

RESEARCH ARTICLE

Open Access

Consensus-Making Algorithms for Cognitive Sharing of Object in Multi-Robot Systems

Shodai Tomita^{1*}, Kosuke Sekiyama¹ and Toshio Fukuda^{2,3}

Abstract

Background: Visual recognition in multi-robot systems is afflicted with a peculiar problem that observations made from different viewpoints present different perspectives. Hence, realizing cognitive sharing of the object among robots in an unconstructed environment has become challenging. To cope with these issues, we have proposed the Hierarchical Invariants Perception Model (HIPM) in which multiple representations of the target are dynamically evaluated and selected by the robot. In this paper, we propose consensus-making algorithms to acquire a viewpoint-invariant representation of the geometric relation, which is an unaddressed issue in the HIPM.

Methods: The target is described by a combination of three representations: color, shape, and geometric relation. In terms of geometric relation, we employ relative positions between the target and the salient objects, which we call a geometric-relation-based representation (GRR). A GRR is regarded as viewpoint-invariant when satisfying two conditions: (i) It consists of sharable objects and (ii) the number of target candidates, which is reduced by using the GRR, is equivalent among the robots. Based on this definition, consensus-making algorithms are formulated

Results: Experiments with real-world robots demonstrated that robots perceived the viewpoint-invariant GRR even when objects are occluded or their appearance changes. The experiment also demonstrated that the proposed algorithms were able to reduce the candidates without succumbing to an infinite loop. The success rate of cognitive sharing was about 60%. However, the success rate was 100% when GRR was shared. As long as robots can share the GRR, cognitive sharing may be realized even if the environment is more unstructured and uncertainties increase.

Keywords: Cognitive sharing; Multi-robot; Consensus making

Background

Cognitive sharing of an object is a primary issue in multi-robot task execution, where robots with different perspective are expected to cooperate in our daily unstructured environment. As robots engage in more varied and difficult tasks, they will become a ubiquitous part of our daily life in the future [1]. Researchers generally agree that multi-robot systems of inherently distributed character may behave more robustly and effectively and accomplish cooperative tasks that are not possible for single-robot systems (e.g., carrying a heavy object) [2]. In this paper, we refer to the robot that makes a request to another robot

for a task execution as the *client robot* and to the task execution robot as the *server robot*.

Two methods are needed to realize cognitive sharing: a method for describing suitable representations of the target object so that the server robot uniquely identifies it and a method for sharing the representation with another robot. Cognitive sharing of an object would be achieved when the client robot and the server robot share representations of the object. However, visual recognition in multi-robot systems is afflicted with the peculiar problem that observations made from different viewpoints give different perspectives. Therefore, the server robot may mistake a target for background false positive objects and not all representations of the object can be shared among the robots.

Although cognitive sharing has been realized in conventional work, these methods are based on structured

*Correspondence: tomita@robo.mein.nagoya-u.ac.jp

¹ Department of Micro-Nano Systems Engineering, Nagoya University, Furo-cho, Chikusa-ku, 464-8603 Nagoya, Japan

Full list of author information is available at the end of the article

environments. With the help of an artificial marker such as an RFID tag [3] or QR code [4], or by using a single-colored sphere as a target [5], a target object is identified uniquely and its representation becomes viewpoint-independent. However, in unstructured environments, a predefined marker is difficult to use and the appearance of the target may change according to the viewpoint.

In this paper, to describe suitable representations of a target object, the target is described by a combination of multiple representations by utilizing RGB-D data. RGB-D data give many kinds of representation of an object (e.g., color, shape, label, relative positions, and semantic relations). We deem that, by using a combination of representations, robots can verify whether or not they observe the same object. Although cognitive sharing may be realized by sharing a global position of the target, the global coordinate has to be shared in advance. Even if the global coordinate is shared, localization errors by robots will lead to an ambiguous global position for the target and cause misrecognition.

However, visual representations are highly affected by the robots' viewpoint and environmental condition. Different viewpoints lead to ambiguity of representations as objects are occluded or their appearance changes. The representations described by the robots are often embodiment-specific even if the sensor of the robots is the same e.g., the camera color models may differ slightly [6]. Also, unstructured environments abound with unavoidable disturbances, such as illumination changes, object occlusion, and sensor faults, that would disturb cognitive sharing and even object recognition.

To share representations of the target and to track the target robustly, we have proposed a Hierarchical Invariants Perception Model (HIPM) (Figure 1). This model

is premised on a cluttered, unstructured environment. The purpose of this model is to realize cognitive sharing of an object between two robots and robust object tracking simultaneously. This model has three important processes: describing representations of the target, calculating representational priority, and sharing a viewpoint-invariant decision tree. In calculating representational priority, representations are evaluated with two indicators: *ambiguity*, which estimates the risk of recognition failure for the target in a region of interest (ROI), and *stationarity*, which indicates a steadiness of the ambiguity over time. Robust tracking is realized by selecting representations dynamically based on the ambiguity and stationarity. In sharing a viewpoint-invariant decision tree, the robots try to perceive which representations do not cause a cognitive gap, i.e., which representations are viewpoint-invariant. Cognitive sharing of an object is realized by sharing a suitable combination of viewpoint-invariant representations for identifying the target, which we call a viewpoint-invariant decision tree.

In this paper, we deal with unaddressed issues in the HIPM concept: the definition of a relation-based representation and sharing a viewpoint-invariant decision tree. Although we have proposed the technique of autonomous landmark generation [7], in which some peripheral objects near the target are selected as landmarks in real time to relate the target to the surroundings, problems resulting from differences of viewpoint are not considered and a relation-based representation among the target and multiple landmarks had not been used.

A relation between the target and the surroundings has been investigated for object tracking. Yang et al. [8] formulated a relation between a target and its surroundings under a Markov network with special topology and

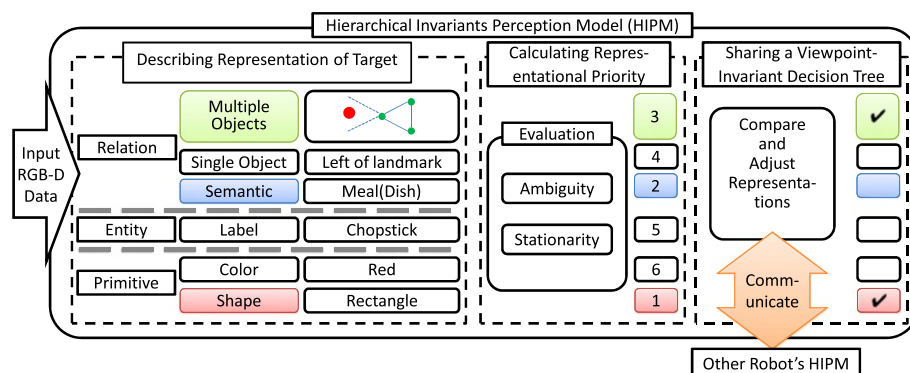


Figure 1 Hierarchical Invariants Perception Model. The client robot describes different classes of representations from RGB-D data (Describing representations of the target). The client robot evaluates the ambiguity and stationarity and selects unique and stable representations (Calculating Representational Priority). The client robot selects a suitable combination of representations for identifying the target to construct a decision tree and sends it to the server robot. By comparing and adjusting representations in the decision tree through communication, the robots reach a consensus (i.e., they determine what representations can be shared). Finally, when the robots share the same decision tree, the target is identified (Sharing-viewpoint-invariant decision tree).

realized robust tracking against occlusion, similar objects, and cluttered backgrounds. However, the relation would change from another viewpoint. On the other hand, relative positions between the target and landmarks are utilized as a relation-based representation of the target. Gohring et al. [5] present a novel approach for multi robot object tracking and self localization using spatial relations of the target with respect to stationary objects. Ahmad et al. [9] represent the problem of cooperative localization and target tracking as a graph, where the edges are relative positions between the target and static landmarks, and solve the problem using sparse optimization methods. However, these methods are based on the use of static predefined landmarks. In unstructured environment, the robots have to select and share undefined landmarks autonomously. A novel approach for sharing the landmarks is proposed in this paper.

As a relation-based representation of the target, we also employ relative positions between the target and the objects near the target, which we call a geometric-relation-based representation (GRR). Use of relative position offers two advantages: 1. The error for the relative positions of objects is comparably small and 2. The information is independent of robot localization and odometry [5]. Therefore, this relation is viewpoint-invariant.

We propose consensus-making algorithms to acquire viewpoint-invariant GRR. Through communication, robots compare and adjust their representations of a target and perceive which representation can be shared and what information is needed gradually. Although relative positions are viewpoint-invariant, not all components of GRR can be shared owing to occlusion or changes in object appearance. Robots can recognize a specific object regardless of viewpoint by means of a three-dimensional model [10]. However, *a priori* knowledge cannot be used in unstructured environments and three-dimensional

models require significant computation in a cluttered environment.

Methods

System overview

The purpose of the proposed approach is to share one target object in an unstructured environment between two robots: the client robot and the server robot. Several assumptions are made as follows.

- Environment: Some objects in the environment are similar to the target. The appearance of objects (e.g., its color and shape) may change as the result of a change in viewpoint. Occlusion of objects may occur. Objects do not move. Drastic illumination changes do not occur.
- Robots: The client robot recognizes a target object. The server robot does not know any target information. The robots are equipped with an RGB-D sensor and can execute translational and rotational movement.

The proposed approach based on the HIPM has two important features:

- Describing representations of the target** The client robot describes representations of the target from an RGB-D image. Color and shape feature are employed as the basic representations to recognize an object; which are referred to as *primitive representation*. Also, GRRs are employed as relation-based representations. The procedure for describing representations of the target is illustrated in Figure 2.
- Sharing the viewpoint-invariant decision tree** The client robot constructs a decision tree, which is a suitable combination of representations for identifying the target from the viewpoint of the client

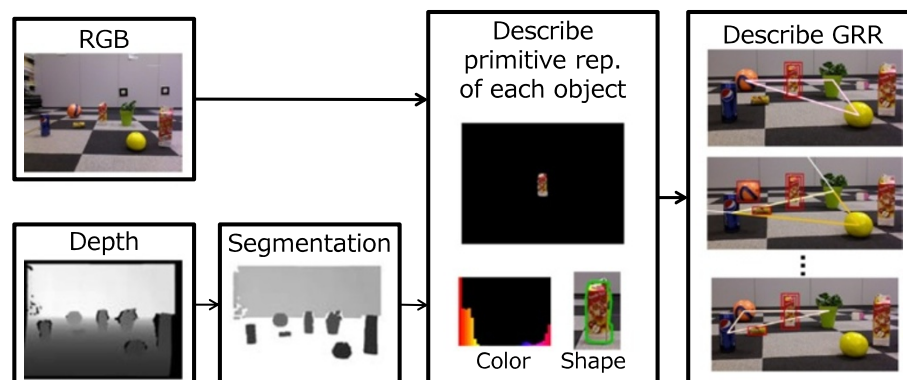


Figure 2 Illustration of describing representations of target. The RGB-D image is employed as input. First, the depth image is divided into objects and other regions. The color and shape representation of each object are then described by using the RGB image and the segmented region. Finally, by using all the objects, geometric-relation-based representations are described.

robot. The server robot receives this decision tree and decides which representations are viewpoint-invariant. If the server robot concludes that the received decision tree includes non-viewpoint-invariant representations, the client robot sends a new decision tree. Through this process, robots share a viewpoint-invariant decision tree gradually.

Calculating representational priority is addressed in [7]. Since we do not assume drastic illumination changes, this component is not discussed in this paper.

Describing representations of the target

Object segmentation

To describe the primitive representation of each object, the input image has to be divided into objects and other areas, and the boundaries have to be contours of objects. In general, a computationally efficient segmentation method is required because the robots are supposed to move around. In this paper, we employ a segmentation method based on the depth information obtained using an RGB-D sensor [11]. RGB-D sensors (Kinect or Xtion) can output sensing data at a frame rate of 30 Hz and with this method one can extract accurate contours by using the depth information.

Primitive representation

We employ color and shape features as the primitive representation. In visual recognition, image features (e.g., color [12], shape [13], and feature points [14-17]) have been used to recognize an object. Although feature points may be salient and therefore suitable for object recognition, they are susceptible to viewpoint changes. However, color and shape features tend to be robust against viewpoint changes.

Because the robots have different viewpoints, the primitive representation should be invariant with respect to scale and illumination changes in the visual recognition. A hue histogram is known to be an invariant representation with respect to scale, illumination direction, and angle changes [18]. In this paper, the histogram similarity function is expressed by histogram intersection [19]. The histogram axis is divided into 32 sections for computational efficiency and recognition accuracy. The similarity of color representation S_c is calculated from

$$S_c(H^a, H^b) = \sum_{i=1}^{32} \min(H^a(i), H^b(i)), \quad (1)$$

where H^a and H^b represent the hue histograms of objects O_a and O_b , respectively, and $H(i)$ represents the value of the i th histogram's bin. Equation (1) is computationally efficient and robust against partial occlusion and resolution changes.

Also, Hu moments [20] of a contour are used as a shape representation in this paper. Hu moments are invariant with respect to scale changes and rotation. Fortunately, because the depth information can be captured under any ambient light conditions, the shape representation is invariant against arbitrary illumination changes. By following the definition in [21], the similarity between two moments can be calculated by

$$S_s(h^a, h^b) = \sum_{i=1}^7 \left| \frac{m^a(i) - m^b(i)}{m^a(i)} \right|, \quad (2)$$

where

$$m(i) = \text{sign}(h(i)) \cdot \log |h(i)|,$$

h^a and h^b represent the Hu moments of objects O_a and O_b respectively, and $h(i)$ represents the value of the i th Hu moments.

Similarity primitive representation from different viewpoints

Before explaining how to describe a GRR, the definition of a similar primitive representation has to be addressed. To identify the components of the GRR and the target, the server robot must determine which representation is similar to the representation sent from the client robot.

In this paper, similar primitive representations are determined by thresholding the similarity value defined by equations (1) and (2). In general, a primitive representation will be highly affected by changes in viewpoint. However, empirically, a similarity of primitive representations between the same object from different viewpoints lies within a certain range.

Assume a color representation H^a and a shape representation h^a of an object O_a are sent from the client robot. When a color representation H^b of object O_b , which is perceived by the server robot, satisfies

$$S_c(H^a, H^b) \geq 0.7, \quad (3)$$

we define O_b as having a similar color representation to H^a . Also, when a shape representation h^b of O_b satisfies

$$S_s(h^a, h^b) \leq 0.3, \quad (4)$$

we define O_b as having a similar shape representation to h^a .

Geometric-relation-based representation

Two processes are needed to form a GRR as follows:

- (i) Select candidate objects of GRR components that have a salient primitive representation. An object that has no similar representation from the viewpoint of the client robot is likely to be identified uniquely. Therefore, such objects are suitable for candidates of GRR components.

- (ii) Describe the relative positions among the target and components of the GRR based on distance information. Three objects of the same kind of primitive representation (e.g., three objects that have the same salient color representations) are selected from the candidates, and a triangle is formed. The reason why the same kind of primitive representation should be selected is that color and shape representations are invariant with respect to different disturbances (e.g., color representation is invariant to partial occlusion and shape representation is invariant to changes in lighting conditions). The reason why a triangle is chosen is that it is the minimum unit needed to divide a flat space into a closed area and other geometric shapes can be represented by a combination of triangles.

A GRR divides the recognition area into 7 areas $A_a (a \in \{1, 2, \dots, 7\})$. We denote the decomposed area where the target belongs by A_t , which is uniquely represented by the triple set of + and - signs as shown in Figure 3. If we assume l candidates for the GRR components, the number of constructed GRRs, m , is lC_3 .

Sharing-viewpoint-invariant decision tree

Construction of the decision tree

We use a binary decision tree that consists of combined representations to identify the target because it is rather rare when the target object can be uniquely identified by means of a single representation. Such a limited case is the following:

- (i) When a primitive representation is employed, no similar representation is found close to the target object.
- (ii) When a GRR is employed, no other objects belong to A_t of a GRR.

Conventional methods (e.g., C4.5 and CART) tend to construct a large tree when learning data are not adequate. Cognitive sharing of an object requires classifying two types of objects: the target object and other objects. Because the target object is unique in the environment, target object class data are not adequate. In this case, taking C4.5 as an example, some information lead to the same value and C4.5 cannot select an effective node.

To construct a decision tree, we employ the branch and bound algorithm. The branch and bound algorithm has the advantage of constructing a decision tree that can minimize the required number of nodes. Redundant information may increase searching time because not all representation can be shared owing to appearance changes and occlusion. The solution to the problem resulting from viewpoint changes is discussed in the next section on consensus-making algorithms.

The branch and bound algorithm is given as follows: Assume a set of n objects $O = \{O_1, O_2, \dots, O_n\}$, which is perceived by the client robot. An object $O_i (\in O)$ is described by a color representation, i.e., hue histogram H^i , a shape representation, i.e., Hu moments h^i , and m GRRs $g_j^i (j \in M = \{1, 2, \dots, m\})$, where i denotes the representation of the i th object and g_j^i denotes the j th GRR of O_i . From the viewpoint of the client robot, the number of candidate target objects $O_t (\in O)$ can be reduced by using the target's representations H^t, h^t, g_j^t according to

$$R_j = \{O_i \in O | O_i \in A_t^j (j \in M)\}, \quad (5)$$

$$R_{m+1} = \{O_i \in O | S_c(H^t, H^i) \leq 0.7\}, \quad (6)$$

$$R_{m+2} = \{O_i \in O | S_s(h^t, h^i) \geq 0.3\}. \quad (7)$$

Here R_j represents a set of candidates that belong to A_t^j , A_t^j denotes the decomposed area of g_j^t , and R_{m+1} and R_{m+2} represent a set of candidates reduced by using similarity criteria h^t and H^t , respectively. The thresholds of

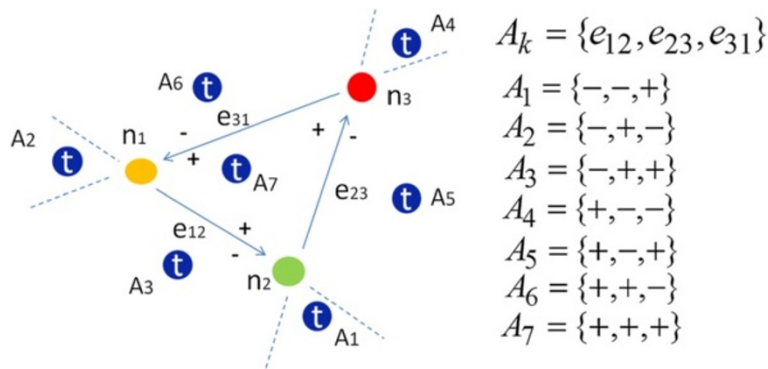


Figure 3 Representation of target position. The components of a GRR are denoted by n_1, n_2, n_3 and connected in the counterclockwise direction. A link vector (e_{12}, e_{23}, e_{31}), which connects to the components, will decompose the recognition area into two domains. The left side of the link vector where each vector is linked counterclockwise is denoted with a positive sign (+) and the right side of each vector is denoted with a negative sign (-).

equations (6) and (7) are defined based on equations (3) and (4) as follows:

(i) Objective Function and Constraint Condition

Let us denote a collection of the target's candidate set by $R = \{R_1, R_2, \dots, R_m, R_{m+1}, R_{m+2}\}$. The goal is to find a combination of the target's representations that can reduce the number of candidate objects of the target to 1 such that the number of tree nodes is minimized. The objective function and constraint condition are defined as follows:

$$\begin{aligned} &\text{minimize } |X|, \text{ where } X = \cap_{R_k \in R'} R_k, R' \subseteq R, & (8) \\ &\text{subject to } |X| \geq 1, |R'| \leq 3, & (9) \end{aligned}$$

where $|\cdot|$ represents the cardinality of a set. To reduce redundant information, the number of nodes $|R'|$ is limited to 3. Also, to share a ROI, a GRR is always employed as a node of the decision tree if there are sharable GRRs. The reason why a GRR is employed for sharing ROI is discussed in the next section on consensus-making algorithms. An illustration of constructing a decision tree is shown in Figure 4.

(ii) Branching

Subproblem P_i (breadth first search)

Minimize $|X|$ subject to $|R'| = i$ ($i \in \{1, 2, 3\}$).

(iii) Bounding

Prune if $|X| \leq z$.

z is the minimum upper bound seen among subproblems examined so far.

Finish if $|X| = 1$.

Viewpoint-invariant representation

Before proceeding to the next section on consensus-making algorithms we need to define a viewpoint-invariant representation. A representation of the target object is regarded as viewpoint-invariant when it satisfies two conditions from the viewpoint of the server robot:

- (i) The server robot finds a similar representation in a ROI.
- (ii) The number of objects that satisfy the similarity criteria of the representation is equivalent from the views of both the client and server robots.

When the second condition is not satisfied, the server robot may mistake the target because the server robot cannot perceive which object is occluded or changes in its appearance.

Next, we define viewpoint-invariant color, shape, and geometric-relation-based representation. Assume a set of n objects $O' = \{O'_1, O'_2, \dots, O'_n\}$, which is perceived by the server robot. For the target, we have color representation H , shape representation h , and GRR g being sent from the client robot, and the number of representations similar to H and h from the viewpoint of the client robot is α and β , respectively. The number of candidates that belong to A_t of g from the viewpoint of the client robot is γ .

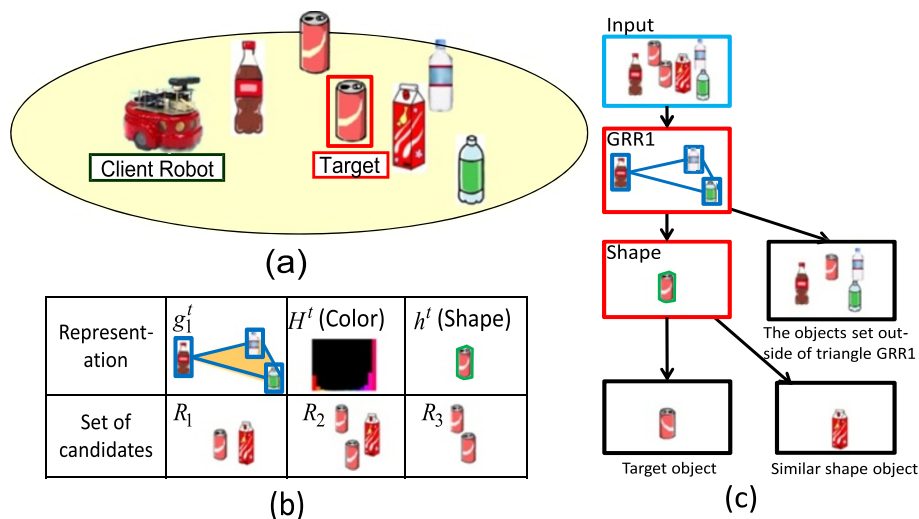


Figure 4 Illustration of constructing a decision tree. (a) Environment. The object enclosed within the red rectangle is the target. (b) Representations of the target and sets of candidates. A GRR (g_1^t), color (H^t), and shape (h^t) are described as representations of the target (top row). R_1 is a set of candidates that belong to the same decomposed area A_1^t of g_1^t . R_2 and R_3 are sets of candidates reduced by using the similarity criteria of color and shape, respectively (bottom row). (c) Example of a decision tree. Assume g_1^t and h^t are employed as nodes of the decision tree. The blue rectangle represents input to the decision tree; red rectangles represent nodes of the decision tree. The number of inputted candidates is reduced to $R_1 = 2$ by using GRR1. Then, the number of candidates (R_1) is reduced to $R_1 \cap R_2 = 1$ by using the shape representation.

H is viewpoint-invariant when the following equation is satisfied:

$$|C| = \alpha, \text{ where } C = \{O'_i \in O' | S_c(H, H^i) \leq 0.7\}. \quad (10)$$

h is viewpoint-invariant when the following equation is satisfied:

$$|S| = \beta, \text{ where } S = \{O'_i \in O' | S_s(h, h^i) \geq 0.3\}. \quad (11)$$

g is viewpoint-invariant when the following equation is satisfied:

$$|G| = \gamma, \text{ where } G = \{O'_i \in O' | O'_i \in A_t\} \quad (12)$$

and the primitive representation of the component is viewpoint-invariant.

Consensus-making algorithms

As mentioned in the preceding section for construction of a decision tree, a decision tree sent from the client robot can include a non-viewpoint-invariant representation. In this section, we discuss how two robots perceive viewpoint-invariant GRRs and share a viewpoint-invariant decision tree through communication. We define a decision tree consisting of viewpoint-invariant representations as a viewpoint-invariant decision tree. Figure 5 shows a state transition diagram of the consensus-making algorithms.

The consensus-making algorithms are composed of four functional parts: sharing ROI, searching candidates, perceiving invariants, and adjusting the decision tree.

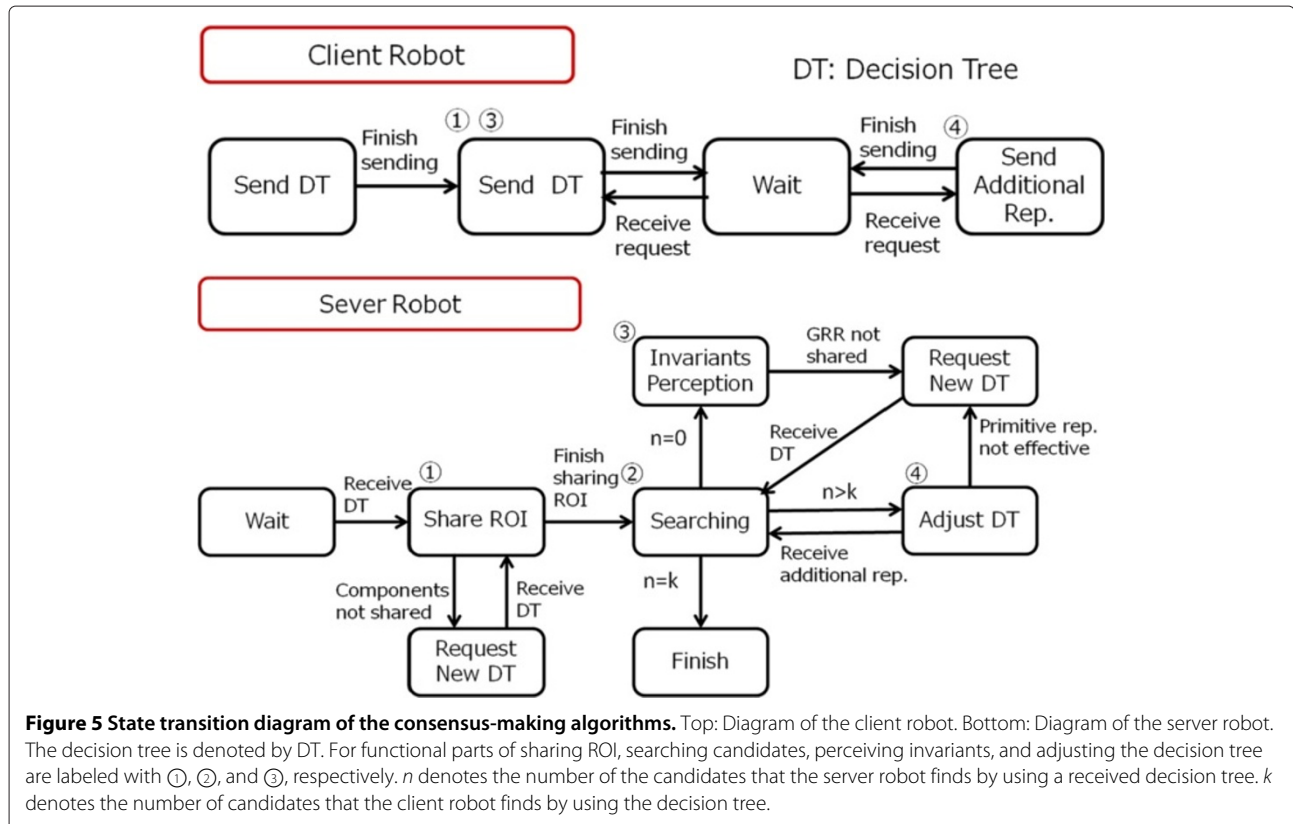
Sharing ROI

The server robot starts sharing ROI while rotating when receiving the decision tree (Figure 5 ①). A ROI has to be shared to verify whether or not a representation is viewpoint-invariant. Without sharing a ROI, robots cannot determine the cause of the misrecognition by searching another region or a non-viewpoint-invariant representation.

In this paper, if the server robot finds one or more components of the GRR, a ROI is considered to be shared roughly. If the server robot does not find components during one rotation, the server robot requests the client robot to send another decision tree. This process lasts until a ROI is shared.

Searching candidates

After sharing the ROI, the server robot searches around the ROI for the target. The state of the server robot transitions according to the number of target candidates found by using the received decision tree. Let us denote the number of the candidates the server robot finds using a received decision tree by n and the number of candidates the client robot finds using the decision tree by k . When



$n = 0$, the state transitions to perceiving invariants. When $n > k$, the state transitions to adjusting the decision tree. When $n = k$, the state transitions to finishing cognitive sharing (Figure 5 ②).

Perceiving invariants

The server robot regards a GRR as non-viewpoint-invariant under three conditions: (i) when the GRR includes the components that cannot be found in the ROI, (ii) when the GRR includes components whose candidates are found more than once in the ROI, and (iii) when the number of candidates existing inside of A_t of the GRR, where the target object is found, is not equivalent between the client and server robots. Because the client robot selects the objects that have a salient primitive representation as GRR components, such components should be identified uniquely. However, either occlusion or appearance changes may occur as a result of difference of the viewpoints when the components are not found in the ROI. Also, objects near the components may become similar to the component owing to appearance change when candidates of the component are found more than once. The number of candidates differs when objects out of the client robot's view exist. In this situation, a decision tree has difficulty in classifying the objects correctly.

If the received decision tree includes non-viewpoint-invariant GRRs, the server robot request the client robot to send another decision tree that does not include unidentified components. The robots iterate this process until they share a viewpoint-invariant GRR (Figure 5 ③).

Adjusting the decision tree

Even though the robots successfully share a viewpoint-invariant GRR, the server robot sometimes may fail to identify the target when the primitive representation is not viewpoint-invariant. For example, this would occur when an object out of the client robot's view exists in A_t of the GRR or when the objects near the target in A_t change their appearance.

In these situations, the server robot has to determine what information is needed autonomously. The server robot calculates the similarity of primitive representations among the candidates as reduced by using the received decision tree. If the primitive representations of the candidates are not similar, the server robot requests the primitive representation. Then, the server robot adds the primitive representation to the decision tree and tries to identify the target. The server robot also requests another decision tree if the number of candidates cannot be reduced by using only the primitive representation. The server robot will finish searching when the number of candidates is reduced to one or when the number of candidates obtained by using the decision tree is the same between client and server robots (Figure 5 ④).

Results

Experimental settings

Experiments are conducted to validate the algorithm for use in real-world robots with basic kinds of objects from different viewpoints of the server robot as shown in Figure 6. The first two sections demonstrate that the proposed approach has robustness against the following problems:

- The server robot mistakes the target for a similar object.
- Not all representations can be shared because of object occlusion and appearance changes resulting from differences of viewpoint.

In the third section, termination of the algorithm is discussed. Finally, the algorithm is evaluated quantitatively.

Each robot (client robot: Pioneer 3-DX; server robot: Amigo Bot) is equipped with a PC (Intel Core i5 2.4 GHz), an RGB-D camera (Kinect), and a communication module (OKI UDv4). The client robot recognizes the target and the server robot does not know *a priori* target information. Neither robot knows the current position of the other robot. Also, we assume that dramatic illumination changes and movement of objects will not occur in the time frame of communication.

Perception of viewpoint-invariant GRR

This experiment demonstrates that the client robot can select a suitable combination of representations for identifying the target and that the server robot can perceive the viewpoint-invariant GRR. The experimental condition is shown in Figure 7. The relative positions of the robots corresponds to Figure 6(a).

Snapshots of the experiment are shown in Figure 8. A decision tree is shown in the upper left of each image. The robot's action and communication condition are shown above each image. The red bounding box in the image represents the target and the blue bounding boxes represent identified components of the GRR. White crosses represent the objects that have viewpoint-invariant representations.

Sharing ROI

The client robot selected a GRR as a node of the decision tree to share the ROI. By $t = 4$ [s], the server robot succeeded in receiving a decision tree and started searching the components. By $t = 13$ [s], the server robot found two out of three components and succeeded in sharing the ROI.

Perceiving invariants: appearance change

The server robot searched the last component of the GRR but could not find it. Therefore, by $t = 18$ [s], the server

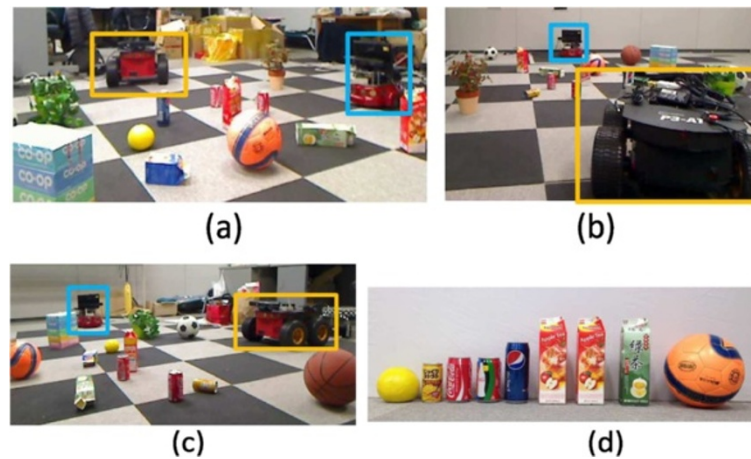


Figure 6 Examined viewpoints of the server robot and experimental objects. (a), (b), (c) Examined positional relation between the client and server robots. The client and server robots are enclosed with yellow and blue rectangles, respectively. The server robot is placed so that the client–server–target angle is about 90, 180, and 270 degrees, respectively, in the clockwise direction from the client robot. (d) Nine experimental objects as the target.

robot regarded the last component as non-viewpoint-invariant and requested the client robot to send another decision tree that did not include the unidentified component. Actually, the color for the last component from the viewpoint of the client robot was different from that of the server robot.

Perceiving invariants: occlusion

The client robot selected a GRR as a decision tree because the target could not be determined by using only the primitive representation owing to presence of similar object. By $t = 34$ [s], the server robot did not find one of the components because of occlusion. The server robot regarded it as non-viewpoint-invariant and requested the client robot to send another decision tree.

Perceiving invariants: similar representation

By $t = 45$ [s], the server robot requested another decision tree because one of the components had a similar

representation in the environment owing to appearance change and another component was not found.

Cognitive sharing

By $t = 49$ [s], the server robot reduced the number of candidates to one because the received decision tree consisted of only a viewpoint-invariant GRR, and it succeeded in cognitive sharing.

Adjusting the decision tree

This experiment, which is conducted in the same environment as depicted in Figure 7, demonstrates the effect of adjusting the decision tree to share the viewpoint-invariant decision tree. The result is shown in Figure 9. By $t = 95$ [s], the robots shared the decision tree that consists of the viewpoint-invariant representations. However, a similar color object appeared in the decomposed area because of an appearance change; therefore, server robot could not distinguish the candidate objects. The



Figure 7 Experimental condition. The client robot recognizes the target in the middle of the figure. An object similar to the target exists in the middle right of the figure (labeled as “Dummy”).

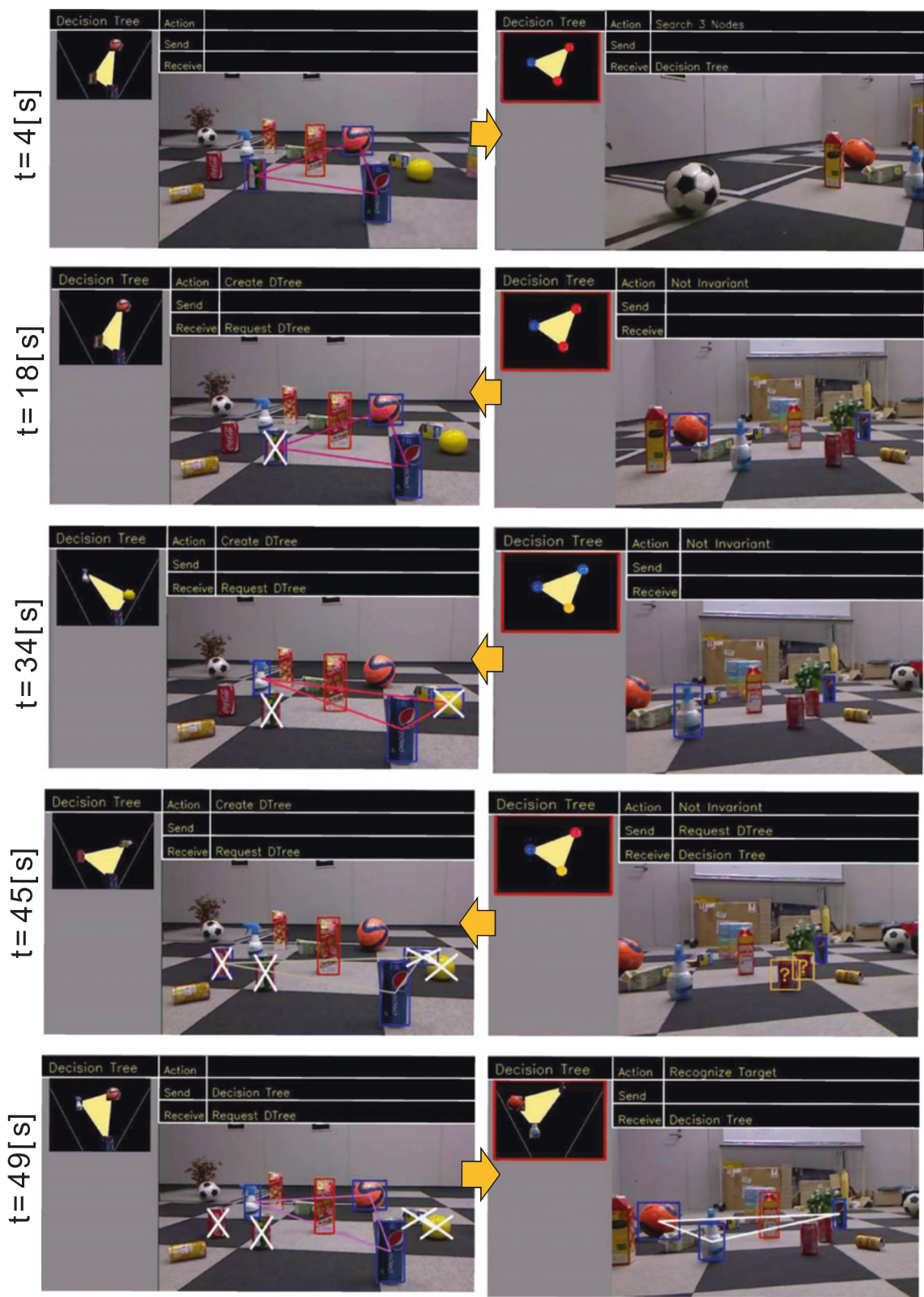


Figure 8 Snapshots of perceiving viewpoint-invariant GRR. Left column: View of the client robot. Right column: View of the server robot. Yellow arrows represent communication flow.

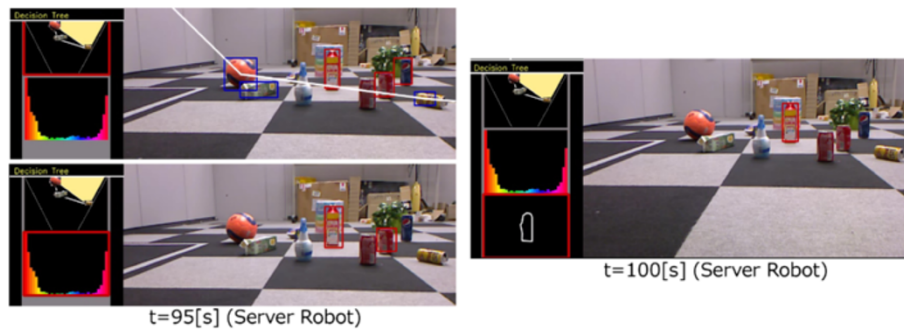


Figure 9 Snapshots of adjusting the decision tree. Top of left column: Candidates are reduced to three with the GRR. Bottom of left column: The candidates are reduced to two by using color representation. Right column: The server robot identifies the target by adding shape representation to the decision tree.

server robot determined the necessary representation autonomously and requested its shape representation. By $t = 100$ [s], the server robot succeeded in cognitive sharing by adding the additional representation.

Termination of consensus-making algorithms

This experiment demonstrates that the proposed algorithms reduce the number of candidates as many as possible by using sharable representations. In this experiment, the relative position of the robots corresponds to Figure 6(b) and the target corresponds to the object labeled “Dummy” in Figure 7.

Snapshots of the experiment are shown in Figure 10. By $t = 12$ [s], the server robot found one component of the GRR and succeeded in sharing a ROI. However, the other components were not found owing to an appearance change. Hence, the server robot regarded the GRR as non-viewpoint-invariant and requested another decision

tree from the client robot. By $t = 14$ [s], the client robot sent a new decision tree. Because two components were regarded as non-viewpoint-invariant, there was no sharable GRR. Therefore, the client robot selected the primitive representation as a decision tree. This decision tree reduced the number of candidates to two in the view of the client robot. By $t = 19$ [s], the server robot terminated cognitive sharing because the number of found candidates is two. Although the target was not identified, the proposed approach was able to reduce the number of candidates without succumbing to an infinite loop.

Quantitative evaluation

The total number of experiments conducted was 27. Experimental results are shown in Table 1. We deem that cognitive sharing is successful when the server robot identifies the target uniquely. The top two rows indicate environmental conditions, whether or not similar objects,

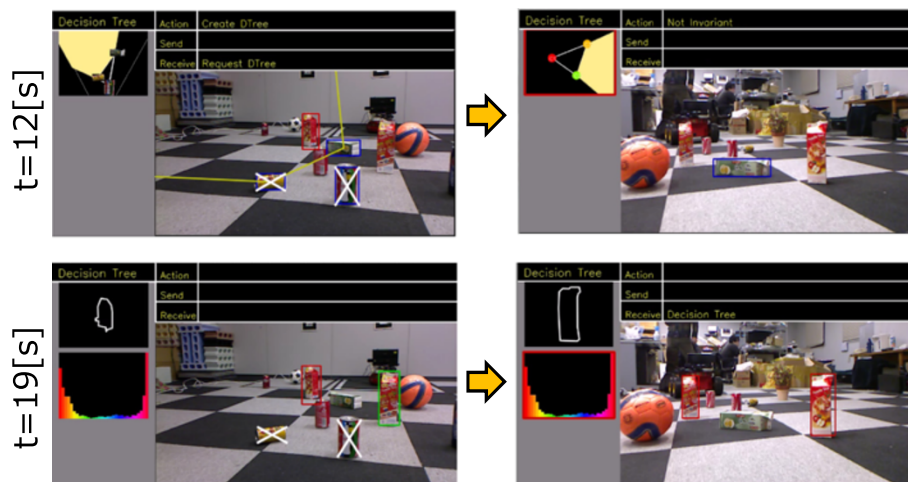


Figure 10 Snapshots of termination of consensus-making algorithms. Left column: View of the client robot. Right column: View of the the server robot. Yellow arrows represent communication flow. The object in the green rectangle represents the candidate object classified by using the decision tree from the viewpoint of the client robot.

Table 1 Success rate of cognitive sharing

Condition	SO ^a exist		SO do not exist	
	Invariant ^b	Variant	Invariant	Variant
Share GRR	4/4 ^c	1/1	4/4	0/0
Do not share GRR	2/8	0/1	5/6	0/3

^aSO: similar objects.

^bThe target's representation is viewpoint invariant.

^cThe number of successes divided by the number of situations.

and whether or not primitive representations of the target are viewpoint-invariant. The left column indicates whether or not the robots share a GRR. When robots can share the GRR, they achieve cognitive sharing regardless of environmental conditions. As long as both robots can share the GRR, cognitive sharing may be realized even if the environment is more unstructured and uncertainties increase.

Discussion

The experiments in Figure 8 and 10 demonstrated that the consensus-making algorithms allow the server robot to perceive non-viewpoint-invariant GRRs under three conditions: (i) The appearance of GRR components changed, (ii) a part of the components were occluded, and (iii) peripheral objects were similar to a part of the components in the view of the server robot. There may be some difficult cases where the robots fail to regard GRRs as viewpoint-invariant correctly. For example, a similar object to one of the components is occluded in the view of the client robot while the component is occluded in the view of the server robot, and the number of candidates existing inside of A_t of the identified GRR is equivalent between the robots. However, such case is rare in the real environment.

The experimental results in Table 1 show the effect of combined representations for identifying the target. The success rate of cognitive sharing was 100% in the case that the robots can share the GRR, while the success rate is 38% in the case that the robots cannot share the GRR, i.e., the decision tree only includes the primitive representations.

Conclusions

We proposed a cognitive sharing algorithm based on visual information. A decision tree including a geometric-relation-based representation allows multiple robots to share precise ROI and avoids confusing the target for a similar object. The consensus-making algorithms serve to acquire viewpoint-invariant GRRs under conditions in which occlusion and object appearance changes occur. By adjusting the decision tree, the server robot, i.e., the robot requested to execute the task, identifies the target even if some primitive representations are not viewpoint-invariant. The consensus-making algorithms can reduce

the number of candidates without succumbing to an infinite loop.

For future work, an important issue will be to integrate one of the features in HIPM, i.e., calculating representational priority with the consensus-making algorithms, thereby enabling robots to realize tracking and cognitive sharing of an object simultaneously in spite of disturbances.

Abbreviations

HIPM: Hierarchical invariants perception model; GRR:

Geometric-relation-based representation; ROI: Region of interest.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

ST formulated concepts and ideas and drafted the manuscript. KS and TF contributed concepts and edited and revised the manuscript. All authors read and approved the final manuscript.

Author details

¹Department of Micro-Nano Systems Engineering, Nagoya University,

Furo-cho, Chikusa-ku, 464-8603 Nagoya, Japan. ²Institute for Advanced

Research, Nagoya University, Furo-cho, Chikusa-ku, 464-8603 Nagoya, Japan.

³Faculty of Science and Engineering, Meijo University, 1-501, Shiogamaguchi, Tenpaku-ku, 464-8502 Nagoya, Japan.

Received: 15 January 2014 Accepted: 12 June 2014

Published online: 04 September 2014

References

- Gates B (2007) A robot in every home. *Sci Am* 296:58–65
- Arai T, Pagello E, Parker LE (2002) Editorial: advances in multi-robot systems. *IEEE Trans Rob Autom* 18(5):655–661
- Tan KG, Wasif AR, Tan CP (2008) Objects tracking utilizing square grid RFID reader antenna network. *J Electromagn Waves Appl* 22:27–38
- Xue Y, Tian G, Li R, Jiang H (2010) A new object search and recognition method based on artificial object mark in complex indoor environment. In: *World congress on intelligent control and automation*. IEEE, pp 6648–6653
- Gohring D, Homann J (2006) Multi robot object tracking and self localization using visual percept relations In: *Proceedings of IEEE/RSJ international conference of intelligent robots and systems*. IEEE, pp 31–36
- Kira Z (2010) Inter-robot transfer learning for perceptual classification. In: *Proceedings of the 9th international conference on autonomous agents and multiagent systems*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, pp 13–20
- Umeda T, Sekiyama K, Fukuda T (2012) Vision-based object tracking by multi-robots. *J Rob Mechatron* 24(3):531–539
- Yang M, Wu Y, Hua G (2009) Context-aware visual tracking. *IEEE Trans. Pattern Anal Mach Intell* 31(7):1195–1209
- Ahmad A, Tipaldi GD, Lima P, Burgard W (2013) Cooperative robot localization and target tracking based on least squares minimization. In: *IEEE International Conference on Robotics and Automation 2013 (ICRA)*, IEEE, pp 5696–5701
- Waibel M, Beetz M, Civera J, d'Andrea R, Elfing J, Galvez-Lopez D, Haussermann K, Janssen R, Montiel JMM, Perzylo A, Schiesle B, Tenorth M, Zweigle O, van de Molengraft R (2011) A World Wide Web for Robots. *IEEE Rob Autom Mag* 18(2):69–82
- Filliat D, Battesti E, Bazeille S, Duceux G, Geppert A, Harrath L, Jebbari I, Pereira R, Tapus A, Meyer C, Ieng S-H, Benosman R, Cizeron E, MAManna J-C, Pothier B (2012) RGBD object recognition and visual texture classification for indoor semantic mapping. In: *IEEE international conference on technologies for practical robot applications (TePRA)*. IEEE, pp 127–132
- Comaniciu D, Ramesh V, Meer P (2003) Kernel-based object tracking. *IEEE Trans Pattern Anal Mach Intell* 25(5):564–577

13. Isard M, Blake A (1996) Contour tracking by stochastic propagation of conditional density. In: Buxton B, Cipolla R (eds) *Computer Vision -ECCV '96*. Lecture Notes in Computer Science, vol. 1064. Springer, Berlin Heidelberg, pp 343–356
14. Shi J, Tomasi C (1994) Good features to track. In: *Proceedings of IEEE conference of computer vision and pattern recognition*. IEEE, pp 593–600
15. Lowe DG (1999) Object recognition from local scale-invariant features. *Proc Seventh IEEE Int Conf Comput Vision* 2:1150–1157
16. Mikolajczyk K, Schmid C (2001) Indexing based on scale invariant interest points. *Proc Eighth IEEE Int Conf Comput Vision* 1:525–531
17. Fitzgibbon A, Zisserman A (2002) On affine invariant clustering and automatic cast listing in movies. *Proc. Seventh Eur Conf Comput Vision* 3:304–320
18. Gevers T, Smeulders AWM (1999) Color based object recognition. *Pattern Recognit* 32(March):453–465
19. Swain MJ, Ballard DH (1991) Color Indexing. *Int J Comput Vision* 7(1):11–32
20. Hu M (1962) Visual pattern recognition by moment invariants. *IRE Trans Inf Theory* 8(2):179–187
21. Gary B, Kaehler A (2008) *Learning OpenCV: Computer vision with the OpenCV library*. O'Reilly Media, Sebastopol, CA

doi:10.1186/s40648-014-0007-6

Cite this article as: Tomita et al.: Consensus-Making Algorithms for Cognitive Sharing of Object in Multi-Robot Systems. *ROBOMECH Journal* 2014 **1**:7.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com