). Because $\mathbf{C}_t$ is diagonal, $\mathbf{U}_t$ and $\mathbf{V}_t$ are under relation $\mathbf{U}_t = \mathbf{V}_t$. We define i-th row of $\mathbf{V}_t$ as $\boldsymbol{v}_i$ and substitute them into Eq. (3), then,

$$\begin{aligned} S_t \quad &= \quad \boldsymbol{x}^{\mathrm{T}}\,\mathbf{V}_t\,\mathbf{S}_t\,\mathbf{V}_t^{\mathrm{T}}\,\boldsymbol{x} \\ &= \quad \left(\,\mathbf{V}_t^{\mathrm{T}}\,\boldsymbol{x}\,\right)^{\mathrm{T}}\mathbf{S}_t\,\left(\,\mathbf{V}_t^{\mathrm{T}}\,\boldsymbol{x}\,\right), \end{aligned} \tag{7}$$

where ${\mathbf{V}_t}^{\mathrm{T}}\,\boldsymbol{x}$ means rotation of vector because the right singular matrix $\mathbf{V}_t$ is a orthonormal matrix.

Here we represent a unit vector $\boldsymbol{x}$ to be identified as a linear combination of $\boldsymbol{v}_i$, i.e. with coefficient vector $\boldsymbol{a}$ as,

$$\boldsymbol{x} = \mathbf{V}_t\,\boldsymbol{a}, \tag{8}$$

$$\text{under } \boldsymbol{a}^{\mathrm{T}}\boldsymbol{a} = 1. \tag{9}$$

Substituting this equation to Eq. (7), S_t becomes as,

$$S_t \quad = \quad \boldsymbol{a}^{\mathrm{T}}\,\mathbf{S}_t\,\boldsymbol{a}. \tag{10}$$

With a undetermined multiplier k, Eq. (10) can be represented as,

$$S_t \quad = \quad \boldsymbol{a}^{\mathrm{T}}\,\mathbf{S}_t\,\boldsymbol{a} + k\,\left(\,1 - \boldsymbol{a}^{\mathrm{T}}\boldsymbol{a},\right). \tag{11}$$

The partial differential of this equation by $\boldsymbol{a}$ is as,

$$\frac{\partial S_t}{\partial \boldsymbol{a}} \quad = \quad 2(\,\mathbf{S}_t - k\,\mathbf{E}\,)\,\boldsymbol{a}, \tag{12}$$

where the unit matrix $\mathbf{E} = [\, \boldsymbol{e}_1, \ldots, \boldsymbol{e}_{\mathrm{N}} \,]$. Because $\mathbf{S}_t$ is diagonal, $k = \lambda_i$ and $\boldsymbol{a} = \boldsymbol{e}_i$ are necessary to have Eq. (12) be 0; i.e. $S_t = \lambda_i$ at extremal. Therefore, the minimal S_t is λ_{N} and then $\boldsymbol{x} = \boldsymbol{v}_{\mathrm{N}}$ (unit vector). In other words, the column vector $\boldsymbol{v}_{\mathrm{N}}$ of right singular matrix corresponding to minimal singular value λ_{N} of variance-covariance matrix $\mathbf{C}_t$ is the directional vector which minimize mean square error S_t, and the minimal error equals to λ_{N}.

Obtained $\boldsymbol{v}_{\mathrm{N}}$ is to be scaled by other constraint condition as $\boldsymbol{x} = a\,\boldsymbol{v}_{\mathrm{N}}$ with a coefficient a. Now Eq. (7) is transformed as,

$$\begin{aligned} S_t &= \left(\mathbf{V}^{\mathrm{T}}\, a\, \boldsymbol{v}_{\mathrm{N}} \right)^{\mathrm{T}} \mathbf{S} \left(\mathbf{V}^{\mathrm{T}}\, a\, \boldsymbol{v}_{\mathrm{N}} \right) \\ &= a^2 \lambda_{\mathrm{N}}. \end{aligned} \tag{13}$$

Finally, the mean squared error is estimated as $S_t = a^2\lambda_{\mathrm{N}}$ in the identified coordination.

As discussed above, we can define the reliability of parameter identification by (a) how much λ_{N} is close to 0, (b) how smaller λ_{N} is than $\lambda_{\mathrm{N}-1}$, and (c) how much $\boldsymbol{v}_{\mathrm{N}}$ satisfies constraint conditions. Meanwhile, if the coefficient a normalizing obtained $\boldsymbol{v}_{\mathrm{N}}$ is determined, a robot can estimate the mean square error is $a^2\lambda_{\mathrm{N}}$ under obtained $\boldsymbol{x}$, which is a reference to determine the threshold σ.

Our method has features listed below:

- real-time:
  $\mathbf{C}$ is fixed size, and computation time for its singular decomposition is estimated under a constant value;
- numerically stable:
  Because $\mathbf{C}$ is a variance-covariance matrix of input data whose mean is tentatively set 0, it is hard to include numerical error;
- easy to check termination condition:
  reliability of identification is easy validated using singular values as criterion;
- easy to check starting condition:
  The threshold is automatically derived from parameter identification as the mean square error;
- incoherent to other identification precesses:
  Renewing $\mathbf{C}$ at the beginning of each identification makes the calculation incoherent to the previous identification; and,
- optimal in principle:
  Identified parameters are optimal as far as it minimizes the mean square error.

3.2 Constraint condition for identification including rotation matrix

When we linearize a spatial sensory-motor coordination such as Eq. (1), the parameters to be identified often include rotation matrix $\mathbf{R}$, which represents difference of attitude between two coordinate systems.

We define $\mathbf{R}$ as,

$$\mathbf{R} \;=\; \left[\; \boldsymbol{\gamma}_1 \quad \boldsymbol{\gamma}_2 \quad \boldsymbol{\gamma}_3 \;\right]. \tag{14}$$

Here, we assume parameters to be identified as,

$$\boldsymbol{x} \;=\; \left[\; {\boldsymbol{\gamma}_1}^{\mathrm{T}} \quad {\boldsymbol{\gamma}_2}^{\mathrm{T}} \quad {\boldsymbol{\gamma}_3}^{\mathrm{T}} \;\right]^{\mathrm{T}}, \tag{15}$$

and parameter vector obtained as a result of the method in Sec. 3.1 at each data input as $\hat{\boldsymbol{x}} = \left[\; {\hat{\boldsymbol{x}}_1}^{\mathrm{T}} \quad {\hat{\boldsymbol{x}}_2}^{\mathrm{T}} \quad {\hat{\boldsymbol{x}}_3}^{\mathrm{T}} \;\right]^{\mathrm{T}}$. We show in this subsection a method to determine whether $\hat{\boldsymbol{x}}_1$, $\hat{\boldsymbol{x}}_2$ and $\hat{\boldsymbol{x}}_3$ are elements of rotation matrix.

For the determination, we define a matrix $\hat{\mathbf{X}}$ whose elements are $\hat{\boldsymbol{x}}_1$, $\hat{\boldsymbol{x}}_2$ and $\hat{\boldsymbol{x}}_3$ and calculate the rotation matrix $\hat{\mathbf{R}}$ which are mostly approximate to $\hat{\mathbf{X}}$. Our derivation method applies Challis's method [50]. If $\hat{\boldsymbol{x}}_1$, $\hat{\boldsymbol{x}}_2$ and $\hat{\boldsymbol{x}}_3$ are elements of $\hat{\mathbf{R}}$, with $\boldsymbol{e}_1$, $\boldsymbol{e}_2$ and $\boldsymbol{e}_3$ which are elements of the unit matrix $\mathbf{E}$, we can represent the relationship among those values as,

$$\hat{\mathbf{R}}^{\mathrm{T}} \hat{\boldsymbol{x}}_i \;=\; \boldsymbol{e}_i. \tag{16}$$

Accordingly, square error S_R is as,

$$\begin{aligned}
S_R \;&=\; \sum_{i=1}^{3} || \, \hat{\mathbf{R}}^{\mathrm{T}} \hat{\boldsymbol{x}}_i \,-\, \boldsymbol{e}_i \, ||^2 \\
&=\; \sum_{i=1}^{3} \left(\hat{\boldsymbol{x}}_i^{\mathrm{T}} \hat{\mathbf{R}} \,-\, \boldsymbol{e}_i^{\mathrm{T}} \right) \left(\hat{\mathbf{R}}^{\mathrm{T}} \hat{\boldsymbol{x}}_i \,-\, \boldsymbol{e}_i \right) \\
&=\; \sum_{i=1}^{3} \left(\hat{\boldsymbol{x}}_i^{\mathrm{T}} \hat{\mathbf{R}} \, \hat{\mathbf{R}}^{\mathrm{T}} \hat{\boldsymbol{x}}_i \,-\, \boldsymbol{e}_i^{\mathrm{T}} \hat{\mathbf{R}}^{\mathrm{T}} \hat{\boldsymbol{x}}_i \,-\, \hat{\boldsymbol{x}}_i^{\mathrm{T}} \hat{\mathbf{R}} \, \boldsymbol{e}_i \,+\, \boldsymbol{e}_i^{\mathrm{T}} \boldsymbol{e}_i \right) \\
&=\; \sum_{i=1}^{3} \left(\hat{\boldsymbol{x}}_i^{\mathrm{T}} \hat{\boldsymbol{x}}_i \right) \,-\, 2 \sum_{i=1}^{3} \left(\boldsymbol{e}_i^{\mathrm{T}} \hat{\mathbf{R}}^{\mathrm{T}} \hat{\boldsymbol{x}}_i \right) \,+\, \sum_{i=1}^{3} \left(\boldsymbol{e}_i^{\mathrm{T}} \boldsymbol{e}_i \right).
\end{aligned} \tag{17}$$

In Eq. (17), the second term of right-hand side $2 \sum_{i=1}^{3} \boldsymbol{e}_i^{\mathrm{T}} \hat{\mathbf{R}}^{\mathrm{T}} \hat{\boldsymbol{x}}_i$ should be maximal to minimize S_R because the first and third terms are constant.

We transform the second term of right-hand side of Eq. (17) as,

$$\begin{aligned}
\sum_{i=1}^{3} \boldsymbol{e}_i^{\mathrm{T}} \hat{\mathbf{R}}^{\mathrm{T}} \hat{\boldsymbol{x}}_i \;&=\; \sum_{i=1}^{3} \mathrm{trace} \left(\hat{\mathbf{R}} \, \boldsymbol{e}_i \hat{\boldsymbol{x}}_i^{\mathrm{T}} \right) \\
&=\; \mathrm{trace} \left(\hat{\mathbf{R}} \sum_{i=1}^{3} \boldsymbol{e}_i \hat{\boldsymbol{x}}_i^{\mathrm{T}} \right) \\
&=\; \mathrm{trace} \left(\hat{\mathbf{R}} \, \hat{\mathbf{X}}^{\mathrm{T}} \right).
\end{aligned} \tag{18}$$

Here, we also assume the singular value decomposition $\hat{\mathbf{X}} = \mathbf{U} \, \mathbf{S} \, \mathbf{V}^{\mathrm{T}}$ is obtained, and substitute it into Eq. (18); then,

$$\begin{aligned}
\mathrm{trace} \left(\hat{\mathbf{R}} \, \hat{\mathbf{X}}^{\mathrm{T}} \right) \;&=\; \mathrm{trace} \left(\hat{\mathbf{R}} \, \mathbf{V} \, \mathbf{S} \, \mathbf{U}^{\mathrm{T}} \right) \\
&=\; \mathrm{trace} \left(\mathbf{U}^{\mathrm{T}} \, \hat{\mathbf{R}} \, \mathbf{V} \, \mathbf{S} \right) \\
&=\; \mathrm{trace} \left\{ \left(\mathbf{U}^{\mathrm{T}} \, \hat{\mathbf{R}} \, \mathbf{V} \right) \mathbf{S} \right\}.
\end{aligned} \tag{19}$$

We can say from Eq. (19), only diagonal section of $\mathbf{U}^{\mathrm{T}}\,\hat{\mathbf{R}}\,\mathbf{V}$ contributes to Eq. (18) because $\mathbf{S}$ is diagonal whose elements are all positive. Here $\mathbf{U}$, $\hat{\mathbf{R}}$ and $\mathbf{V}$ are all orthonormal matrices; therefore Eq. (18) becomes maximal when $\mathbf{U}^{\mathrm{T}}\,\hat{\mathbf{R}}\,\mathbf{V} = \mathbf{E}$. As discussion above, we can use $\hat{\mathbf{R}} = \mathbf{U}\,\mathbf{V}^{\mathrm{T}}$ using the singular decomposition of obtained $\hat{\mathbf{X}}$ as the rotation matrix. Due to discussion above, we should use $\hat{\mathbf{R}} = \mathbf{U}\,\mathbf{V}^{\mathrm{T}}$ as the rotation matrix which is mostly approximate to the identified $\hat{\mathbf{X}}$ with the least square error.

Next, we show a way to derive a coefficient for normalization a exploiting approximated $\hat{\mathbf{R}}$. We define elements of $\hat{\mathbf{R}}$ as,

$$\hat{\mathbf{R}} \;=\; \left[\begin{array}{ccc} \hat{\boldsymbol{\gamma}}_1 & \hat{\boldsymbol{\gamma}}_2 & \hat{\boldsymbol{\gamma}}_3 \end{array}\right]. \tag{20}$$

With a vector $\hat{\boldsymbol{\gamma}} = \left[\begin{array}{ccc} \hat{\boldsymbol{\gamma}}_1{}^{\mathrm{T}} & \hat{\boldsymbol{\gamma}}_2{}^{\mathrm{T}} & \hat{\boldsymbol{\gamma}}_3{}^{\mathrm{T}} \end{array}\right]^{\mathrm{T}}$, a is obtained as,

$$a \;=\; \frac{\left(\hat{\boldsymbol{e}}^{\mathrm{T}}\hat{\boldsymbol{x}}\right)}{\left(\hat{\boldsymbol{x}}^{\mathrm{T}}\hat{\boldsymbol{x}}\right)}, \tag{21}$$

which minimizes square error. We can calculate Sum of Square Difference (SSD) for each element of $a\hat{\mathbf{R}}^{\mathrm{T}}\hat{\mathbf{X}} - \mathbf{E}$, and regard it as closeness between the approximated $\hat{\mathbf{R}}$ and $\hat{\mathbf{X}}$. If this SSD is enough close to 0, it can be determined to satisfy the constraint condition of solution.

3.3 Visual body detection and its binding to body motion based on motion contingency

visual-motor coordination changes as the body changes because of tool holding, object attachment, or injury. To adapt to the changed coordination, it is necessary to detect the body in view. A robot should do it without preliminary knowledge because unfortunately we can not design and model all case of the changes.

We human can easily detect actively "controllable" objects (self-body, worn clothes or held tool) even when a lot of irrelevant objects are in view. We can do it even if the view is reflected or refracted, or even if the objects are not physically connected to the body [48]. Moreover, we can also find what motion controls the object. Even from infant, we detect a mobile tied to the body as well as the body itself [51]. Detection of self-body in view seems fundamental not only for robot but for our cognitive functions.

Several methods were proposed according to the capability of self-body detection. Yoshikawa [4] showed a method to leave image area invariant to motion. However, he did not include a way to bind the area to what motion causes it, nor his robot reuse the identification for motion generation. To address this binding problem, Fitzpatrick et al. [7] proposed a method using correlation between link motion and optical flow (unfortunately, they did not give their technical detail).

A method of Natale et al. [9] exploited periodic body motion. Our previously proposed method [2] focused on linear time-invariant relationship between coordinates of a body part and a object fixed to

the body part, and used the canonical correlation. These methods can actively detect a pair of body motion and its attached object. However, the method of Natale et al. takes enough time to determine the period of motion and restricts the robot motion to periodic one; meanwhile, our previous method takes less time but restricts the motion to planar one to calculate canonical correlation.

To compensate the previous methods, we believe it is required to support non-linear transformation such as refraction, reflection and perspective. In this subsection, we show a method satisfying the requirement using easy and fast but robust criterion.

3.3.1 Contingency detection with motion coincidence criterion

We assume a robot and its camera are stationary in the environment, and it can move only its manipulator. The environment includes objects with various motion, which are to be captured by the robot's camera. The robot should detect its body part in the view.

The body is a collection of particles. Here we define a point of particle in the 3D head coordinate system as $\boldsymbol{r}$, and the point in the 2D camera coordinate system as $\boldsymbol{x}$. With a function $\boldsymbol{f}$, which is derived by the camera parameters and the environment, $\boldsymbol{x}$ can be represented as,

$$\boldsymbol{x} \quad = \quad \boldsymbol{f}\,(\,\boldsymbol{r}\,). \tag{22}$$

We denote a point on "controllable" object p by $\boldsymbol{r}^{\mathrm{p}}$, sensor vector of an actuator b (e.g. 3D coordinates of body part, joint torque, or muscle tension) by $\boldsymbol{a}^{\mathrm{b}}$. The relationship between $\boldsymbol{a}^{\mathrm{b}}$ and $\boldsymbol{r}^{\mathrm{p}}$ can be represented with a function $\boldsymbol{g}$ as,

$$\boldsymbol{r}^{\mathrm{p}} \quad = \quad \boldsymbol{g}\,(\,\boldsymbol{a}^{\mathrm{b}}\,). \tag{23}$$

$\boldsymbol{g}$ is kinematics, which is usually non-linear if $\boldsymbol{a}^{\mathrm{b}}$ is joint angles; meanwhile, $\boldsymbol{g}$ is linear if $\boldsymbol{a}^{\mathrm{b}}$ is the position of a body part and p is attached to the body part. If both $\boldsymbol{f}$ and $\boldsymbol{g}$ were linear, the "controllable" object could be identified by the canonical correlation [2]. However, $\boldsymbol{f}$ often includes refraction, reflection or perspective depending on the environment; i.e. $\boldsymbol{f}$ and $\boldsymbol{g}$ are potentially non-linear and hard to calculate correlation.

Here we assume $\boldsymbol{f}$ and $\boldsymbol{g}$ are time-differentiable at $\boldsymbol{r}^{\mathrm{p}}$ and $\boldsymbol{a}^{\mathrm{b}}$ respectively, and apply time-differentiation to Eq. (22) and Eq. (23); then,

$$\dot{\boldsymbol{x}}^{\mathrm{p}} \quad = \quad \mathbf{J}^{\boldsymbol{f}}(\,\boldsymbol{r}^{\mathrm{p}}\,)\,\dot{\boldsymbol{r}}^{\mathrm{p}}, \tag{24}$$

$$\dot{\boldsymbol{r}}^{\mathrm{p}} \quad = \quad \mathbf{J}^{\boldsymbol{g}}(\,\boldsymbol{a}^{\mathrm{b}}\,)\,\dot{\boldsymbol{a}}^{\mathrm{b}}, \tag{25}$$

where $\mathbf{J}^{\boldsymbol{f}}$ and $\mathbf{J}^{\boldsymbol{g}}$ are Jacobian matrices of $\boldsymbol{f}$ and $\boldsymbol{g}$ respectively; and a dot over variable indicates its time-differentiation.

Due to Eq. (24) and Eq. (25), we can say independently from function form of $\boldsymbol{f}$ and $\boldsymbol{g}$,

$$\dot{\boldsymbol{a}}^{\mathrm{b}} = 0 \implies \dot{\boldsymbol{r}}^{\mathrm{p}} = 0 \implies \dot{\boldsymbol{x}}^{\mathrm{p}} = 0. \tag{26}$$

With a body composed of multiple links, a robot can actively move them in multiple direction. If we assume this universality of motion, $\mathbf{J}^{\boldsymbol{f}}$ and $\mathbf{J}^{\boldsymbol{g}}$ are full rank ($n(\mathbf{J}^{\boldsymbol{f}}) = 0$, $n(\mathbf{J}^{\boldsymbol{g}}) = 0$) as is often the case. Accepting this assumption, Eq. (26) could be equivalence relationship. That is, "generally," $\dot{\boldsymbol{a}}^{\mathrm{b}} \neq 0 \implies \dot{\boldsymbol{x}}^{\mathrm{p}} \neq 0$ stands. This equivalence means $\dot{\boldsymbol{a}}^{\mathrm{b}}$, $\dot{\boldsymbol{r}}^{\mathrm{p}}$ and $\dot{\boldsymbol{x}}^{\mathrm{p}}$ get 0 or get not 0 simultaneously (**motion coincidence**).

Even when $\mathbf{J}$ is non-linear and unknown, Eq. (26) stands; i.e. only Eq. (26) is "generally" valid. That is why the motion coincidence should be used as an invariant clue to detect body and attached object. The motion coincidence clue indicates naive feeling that "there is no body movement in view without any body motion;" the derivation of Eq. (26) corresponds to theoretical grounds for the naive feeling.

3.3.2 Visual body detection with motion coincidence

The motion coincidence clue derived in Sec. 3.3.1 is applicable to actively detect self-body, worn clothes and held tools. Here, we show a method to detect body parts in view using the motion coincidence in Fig. 1.

We denote temporal change of i-th actuator as $\dot{\boldsymbol{a}}^i$, temporal feature change in view of j-th object as $\dot{\boldsymbol{s}}^j$, and likelihood that j-th object is moved by i-th actuator as ϵ^{ij}. The likelihood ϵ^{ij} is incrementally updated according to the equation as,

$$\Delta\epsilon^{ij} \quad = \quad \rho^a\,(\,\dot{\boldsymbol{a}}^i\,) \,\cap\, \rho^s\,(\,\dot{\boldsymbol{s}}^j\,), \tag{27}$$

where ρ^a and ρ^s are functions whose output is 1 if any element of input vector rises up from 0 during duration τ^a and τ^s respectively; otherwise, output is 0; and the symbol $\cap$ is a operator whose output is 1 if remaining time of left term is less than that of right term; otherwise, output is 0. Eq. (27) indicates the likelihood increases if visual motion occurs immediately after body motion, in which causal order is reflected.

The duration τ assures robustness over temporal difference of transmission depending on a type of each sensor; i.e. τ^a and τ^s defines the performance of Eq. (27). We believe those values could be given from other experiments or physiological literatures (e.g. [49]). As for human, they seem around $\tau \simeq 300$ [msec] [52].

The increment of ϵ^{ij} is also usable as a trigger of motion which confirms if j-th object is truly "controllable." If ϵ^{ij} is increased when $\dot{\boldsymbol{a}}^i = \dot{\boldsymbol{a}}^i_t$, the robot inputs $\dot{\boldsymbol{a}}^i_t$ into its motion generator, and moves with perturbed $\boldsymbol{a}^i_t$ and $\dot{\boldsymbol{a}}^i_t$. This confirming motion accelerates the determination which element of $\dot{\boldsymbol{a}}^i_t$ causes $\dot{\boldsymbol{s}}^i_t$. Of course, ϵ^{ij} is not necessary to change during the time before the confirming motion. This detection method is more generalized and simpler than the method of Natale [8, 9] using periodic motion.

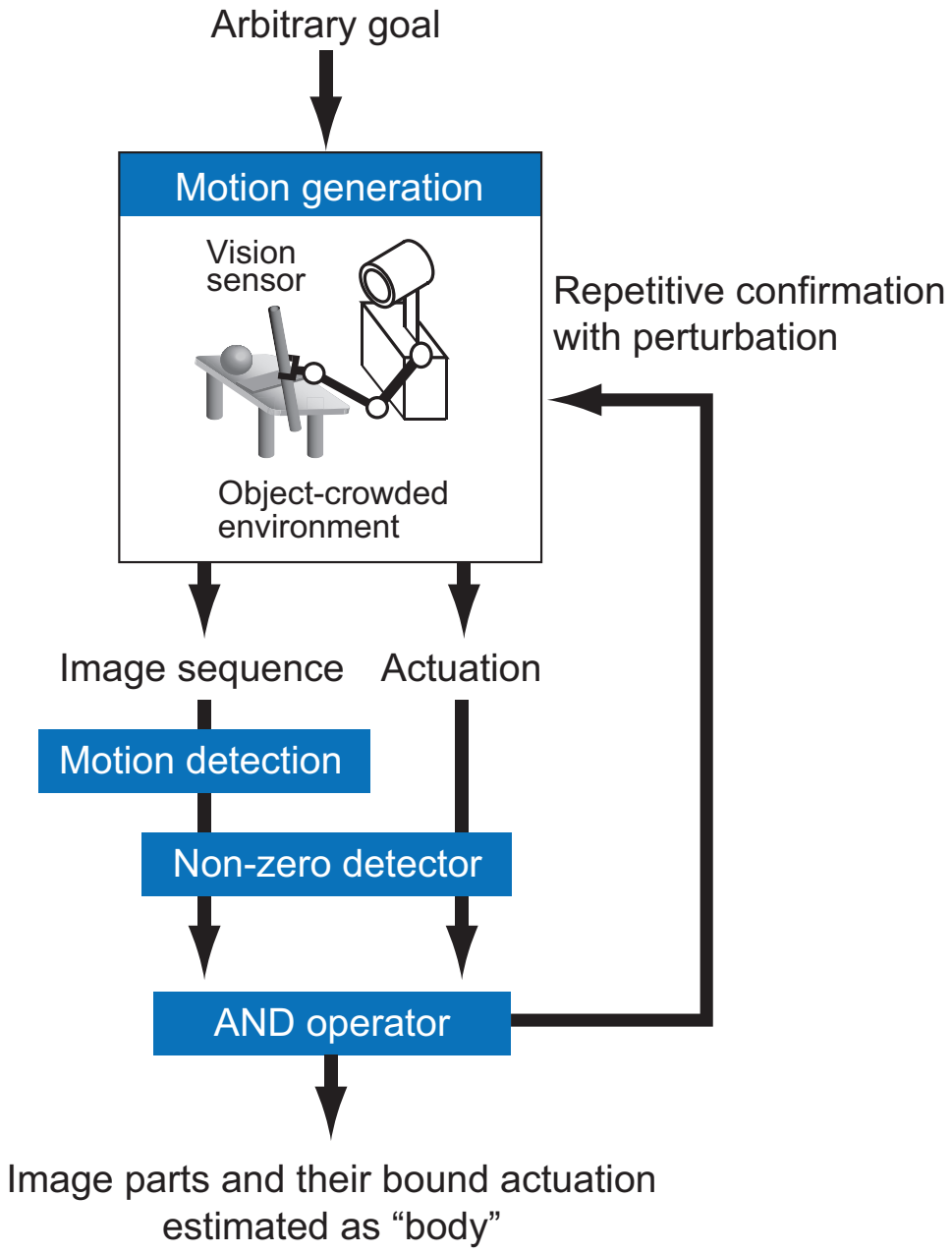

Figure 1: Schematic diagram of body detection based on motion coincidence. It consists of non-zero detectors and AND-like operators. It iteratively discovers which pair in object and actuator has a causal relationship of motion.

3.4 Adaptation of kinematic hand-eye coordination without known marker

Spatial hand-eye coordination is necessary to plan and generate behaviors such as object avoidance, reaching, or handling. That is, sustainable hand-eye coordination is a similar problem of hand-eye calibration, where coordinate transformation is identified as initialization.

Conventional formulation assumes a given marker which has sorted and salient pattern attached to the end-effector [1]; and the given marker coordinate ${}^{\mathcal{C}}_{\mathcal{P}}\mathbf{H}$ (homogeneous transformation matrix) is obtained by sensors e.g. stereo camera. The transformation relationship among the camera coordinate to the world coordinate ${}^{\mathcal{W}}_{\mathcal{C}}\mathbf{H}$ (fixed and unknown), the hand coordinate to the world coordinate ${}^{\mathcal{W}}_{\mathcal{H}}\mathbf{H}$ (variable and known), and the given marker coordinate to the hand coordinate ${}^{\mathcal{H}}_{\mathcal{P}}\mathbf{H}$ (fixed and unknown) is as,

$$ {}^{\mathcal{W}}_{\mathcal{C}}\mathbf{H}\,{}^{\mathcal{C}}_{\mathcal{P}}\mathbf{H} \;=\; {}^{\mathcal{W}}_{\mathcal{H}}\mathbf{H}\,{}^{\mathcal{H}}_{\mathcal{P}}\mathbf{H}, \tag{28} $$

which is shown in Fig. 2-A. The conventional methods formulate the calibration as identification of unknown variables by solving Eq. (28).

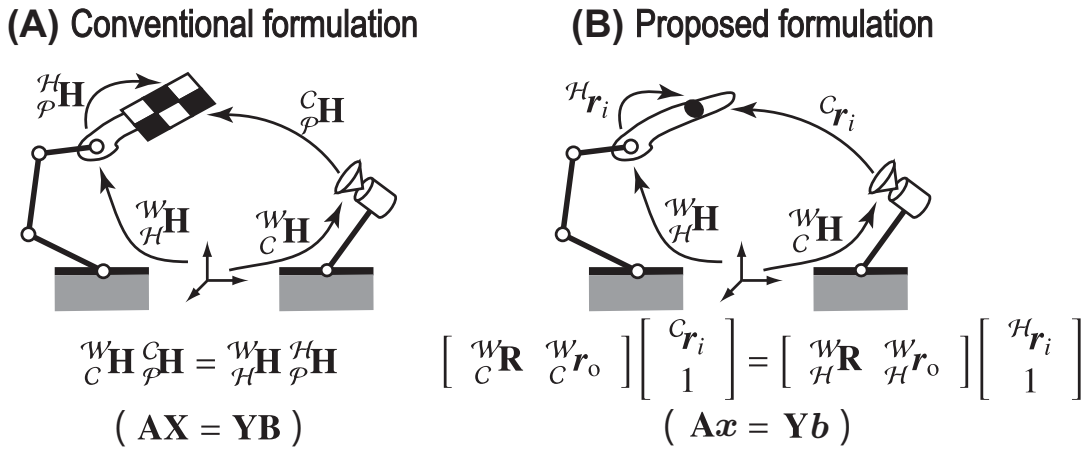

Figure 2: Conventional and proposed formulation of visual-motor coordination.

Eq. (28) has 12 DOF because each homogeneous transformation matrix has 6 DOF. Ideally, 12 equations which compare elements of left and right hand sides are obtained, and one measurement is enough for identification. However, if we replace fixed and unknown matrices as $\mathbf{A}$ and $\mathbf{B}$, and variable and known matrices as $\mathbf{X}$ and $\mathbf{Y}$, Eq. (28) can be represented as,

$$ \mathbf{AX} \;=\; \mathbf{YB}. \tag{29} $$

With normal matrix calculation, Eq. (29) is not solvable; i.e. existence of solution is not assured because variables are wedged between unknown matrices. To assure the existence of solution by multiple measurement, meanwhile, it is not apparent how to exploit $\mathbf{X}$ and $\mathbf{Y}$ samples obtained by each measurement

[1] Conventionally, calibration methods without a marker only assume the camera is mounted on the hand, and the environment is stationary. Under this assumption, camera does not capture visual image of self-body.

in the form of Eq. (29). Therefore, robust estimation has been often used [53, 54]. However, it takes a lot of memory according to measurement times and its computation time is indeterminate; moreover, there are still problems in assurance of solution existence, sampling method from the measurements and termination condition of identification. That is why direct usage of conventional methods is not suitable for sustainable coordination.

In conventional methods, ${}^{\mathcal{C}}_{\mathcal{P}}\mathbf{H}$ is assumed to be given by measurement. To obtain ${}^{\mathcal{C}}_{\mathcal{P}}\mathbf{H}$ from monocular/stereo camera or laser range finder, it is necessary to prepare a marker whose coordinate can be uniquely determined. The methods finds the marker from camera captured images, and estimates the marker coordinate to the camera coordinate. Note that the marker should have at least three distinguishing points for estimation. Unfortunately, in the case of body changes caused by attachment of unexpected object, we can not generally assume the marker or pattern are given; i.e. we need a "marker-free" method for sustainable coordination.

Here, we implement sustainable visual-motor coordination method exploiting body detection method of Sec. 3.3 and parameter identification method of Sec. 3.1. Owing to those methods, our coordination method has preferable properties to its sustainability: less distinguishing point fixed in the hand coordinate (at least 1 point), free motion during calculation, and adequate process beginning and termination based on successive sensor input.

We define i-th point fixed in the hand coordinate as ${}^{\mathcal{H}}\boldsymbol{r}_i$ (fixed and unknown) and ${}^{\mathcal{C}}\boldsymbol{r}_i$ (variable and known) in the hand coordinate and the camera coordinate respectively. Candidates of the point can be detected in view by the method of Sec. 3.3. Now, the relationship between the hand coordinate and the camera coordinate comes as,

$$\begin{bmatrix} {}^{\mathcal{W}}_{\mathcal{C}}\mathbf{R} & {}^{\mathcal{W}}_{\mathcal{C}}\boldsymbol{r}_{\mathrm{o}} \end{bmatrix} \begin{bmatrix} {}^{\mathcal{C}}\boldsymbol{r}_i \\ 1 \end{bmatrix} = \begin{bmatrix} {}^{\mathcal{W}}_{\mathcal{H}}\mathbf{R} & {}^{\mathcal{W}}_{\mathcal{H}}\boldsymbol{r}_{\mathrm{o}} \end{bmatrix} \begin{bmatrix} {}^{\mathcal{H}}\boldsymbol{r}_i \\ 1 \end{bmatrix}, \tag{30}$$

which is the same form as Eq. (28) (Fig. 2-B).

By replacing ${}^{\mathcal{W}}_{\mathcal{C}}\mathbf{R}$ and ${}^{\mathcal{C}}\boldsymbol{r}_i$ in Eq. (30) with,

$${}^{\mathcal{W}}_{\mathcal{C}}\mathbf{R} = \begin{bmatrix} {}^{\mathcal{W}}_{\mathcal{C}}\boldsymbol{\gamma}_1 & {}^{\mathcal{W}}_{\mathcal{C}}\boldsymbol{\gamma}_2 & {}^{\mathcal{W}}_{\mathcal{C}}\boldsymbol{\gamma}_3 \end{bmatrix}, \tag{31}$$

$${}^{\mathcal{C}}\boldsymbol{r}_i = \begin{bmatrix} {}^{\mathcal{C}}x_i & {}^{\mathcal{C}}y_i & {}^{\mathcal{C}}z_i \end{bmatrix}^{\mathrm{T}}, \tag{32}$$

the left side of Eq. (30) can be transformed as,

$$\begin{bmatrix} {}^{\mathcal{W}}_{\mathcal{C}}\mathbf{R} & {}^{\mathcal{W}}_{\mathcal{C}}\boldsymbol{r}_{\mathrm{o}} \end{bmatrix} \begin{bmatrix} {}^{\mathcal{C}}\boldsymbol{r}_i \\ 1 \end{bmatrix} = \begin{bmatrix} {}^{\mathcal{C}}x_i\mathbf{E} & {}^{\mathcal{C}}y_i\mathbf{E} & {}^{\mathcal{C}}z_i\mathbf{E} & \mathbf{E} \end{bmatrix} \begin{bmatrix} {}^{\mathcal{W}}_{\mathcal{C}}\boldsymbol{\gamma}_1{}^{\mathrm{T}} & {}^{\mathcal{W}}_{\mathcal{C}}\boldsymbol{\gamma}_2{}^{\mathrm{T}} & {}^{\mathcal{W}}_{\mathcal{C}}\boldsymbol{\gamma}_3{}^{\mathrm{T}} & {}^{\mathcal{W}}_{\mathcal{C}}\boldsymbol{r}_{\mathrm{o}}{}^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}}. \tag{33}$$

With the transformed left side Eq. (33), we transform Eq. (30) so that unknown and known variables get together respectively; then,

$$\begin{bmatrix} {}^{\mathcal{C}}x_i\mathbf{E} & {}^{\mathcal{C}}y_i\mathbf{E} & {}^{\mathcal{C}}z_i\mathbf{E} & \mathbf{E} & -{}^{\mathcal{W}}_{\mathcal{H}}\mathbf{R} & -{}^{\mathcal{W}}_{\mathcal{H}}\boldsymbol{r}_{\mathrm{o}} \end{bmatrix} \begin{bmatrix} {}^{\mathcal{W}}_{\mathcal{C}}\boldsymbol{\gamma}_1{}^{\mathrm{T}} & {}^{\mathcal{W}}_{\mathcal{C}}\boldsymbol{\gamma}_2{}^{\mathrm{T}} & {}^{\mathcal{W}}_{\mathcal{C}}\boldsymbol{\gamma}_3{}^{\mathrm{T}} & {}^{\mathcal{W}}_{\mathcal{C}}\boldsymbol{r}_{\mathrm{o}}{}^{\mathrm{T}} & {}^{\mathcal{H}}\boldsymbol{r}_i{}^{\mathrm{T}} & 1 \end{bmatrix}^{\mathrm{T}} = 0. \tag{34}$$

Eq. (34) is the same form as Eq. (1), and its variance-covariance matrix updated by each measurement

is,

$$\mathbf{C}_t = \frac{1}{n}\sum_t \mathbf{A}_t^{\mathrm{T}}\mathbf{A}_t \tag{35}$$

$$\text{under } \mathbf{A}_t = \left[\begin{array}{cccccc} {}^{\mathcal{C}}x_i\mathbf{E} & {}^{\mathcal{C}}y_i\mathbf{E} & {}^{\mathcal{C}}z_i\mathbf{E} & \mathbf{E} & -{}^{\mathcal{W}}_{\mathcal{H}}\mathbf{R} & -{}^{\mathcal{W}}_{\mathcal{H}}\boldsymbol{r}_{\mathrm{o}} \end{array}\right].$$

The parameters to be identified is,

$$\left[\begin{array}{ccccccc} {}^{\mathcal{W}}_{\mathcal{C}}\boldsymbol{\gamma}_1{}^{\mathrm{T}} & {}^{\mathcal{W}}_{\mathcal{C}}\boldsymbol{\gamma}_2{}^{\mathrm{T}} & {}^{\mathcal{W}}_{\mathcal{C}}\boldsymbol{\gamma}_3{}^{\mathrm{T}} & {}^{\mathcal{W}}_{\mathcal{C}}\boldsymbol{r}_{\mathrm{o}}{}^{\mathrm{T}} & {}^{\mathcal{H}}\boldsymbol{r}_i{}^{\mathrm{T}} & 1 \end{array}\right]^{\mathrm{T}}. \tag{36}$$

The reliability of identification is calculated as how close to 0 the minimal singular value λ_{16} of $\mathbf{C}_t$ is, how smaller λ_{16} is than λ_{15}, and how close to rotation matrix the matrix whose elements are ${}^{\mathcal{W}}_{\mathcal{C}}\boldsymbol{\gamma}_1$, ${}^{\mathcal{W}}_{\mathcal{C}}\boldsymbol{\gamma}_2$ and ${}^{\mathcal{W}}_{\mathcal{C}}\boldsymbol{\gamma}_3$ in obtained $\boldsymbol{v}_{16}$ (see Sec. 3.2). $\boldsymbol{v}_{16}$ is scaled and normalized so that its last element becomes 0.

Compared to conventional hand-eye calibration methods, our method is marker-free, more real-time, easily decidable of convergence and enough accurate. Moreover, it can determine which body segment has the distinguishing point in view thanks to the detection method of Sec. 3.3. That means our method is easily applicable not only for hand-eye but body-eye calibration; i.e. visual-motor coordination. Definitely, our method is usable for head-eye calibration, which is a subset of hand-eye calibration.

3.5 Adaptation of kinesthetia: inertia identification and tactile extension

Identification of the inertia parameters of held or adhered object is necessary to estimate a touch on it [2]. Inertia identification is a field of robotics, and multiple methods have been proposed (see Sec. 2). However, they are not suitable for sustainable coordination because of direct usages of pseudo-inverse matrix, robust estimation or Kalman filter.

As is expected, motion equation is known to be linearizable with inertia parameters in the same way as visual-motor coordination even when the body consists of multiple links [16]. This indicates our sustainable coordination method of Sec. 3.1 can be applied to inertia identification. With our method, a robot can automatically terminate the identification process and use the obtained parameters just after the termination; e.g. touch estimation on the object whose inertia parameters are identified [2]. We note that our method is also applicable to identification of spatial alignment between kinesthetic (force/torque) sensor coordinate and the hand coordinate; i.e. spatial calibration.

4 Experiments

4.1 Experimental setup

To validate our method, we built a robotic system. The robot has a 6-DOF (5-DOF is for arm, the other is for gripper) robot arm "Katana" [55], a 6-axis force/torque sensor "Mini40 SI-80-4" [56] equipped in the wrist, 2-DOF head "Biclops PT" [57], and two "Firefly MV" [58] for a stereo camera.

We adopted a Linux-2.6.17.3 PC with xenomai patch [59] for real-time sensor measurement, OpenCV [60] and Integrated Performance Primitives (IPP) [61] for image processing, C++ for programming, and In-

tel C++ Compiler for compilation. Any vision process was operated every 33 [msec] tuned to sampling rate of the cameras.

4.2 Visual body detection with motion coincidence

We experiment visual body detection with motion coincidence, which is shown in Sec. 3.3. We do not give the robot knowledges of any optical condition and kinematics of the arm, where the cameras are not calibrated. The robot should instantly detect its body from the view including irrelevant objects and noise.

We implement $\dot{a}^i$ in Eq. (27) as the temporal change of encoder values of joints, and $\dot{s}^j$ as the sum of norm of optical flow within subimage j calculated by cvCalcOpticalFlowPyrLK() function in OpenCV. Note that $\boldsymbol{f}$ and $\boldsymbol{g}$ are non-linear and not given to the robot.

The robot randomly and repeatedly moves its arm and detects its body in view. If ϵ^{ij} is incremented by Eq. (27), it interrupts motion generation, moves the arm to the posture at which ϵ^{ij} was incremented, and moves it again (confirmative motion).

A result is shown in Fig. 3 from the beginning of motion (0 [msec]) to the end of duration (300 [msec]). Blue rectangles in the figure indicate optical flow is non-zero with in the area; and red ones do the area detected as body, where $\sum_{i=0}^{5} \epsilon^{ij} \neq 0$.

At 0 [msec], visual motion of human—irrelevant object did not effect ϵ^{ij}. This is because our method uses only visual motion just after self-body motion. The finally obtained likelihood map was overlapped on the visible area of the body.

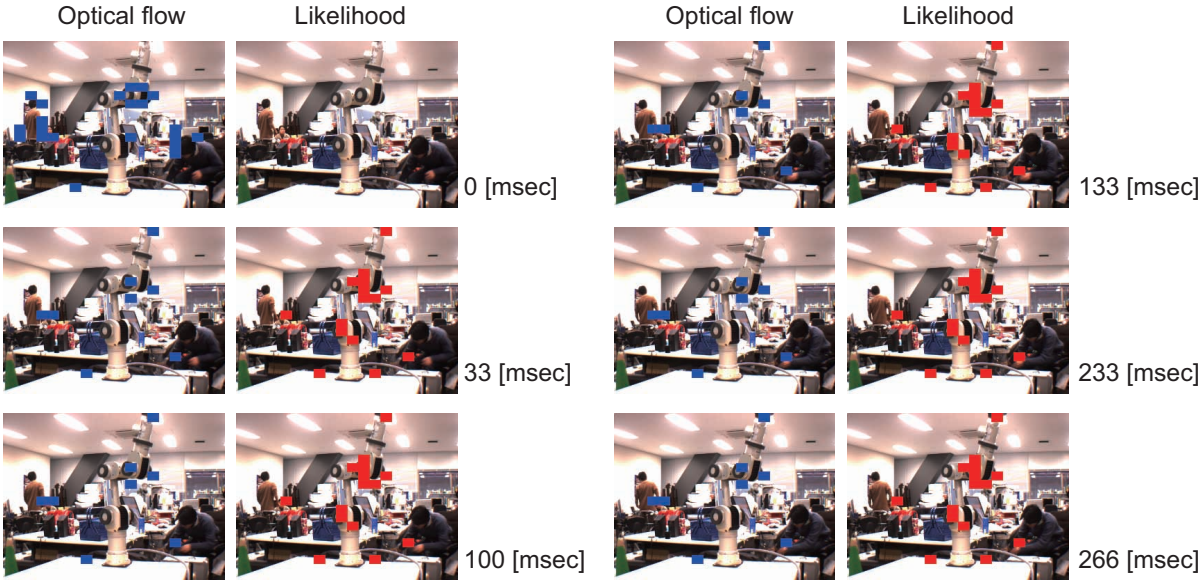

Figure 3: Temporal change of the resultant optical flow and likelihood maps indicating area of body. The blue rectangles indicates the segment whose optical flow is non-zero. The red rectangles are discovered as controllable segments in the image.

Fig. 4 shows resultant $\dot{a}^i$ and ϵ^{ij}. The graph of $\dot{a}^i$ begins right before the beginning of Fig. 3. The values of encoder 0, 1, 3 and 4 simultaneously rise up from 0, and those of encoder 2 and 5 dose not change (encoder 5 is stationary from -266 [msec]). $\dot{a}^i$ properly reflects in the likelihood map

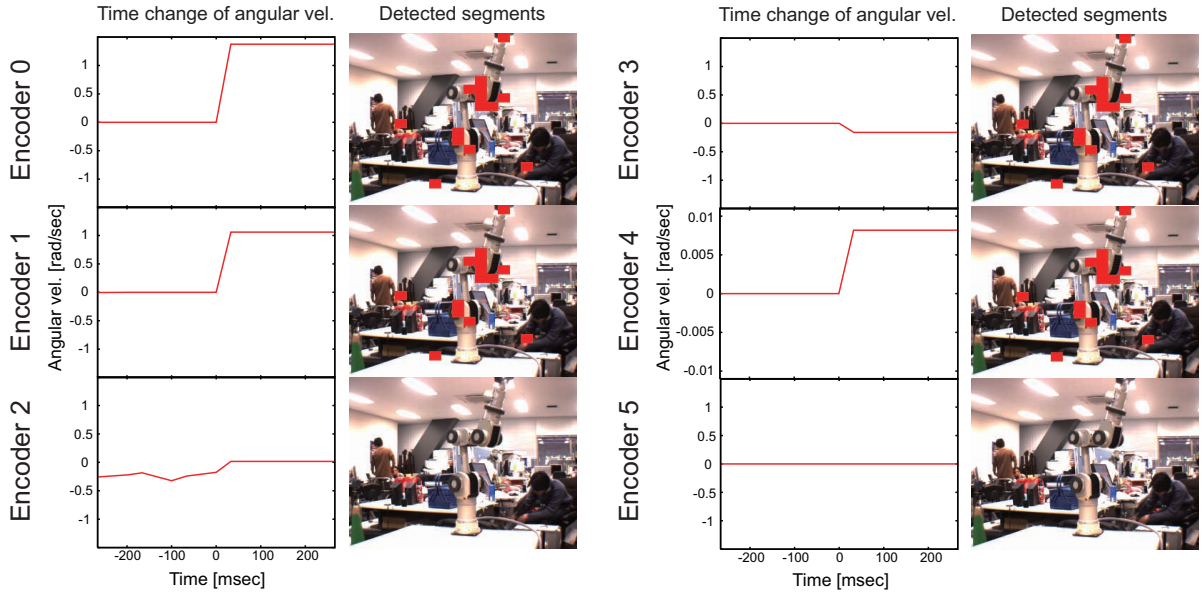

Figure 4: Temporal change of the encoders' output and resultant likelihood maps indicating area of body. The red rectangles are discovered as controllable segments in the image.

Figure 5: Blend image of 15 likelihood maps obtained from 15 iterations of motions. The redder rectangles means the segments stably discovered as controllable.

To verify the convergence of our method, we overlay likelihood maps obtained from 15 randomly generated motions in Fig. 5. Areas which is stably detected as body are indicated redder. Our method detected well in distinguishing points and tips of each links, where optical flow can be accurately calculated.

4.3 Kinematic hand-eye coordination without known marker

We experiment sustainable visual-motor coordination shown in Sec. 3.4. The world coordinate is set at the base of the arm, and other coordinates (the hand coordinate and the camera coordinate) are as shown in Fig. 6. From the beginning of experiment, the robot holds a rod-like object (file, which is usally seen and used in our daily life) with around 0.3 [m] length; the robot moves its arm and file randomly. We assume the stereo camera is calibrated and parallelized.

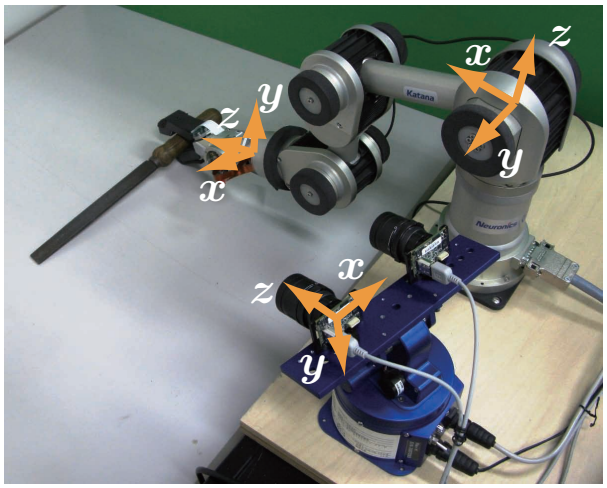

Figure 6: Experimental setup and definition of coordinate systems for sustainable hand-eye coordination.

We show the experimental processing flow in Fig. 7. Visual processing of the left images consists of detection of corners as distinguishing points, tracking of the corners, and detection of corners on body using our method of Sec. 3.3 with input of the tracking information and body motion information. Exploiting right images, depth of those corners on body are calculated. The obtained 3D coordinates of the corners on the body and the motion information are input into our method of Sec. 3.4.

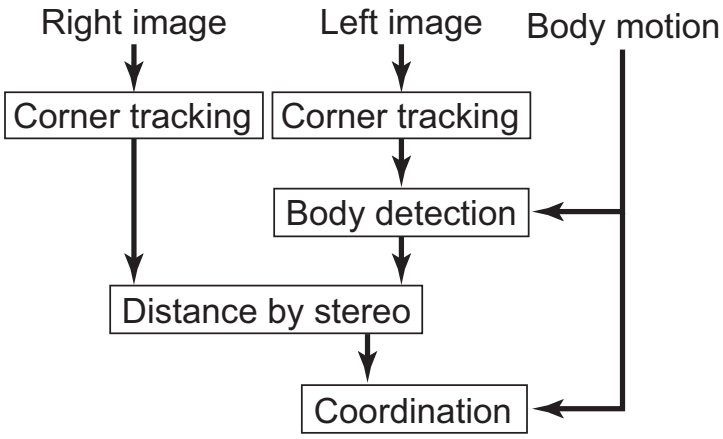

Figure 7: Information flow for experiment of hand-eye coordination.

The identification is calculated at every input, and its termination is determined when the minimal

singular value (λ_{16}) is less than 0.00005 and closeness of solution to rotation matrix (SSD) than 0.1 (see Sec. 3.2). We set $100\lambda_{16}$ as the upper limit of error. The robot stops for 3 [sec] after identification, and randomly moves again. The robot calculates the error of sensor data obtained at each step, and check whether it is under the upper limit. Re-identification begins after 10 consecutive inputs which are over the upper limit. For experiment, the experimenter moves the rod-like object held by the robot; i.e. the hand coordinate changes; and moves the robot's head; i.e. the camera coordinate changes.

Results are shown in Fig. 8. The upper half of each image indicates, starting from the left, the image captured by the left camera, that by the right camera, and internal state. As the internal state, current process (identification or observation) is shown. The lower half is external experimenter's view. In those images, blue circles indicate detected corners; green lines connecting corners do stereo-matched pair with local image around each corner; and red circles in the left images do corners which is detected to be the body. one of red circles changes to a red rectangle after the end of identification, which means the identification using the corner completes its process at first of the others.

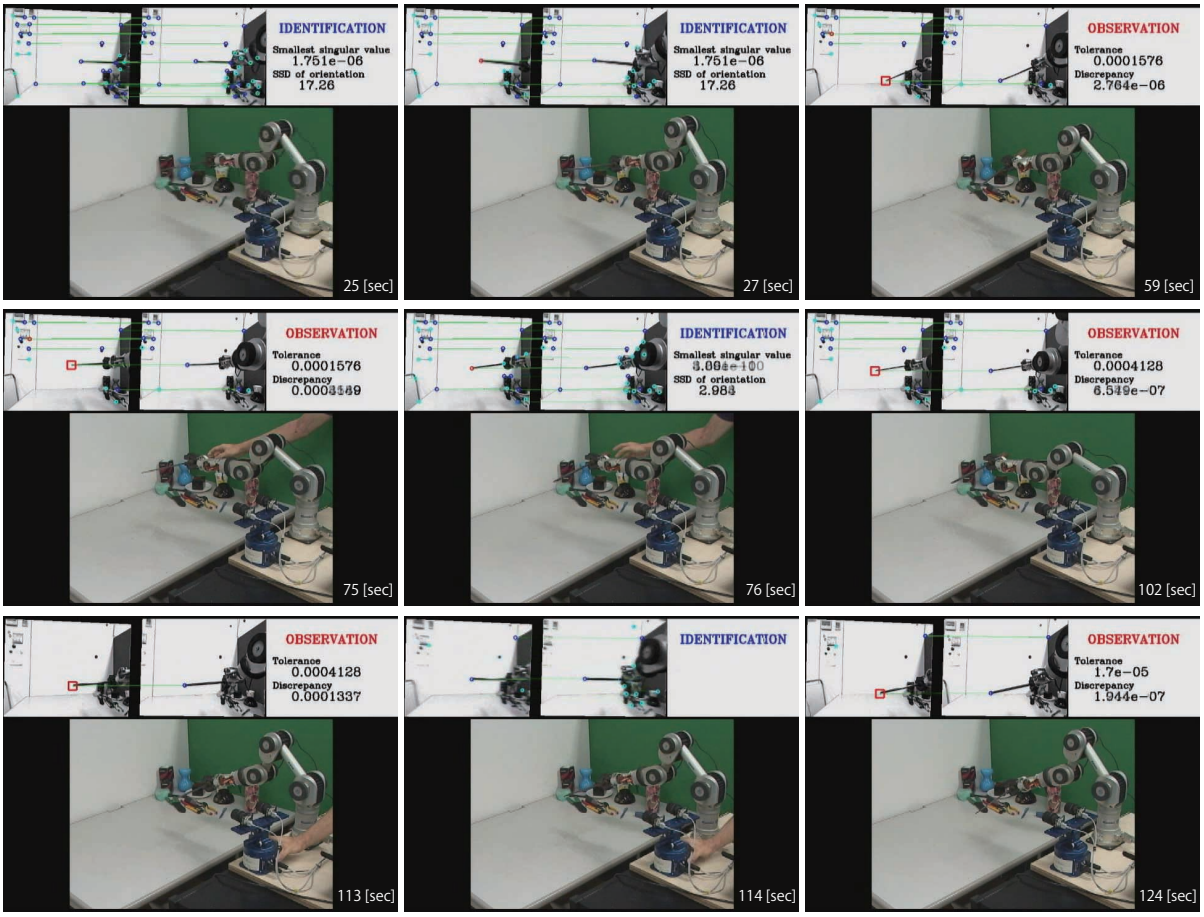

Figure 8: Resultant sequence of experiment of hand-eye coordination.

From 25 [sec] to 27 [sec] in our experiment, the tip of the rod-like object held by the robot wasdetected as the body. The identification converged by 59 [sec] and the process was switched to observation of

the error. The rod-like object was moved furtively by the experimenter at 75 [sec], and the process was changed to identification at 76 [sec]. The renewed identification converged by 102 [sec] and the process was changed to observation again. The experimenter moved the robot head at 113 [sec]. Responding to the change, the process was switched to identification again at 114 [sec]. The third identification converged by 124 [sec] and the process was changed to observation again.

Fig. 9 shows the temporal change of identification reliability in its upper row, and the temporal change of observation error in its lower row. Horizontal axis is time; Vertical axis of the top row is singular values from λ_{13} to λ_{16}; The second top is SSD to rotation matrix in range from 0 to 3.5; the blue line in the bottom is $100\lambda_{16}$ using previously identified λ_{16}; and the red line in the bottom is error calculated by Eq. (2).

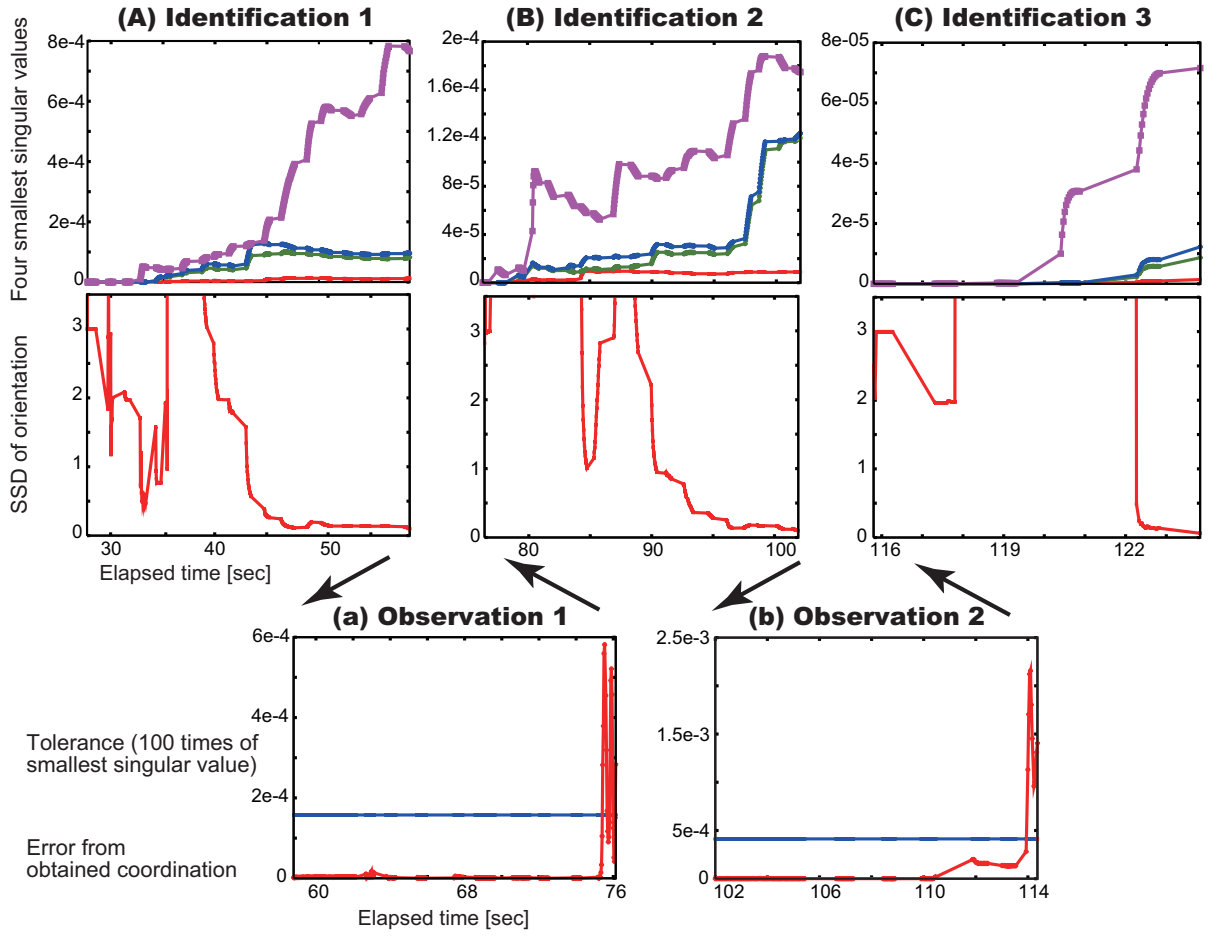

Figure 9: Resultant internal values during hand-eye coordination.

In all identification process, minimal singular value λ_{16} was almost always close to 0; others increased from 0; and SSD to rotation matrix converged to 0. These results indicate our method achieved identification stably. In all observation process, the error never transcended the automatically determined limit before the experimenter moved the rod-like object and changed visual-motor coordination of the robot. The transcending error lasted in several time, and the process switch to identification occurred.

For reference, the parameters obtained in each identification process are,

$$
{}^{\mathcal{W}}_{\mathcal{C}}\mathbf{R} = \left[\begin{array}{ccc} -0.0498065 & -0.998648 & -0.0148783 \\ 0.170921 & 0.00615438 & -0.985266 \\ 0.984025 & -0.0516156 & 0.170383 \end{array}\right], \left[\begin{array}{ccc} -0.00316324 & -0.999583 & -0.0287144 \\ 0.143567 & 0.0279631 & -0.989246 \\ 0.989636 & -0.00725165 & 0.143418 \end{array}\right],
$$

$$
\left[\begin{array}{ccc} 0.00638279 & -0.999756 & 0.0211358 \\ 0.0562159 & -0.020744 & -0.998203 \\ 0.998398 & 0.00755949 & 0.0560698 \end{array}\right],
$$

$$
{}^{\mathcal{W}}_{\mathcal{C}}\boldsymbol{r}_{\mathrm{o}} = \left[\begin{array}{c} 0.114014 \\ 0.360626 \\ -0.118095 \end{array}\right], \left[\begin{array}{c} 0.119858 \\ 0.362371 \\ -0.0912571 \end{array}\right], \left[\begin{array}{c} 0.0691735 \\ 0.37749 \\ -0.0511823 \end{array}\right],
$$

$$
{}^{\mathcal{H}}\boldsymbol{r}_i = \left[\begin{array}{c} 0.287573 \\ -0.0457856 \\ 0.307632 \end{array}\right], \left[\begin{array}{c} 0.299869 \\ -0.0494941 \\ 0.267815 \end{array}\right], \left[\begin{array}{c} 0.29829 \\ -0.0664128 \\ 0.234654 \end{array}\right] [\mathrm{m}].
$$

We can say these values adequately reflected the spatial relationship shown in Fig. 6. Particularly, x value of ${}^{\mathcal{H}}\boldsymbol{r}_i$ was nearly 0.3 [m], which is the length of the rod-like object.

4.4 Kinesthetic-motor coordination and touch estimation

We experiment kinesthetic-motor coordination; i.e. inertia identification and touch estimation of held object (see Sec. 2 and Sec. 3.5). Because the robot has a force/torque sensor in its wrist, we set a task where it should identify inertia parameters of distal part from the wrist (including its held object), and immediately after that, it should estimate touch on the part.

The experiment begins with the robot holding an object. The shape of the object is approximately given, which is necessary to estimate touch in principle [2].

As for the experimental procedure, the robot moves the object, and identifies its inertia parameters. If those parameters are identified, the robot makes the object touch another object. For demonstration, a simple feedback rule is implemented, where the robot repetitively keeps touch if first touch is detected by estimation.

Currently available force/torque sensor can not inevitably cancel fictitious force which is caused by accelerated motion of the sensor itself. We carry out the experiment under low acceleration to ignore its effect. Now we define the variance-covariance matrix updated by each measurement as,

$$
\mathbf{C}_t = \frac{1}{n}\sum {\mathbf{A}_t}^{\mathrm{T}}\mathbf{A}_t \tag{37}
$$

$$
\text{under } \mathbf{A}_t = \left[\begin{array}{ccc} {}^{\mathcal{W}}_{\mathcal{K}}\mathbf{R}\,{}^{\mathcal{K}}\boldsymbol{f} & 0 & \boldsymbol{g} \\ 0 & -\left[{}^{\mathcal{K}}\boldsymbol{f}\times\right] & -{}^{\mathcal{K}}\boldsymbol{\tau} \end{array}\right],
$$

where ${}^{\mathcal{W}}_{\mathcal{K}}\mathbf{R}$ is the kinesthetic sensor coordinate to the world coordinate, ${}^{\mathcal{K}}\hat{\boldsymbol{f}}$ and ${}^{\mathcal{K}}\hat{\boldsymbol{\tau}}$ are touch force/torque in the kinesthetic sensor coordinate respectively, and $\boldsymbol{g}$ is gravity acceleration to the world coordinate. The inertia parameters to be identified are defined as,

$$\left[\begin{array}{ccc} 1/m & {}^{\mathcal{K}}_{\mathcal{T}}\boldsymbol{r}^{\mathrm{T}} & 1 \end{array}\right]^{\mathrm{T}}. \tag{38}$$

Note that these formulation is a minor version of the method which Liu et al. proposed in [16].

Results in the case that a wire-mesh box is held are shown in Fig. 10. The left top of each figure indicates internal view of the robot, where estimated points of touch are represented as red dots. We show the reliability of identification in Fig. 11-(A), where the horizontal axis is time from the beginning of the experiment and the vertical is singular values from λ_2 to λ_5. Resultant 3D coordinates of the points where the robot estimated touch are in Fig. 11-(B), whose projection onto y-z plane is Fig. 11-(b).

The robot automatically determined that the identification finished at 6.6 [sec] because λ_4 reached to the threshold. Just after determination, it performed touching motion to another object (table). From Fig. 11-(A), we can see that λ_4 increased and converged while λ_2 and λ_3 decreased. This could be an effect of object's welter because of non-rigidness of the wire-mesh box. The identified parameters were,

$$\begin{aligned} m &= 0.594409\ [\mathrm{kg}], \\ {}^{\mathcal{K}}_{\mathcal{T}}\boldsymbol{r}^{\mathrm{T}} &= \left[\begin{array}{c} 0.0161846 \\ -0.0214352 \\ 0.0923942 \end{array}\right] [\mathrm{m}]. \end{aligned}$$

The resultant touch points should be on a plane composing the table. However, the error up to 0.1 [m] occurred in the estimated value (see Fig. 11-(B)). We suppose it is because of modelling error of shape and non-rigidness.

5 Discussions

Our results obtained in Sec. 4 indicate computational lightness of our method, which could be implemented even on a microcomputer with small area of memories and low power. Of course, singular value decomposition is costly. Fortunately, matrices to which applied it in our method are small and fixed-size; i.e. we are able to assume an upper limit of computational cost.

In our method, several thresholds should be given as well as conventional methods. However, we suppose these thresholds can be automatically determined according to the scale of input or to necessary precision; e.g. the constraint of rotation matrix is always normalized and we can easily decide its threshold.

To implement the sustainable sensory-motor coordination, we assumed its linearity. There are still problems if we extend our method in case of non-linear coordination; could it work stably and rapidly? We are afraid that such a extension will spoils the features of our method.

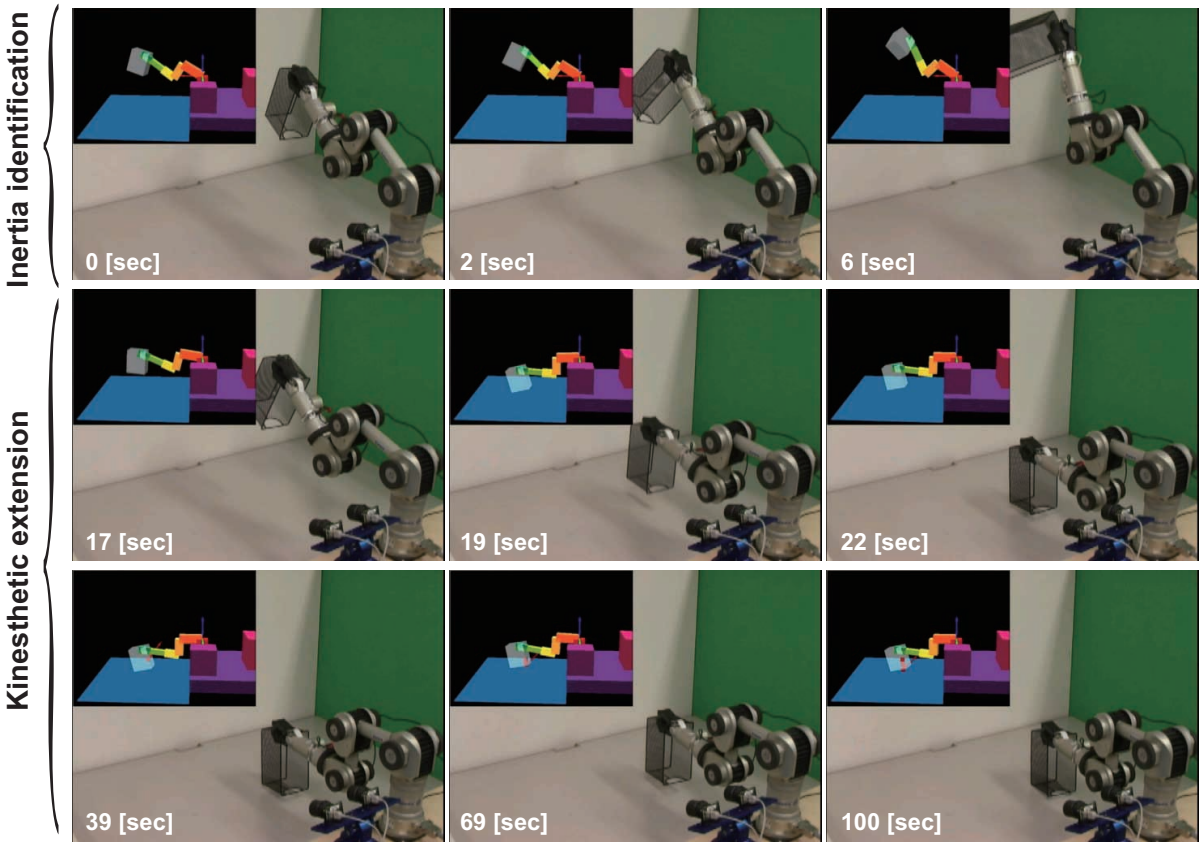

Figure 10: Experiment of kinesthetic coordination and tactile extension of grasped box (tool).

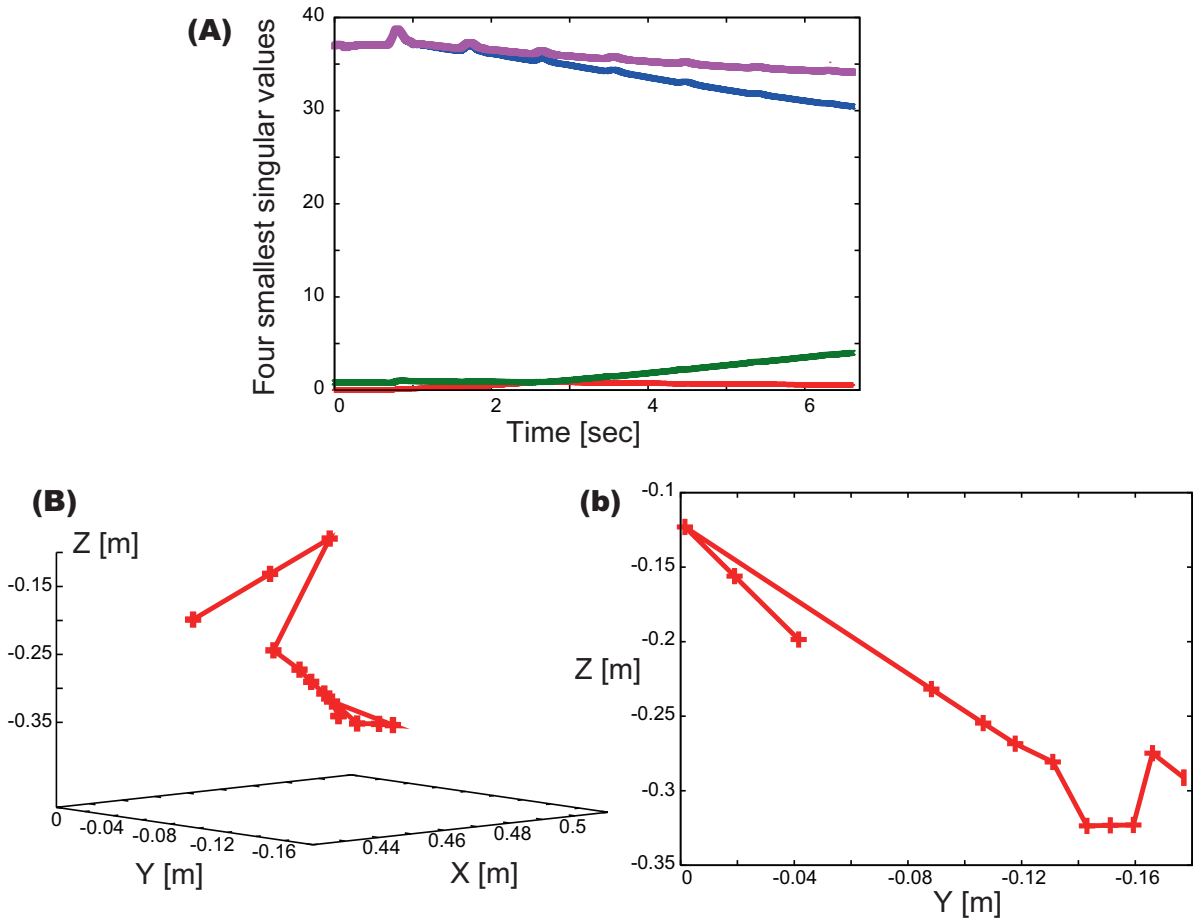

Figure 11: Resultant values of kinesthetic coordination and tactile extension of grasped box (tool).

Meanwhile, it would be possible to extend our method to include motion generation which accelerate its convergence. For instance, orthogonal complement of current variance-covariance matrix seems usable, which could be adopted more easily than a method using mutual information like [18].

Our method of visual-motor coordination sufficiently works only using one tracked point. How can we exploit other points to get the precision better? How can we include tracking by visual feedback without preliminary knowledge? How can it cooperate with camera calibration? These are also problems to enhance our method.

Despite that we ignored accelerated motion in our experiment of inertia identification, it ended up effecting the results. We think a force/torque sensor should be stationary (e.g. placed at the base of arm), if we include accelerated motion in our method. In case that object to be identified is non-rigid, our method can identify inertia parameters approximately. We believe such approximated parameters are usable in our living space.

How can we support the case that kinematics of arm is not given? We are expecting a clue to solve it in ways of animals to obtain body representation. At least, our sensory-motor coordination method in this paper corresponds to simple acquirement of body representation; i.e. we can say it is a computational model of primal body representation. Such a computational modeling will play a crucial role to reveal other cognitive processes of animals.

Identification of coordination potentially involves errors in modeling, estimation, and assumption, which are inevitable in real world. Motion generation should compensate those errors while it subserves identification. Our next goal is cooperation of the identification and the motion generation.

6 Conclusion

Sensory-motor coordination, which is necessary for us to behave consistently in real world, is inconstant because of the change of body when we wear or hold foreign objects such as glasses or tools. This inconstancy makes robots need to have abilities of sustainable coordination. Based on this concept, we proposed a method to achieve sustainable coordination assuming simple and linear coordination. The core idea of our method was to exploit a criterion (the dimension of null-space) which does not directly use error, and to estimate the allowable error of sensory-motor coordination.

We also gave its concrete implementation in visual-motor coordination (hand-eye calibration) and in kinesthetic-motor coordination (inertia identification and touch estimation). Particularly, our implementation of visual-motor coordination endowed a robot marker-free hand-eye calibration under view with irrelevant objects.

Our experiments showed our method is easily convergent and enough accurate. That is, our method is easily applicable to a robotic system. We hope our method is a seed to expand the field where robots autonomously work. We also hope this paper plays a key role to computationally model and understand cognitive functions often seen in animals.

REFERENCES

[1] C. Nabeshima, Y. Kuniyoshi, and M. Lungarella. Adaptive body schema for robotic tool-use. *Advanced Robotics*, 20(10):1105–1126, 2006.

[2] C. Nabeshima, Y. Kuniyoshi, and M. Lungarella. Towards a model for tool-body assimilation and adaptive tool-use. In *Proceedings of The 6th IEEE International Conference on Development and Learning (ICDL-2007)*, 2007.

[3] H. Miyamoto and M. Kawato. A tennis serve and upswing learning robot based on bi-directional theory. *Neural Networks*, 11:1331–1344, 1998.

[4] Y. Yoshikawa. *Subjective Robot Imitation by Finding Invariance.* PhD thesis, Osaka University, 2005.

[5] P. Michel, K. Gold, and B. Scassellati. Motion-based robotic self-recognition. In *Proceedings of 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems*, volume 3, pages 2763–2768, 2004.

[6] C. C. Kemp and A. Edsinger. What can I control?: The development of visual categories for a robot's body and the world that it influences. In *Proceedings of the Fifth International Conference on Development and Learning*, 2006.

[7] P. Fitzpatrick and G. Metta. Grounding vision through experimental manipulation. *Philosophical transactions of the Royal Society.: Mathematical, physical and engineering sciences*, 361(1811):2165–2185, 2003.

[8] L. Natale. *Linking Action to Perception in a Humanoid robot: a Developmental Approach to Grasping.* PhD thesis, the University of Genoa, 2004.

[9] L. Natale, G. Metta, and G. Sandini. A developmental approach to grasping. In *Developmental Robotics: A 2005 AAAI Spring Symposium*, 2005.

[10] A. Stoytchev. Computational model for an extendable robot body schema. Technical Report GIT-CC-03-44, Georgia Institute of Technology, College of Computing Technical Report, 2003.

[11] C. C. Kemp and A. Edsinger. Robot manipulation of human tools: Autonomous detection and control of task relevant features. In *Proceedings of the Fifth International Conference on Development and Learning*, 2006.

[12] C. Schedlinski. A survey of current inertia parameter identification methods. *Mechanical Systems and Signal Processing*, 15(1):189–211, 2001.

[13] D. Ma and J. M. Hollerbach. Identifying mass parameters for gravity compensation and automatic torque sensor calibration. In *Proceedings of IEEE International Conference on Robotics and Automation*, pages 661–666, 1996.

[14] A. Fregolent and A. Sestieri. Identification of rigid body inertia properties from experimental data. *Mechanical Systems and Signal Processing*, 10(6):697–709, 1996.

[15] T. Kodek and M. Munih. An identification of body segment parameters in the upper extremity. In *Proceedings of The IEEE Region 8 EUROCON 2003 Conference: Computer as a Tool*, volume 2, pages 197–201, 2003.

[16] G. Liu, K. Iagnemma, S. Dubowsky, and G. Morel. A base force/torque sensor approach to robot manipulator inertial parameter estimation. In *Proceedings of the 1998 IEEE International Conference on Robotics and Automation*, volume 4, pages 3316–3321, 1998.

[17] M. Gautier and Ph. Poignet. Extended kalman filtering and weighted least squares dynamic identification of robot. *Control Engineering Practice*, 9(12):1361–1372, 2001.

[18] V. A. Sujan and S. Dubowsky. An optimal information method for mobile manipulator dynamic parameter identification. *IEEE/ASME Transactions on Mechatronics*, 8(2):215–225, 2003.

[19] C. Baber. *Cognition and Tool Use: Forms of Engagement in Human and Animal Use of Tools*. Taylor & Francis, 2003.

[20] G. Berlucchi and S. Aglioti. The body in the brain: neural bases of corporeal awareness. *Trends in Neurosciences*, 20(12):560–564, 1997.

[21] J. Paillard. *The Use of Tools by Human and Non-human Primates*, chapter The hand and the tool: the functional architecture of human technical skills. Oxford University Press, New York, 1993.

[22] H. Head and G. Holmes. Sensory disturbances from cerebral lesions. *Brain*, 34:102–245, 1911.

[23] D. G. Smith, J. W. Michael, and J. H. Bowker, editors. *Atlas of Amputations and Limb Deficiencies: Surgical, Prosthetic, and Rehabilitation Principles*. Amer Academy of Orthopaedic, 3rd edition, 2004.

[24] L. M. Koger, J. McIlhattan, and R. Schladetzky. Prosthesis for partially amputated foreleg in a horse. *Journal of the American Veterinary Medical Association*, 156(11):1600–1604, 1970.

[25] T. Breuer, M. Ndoundou-Hockemba, and V. Fishlock. First observation of tool use in wild gorillas. *PLoS Biology*, 3(11):e380, 2005.

[26] A. Iriki, M. Tanaka, and Y. Iwamura. Coding of modified body schema during tool use by macaque postcentral neurons. *Neuroreport*, 7(14):2325–2330, 1996.

[27] S. Obayashi, M. Tanaka, and A. Iriki. Subjective image of invisible hand coded by monkey intraparietal neurons. *Neuroreport*, 11(16):3499–3505, 2000.

[28] A. Iriki, M. Tanaka, S. Obayashi, and Y. Iwamura. Self-images in the video monitor coded by monkey intraparietal neurons. *Neuroscience Research*, 40:163–173, 2001.

[29] S. Yamamoto and S. Kitazawa. Reversal of subjective temporal order due to arm crossing. *Nature Neuroscience*, 4(7):759–765, 2001.

[30] S. Yamamoto and S. Kitazawa. Sensation at the tips of invisible tools. *Nature Neuroscience*, 4(10):979–980, 2001.

[31] S. Yamamoto, S. Moizumi, and S. Kitazawa. Referral of tactile sensation to the tips of L-shaped sticks. *Journal of NeuroPhysiology*, 93(5):2856–2863, 2005.

[32] M. T. Turvey. Dynamic touch. *American Psychologist*, 51:1134–1152, 1996.

[33] P. Schilder. *The Image and Appearance of the Human Body: Studies in the Constructive Energies of the Psyche*. London: Kegan Paul, 1935.

[34] S. Gallagher. Body image and body schema: A conceptual clarification. *Journal of Mind and Behavior*, 7:541–554, 1986.

[35] P. Haggard and D. M. Wolpert. *Higher-Order Motor Disorders*, chapter Disorders of Body Scheme. Oxford University Press, 2005.

[36] M. L. Simmel. The conditions of occurrence of phantom limbs. In *Proceedings of the American Philosophical Society*, volume 102, pages 492–500, 1958.

[37] M. L. Simmel. The absence of phantoms for congenitally missing limbs. *American Journal of Psychology*, 74:467–470, 1961.

[38] V. S. Ramachandran and D. Rogers-Ramachandran. Synaesthesia in phantom limbs induced with mirrors. In *Proceedings of the Royal Society of London*, volume 263, pages 377–386, 1996.

[39] J. R. Lackner. Some proprioceptive influences on the perceptual representation of body shape and orientation. *Brain*, 111:281–297, 1988.

[40] A. Maravita, C. Spence, and J. Driver. Multisensory integration and the body schema: Close to hand and within reach. *Current Biology*, 13:R531–R539, 2003.

[41] N. P. Holmes and C. Spence. The body schema and multisensory representation(s) of peripersonal space. *Cognitive Processing*, 5(2):94–105, 2004.

[42] G. Rizzolatti, L. Fadiga, V. Fogassi, and V. Gallese. The space around us. *Science*, 277(5323):190–191, 1997.

[43] K. Sekiyama, S. Miyauchi, T. Imaruoka, H. Egusa, and T. Tashiro. Body image as a visuomotor transformation device revealed in adaptation to reversed vision. *Nature*, 407:374–377, 2000.

[44] G. G. Gallup Jr. Chimpanzees: Self-recognition. *Science*, 167(3914):86–87, 1970.

[45] J. R. Anderson and J.-J. Roeder. Responses of capuchin monkeys (cebus apella) to different conditions of mirror-image stimulation. *Primates*, 30(4):581–587, 1989.

[46] D. Reiss and L. Marino. Mirror self-recognition in the bottlenose dolphin: A case of cognitive convergence. In *Proceedings of the National Academy of Sciences of the United States of America*, volume 98, pages 5937–5942, 2001.

[47] J. M. Plotnik, F. B. M. de Waal, and D. Reiss. Self-recognition in an asian elephant. In *Proceedings of the National Academy of Sciences of the United States of America*, volume 103, pages 17053–17057, 2006.

[48] P. Rochat. Self-perception and action in infancy. *Experimental Brain Research*, 123(1-2):102–109, 1998.

[49] K. Hiraki. Detecting contingency: A key to understanding development of self and social cognition. *Japanese Psychological Research*, 48(3):204–212, 2006.

[50] J. H. Challis. A procedure for determining rigid body transformation parameters. *Journal of Biomechanics*, 28(5):733–737, 1995.

[51] H. Watanabe and G. Taga. General to specific development of movement patterns and memory for contingency between actions and events in young infants. *Infant Behavior & Development*, 29:402–422, 2006.

[52] S.-J. Blakemore, C. D. Frith, and D. M. Wolpert. Spatio-temporal prediction modulates the perception of self-produced stimuli. *Journal of Cognitive Neuroscience*, 11(5):551–559, 1999.

[53] F. Dornaika and R. Horaud. Simultaneous robot-world and hand-eye calibration. *IEEE Transactions on Robotics and Automation*, 14(4):617–622, 1998.

[54] I. Fassi and G. Legnani. Hand to sensor calibration: A geometrical interpretation of the matrix equation AX=XB. *Journal of Robotic Systems*, 22(9):497–506, 2005.

[55] Neuronics. Katana—intelligent personal robot. `http://www.neuronics.ch/cms_en/web/`.

[56] ATI. F/T Sensor: Mini40. `http://www.ati-ia.com/products/ft/ft_models.aspx?id=Mini40`.

[57] TRACLabs. Pan-tilt Biclops PT. `http://www.traclabs.com/`.

[58] Point Grey. Firefly MV. `http://www.ptgrey.com/`.

[59] Xenomai. Xenomai: Real-time framework for Linux. `http://www.xenomai.org/`.

[60] G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 25(11):120–125, 2000.

[61] Intel. Intel Integrated Performance Primitives (Intel IPP) for Linux. `http://www.intel.com/support/performancetools/libraries/ipp/linux/ia/`.