Toward Video-Guided Robot Behaviors

Alexander Stoytchev
Department of Electrical and Computer Engineering
Iowa State University
Ames, IA 50011, U.S.A.
alexs@iastate.edu

Abstract

This paper shows how a robot can detect its self-image in a TV monitor and use the real-time video to guide its own actions in order to reach an object that can only be observed through the TV image. The robot exploits the temporal contingency between its motor commands and the observed self-movements in the TV monitor to extend its body schema representation. The extended body schema allows the robot to guide its arm movements through the TV image as if it were observing its own arm directly.

1. Introduction

Humans are capable of performing many behaviors in which they receive visual feedback about their own actions only through indirect means, e.g., through a mirror reflection or a real-time video image. Some examples include: driving a car in reverse using the rear view mirrors, playing a video game using a joystick to control a virtual character, and using a computer mouse to position the mouse pointer on a computer monitor. As robots continue to spread to human-inhabited environments the ability to perform video-guided behaviors will become increasingly more important. This problem, however, has not been well addressed by the robotics community.

Some primates can also perform video-guided behaviors. For example, consider the task shown in Figure 1 which is described by Iriki et al. (2001). The hands of the monkey and the incentive object are placed under an opaque panel such that they cannot be observed directly. In order to reach and grasp the incentive object the monkey must use the real-time video feedback of its own movements captured by a camera and projected on a TV monitor.

To solve this problem the monkey must solve at least three sub-problems. First, it must realize that the TV monitor displays a real-time video of its own hands and not, say, a recording of the movements of another monkey. Second, the monkey must figure out the similarity transformation between the position of its real arm (estimated from proprioceptive
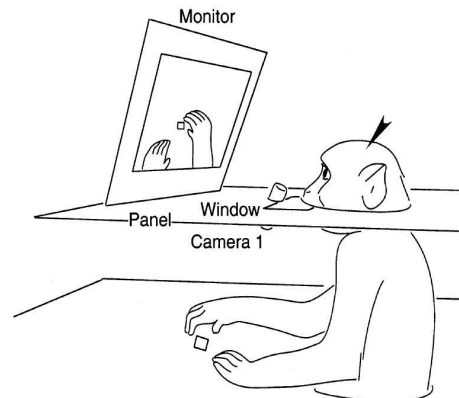

Figure 1: The figure shows the experimental setup that was used by Iriki et al. (2001). The setup consists of a TV monitor that displays real-time images captured by the camera. An opaque panel prevents the monkey from observing the movements of its hands directly. Instead, it must use the TV image to guide its reaching behaviors in order to grasp the food item (a piece of apple). During the initial training phase a transparent window located close to the eye level of the monkey was left open so that it can observe the movements of its hands directly as well as in the TV monitor.

information as it cannot be seen directly) and the image of the arm in the TV monitor. Finally, the monkey must use the video image to guide its hand toward the incentive object (usually a piece of apple).

This paper describes a computational framework that was used successfully by a robot to solve the three sub-problems described above and thus to achieve video-guided behaviors. The robot solves the first sub-problem by detecting the temporal contingency between its own motor commands and the observed self movements in the video image. As the video image is projected in real time the visual self-movements detected in it occur after the expected proprioceptive-to-visual efferent-afferent delay of the robot (a.k.a. the perfect contingency (Watson, 1994)). The second sub-problem is solved by estimating the similarity transformation (translation, rotation, and scale) between two sets of points. The first set consists of the positions of specific locations on the robot's body (e.g., the wrist) which are estimated from proprioceptive information as they can-

not be observed directly. The second set consists of the observed positions of the same body locations in the video. Once the similarity transformation is calculated the robot can perform video-guided grasping by modifying the representation which encodes its own body, i.e., by extending its *body schema*.

As far as we know, this paper describes the first example of video-guided behaviors in the robotics and AI literature.

2. Related Work

2.1 Experiments with Animals

Menzel et al. (1985) reported for the first time the abilities of chimpanzees to perform video-guided reaching behaviors. The chimpanzees in their study were also capable of detecting which of two TV monitors shows their self image and which shows a recording from a previous trial. They succeeded even when the TV image was rotated by 180°.

Experiments in which the visual feedback comes from a mirror instead of a video image have also been performed. Itakura (1987) reported that Japanese monkeys can reach for targets that can only be observed in the mirror image. Epstein et al. (1981) trained pigeons to peck a spot on their body that could be seen only in a mirror. After Gallup (1970) discovered that chimpanzees can self-recognize in the mirror there has been a flood of studies that have used mirrors in primate experiments. These studies are far too numerous to be summarized here. See (Barth et al., 2004) for a comprehensive summary.

More recently, Iriki et al. (2001) have performed reaching experiments with Japanese monkeys (see Figure 1) while simultaneously recording the firing patterns of neurons located in the intraparietal sulcus that are believed to encode the body schema of these monkeys. Their results show that these neurons, which fire when the monkey observes its hand directly, can be trained to fire when the hand is observed in the TV image as well. Furthermore, they showed that the visual receptive fields of these neurons shift, expand, and contract depending on the position and magnification of the TV image. An important condition for learning these skills is that the "monkey's hand-movement had to be displayed on the video monitor without any time delay [...] the coincidence of the movement of the real hand and the video-image of the hand seemed to be essential" (Iriki et al., 2001, p. 166).

2.2 Related Work on Body Schemas

The notion of *body schema* was first suggested by Head and Holmes (1911) who studied the perceptual mechanisms that humans use to perceive their own bodies. They define the *body schema* as a postural model of the body and a model of the surface of the body (Head and Holmes, 1911). It is a perceptual model of the body formed by combining information from proprioceptive, somatosensory, and visual sensors. They suggested that the brain uses such a model in order to register the location of sensations on the body and to control body movements.

Indirect evidence for the existence of a body schema comes from numerous clinical patients who experience disorders in perceiving parts of their bodies - often lacking sensations or feeling sensations in the wrong place (Frederiks, 1969; Head and Holmes, 1911). One such phenomenon called *phantom limb* is often reported by amputees who feel sensations and even pain as if it were coming from their amputated limb (Ramachandran and Blakeslee, 1998).

Direct evidence for the existence of a *body schema* is provided by recent studies which have used brain imaging techniques to identify the specialized regions of the primate (and human) brain responsible for encoding it (Berlucchi and Aglioti, 1997; Graziano et al., 2000; Iriki et al., 1996, 2001). Other studies have shown that body movements are encoded in terms of the body schema (Berthoz, 2000; Graziano et al., 2002). This seems to be the case even for reflex behaviors (Berthoz, 2000).

Perhaps the most interesting property of the body schema is that it is not static but can be modified and extended dynamically in very short periods of time. Such extensions can be triggered by the use of noncorporeal objects such as clothes, ornaments, and tools (Iriki et al., 1996; Maravita and Iriki, 2004). Thus, the body schema is not tied to anatomical boundaries. Instead, the actual boundaries depend on the intended use of the body parts and the external objects attached to the body. For example, when people drive a car they get the feeling that the boundary of the car is part of their own body (Schultz, 2001) but when they get out of the car their body schema goes back to normal.

2.3 Related Work in Robotics and AI

The robotics work on body schemas is still in its infancy as only a few researchers have attempted to tackle this subject.

Yoshikawa et al. (2002) formulated a fully connected neural network model that identified the common firing patterns between tactile, visual, and proprioceptive sensors. Their model was capable of making the right associations between sensory modalities but lacked extensibility properties which are necessary for video-guided behaviors.

Nabeshima et al. (2005, 2006) describe a method for changing the properties of the robot's controller (which is based on inverse kinematics) to accommodate attached tools. The extension is triggered by the coincidence in the firing of tactile sensors (at the hand which is grasping the tool) and the diminishing visual distance between the free end of the tool and some visual landmark. Their extension method requires direct physical contact and thus is also not

suitable for achieving video-guided behaviors.

This paper builds upon our previous work (Stoytchev, 2003) which introduced a computational model for an extendable robot body schema (RBS). The model uses visual and proprioceptive information to build a representation of the robot's body. The visual components of this representation are allowed to extend beyond the boundaries of the robot's body. The proprioceptive representation, however, remains fixed at all times and allows the robot to perform visually-guided movements even when its body representation is extended.

In our previous study the extension of the RBS was triggered by tactile sensations generated by objects that are attached to the robot's body, e.g., tools. The novel extension mechanism described here is triggered by the temporal contingency between the actions of the robot and the observed self-movements in the video image. Thus, the extension mechanism no longer requires direct physical contact as the TV image can be arbitrary translated, rotated, and scaled relative to the robot's body.

3. Experimental Setup

The experimental setup for the robot experiments (which are described below) is shown in Figure 2. All experiments were performed using a CRS+ A251 manipulator arm. The robot has 5 degrees of freedom (waist roll, shoulder pitch, elbow pitch, wrist pitch, wrist roll) plus a gripper. During the experiments, however, the two roll joints were not allowed to move away from their $0°$ positions, i.e., the robot's movements were restricted to the vertical plane.

The experimental setup uses 2 cameras. The first camera (Sony EVI-D30) is the only camera through which the robot receives its visual input. The second camera (Sony Handycam DCR-HC40) was placed between the robot and the first camera such that it can capture approximately $\frac{1}{3}$ of the working envelope of the robot. The frames captured by the second camera were displayed in real-time on a TV monitor (Samsung LTN-406W 40-inch LCD Flat Panel TV).

Six color markers were placed on the robot's body. The positions of the markers were tracked using histogram matching in HSV color space implemented with the openCV library. The same procedure was applied to track the positions of the color markers in the TV. Thus, after the color segmentation is performed the robot sees only a point-light display of the movements of different markers.

Through its camera the robot can observe both its real arm as well as the image of its arm in the TV monitor (see Figure 3.a). During some experiments the robot was not allowed to see its own body. For these experiments the left half of each frame captured by the robot's camera was digitally erased (zeroed) before it was processed (see Figure 3.b).

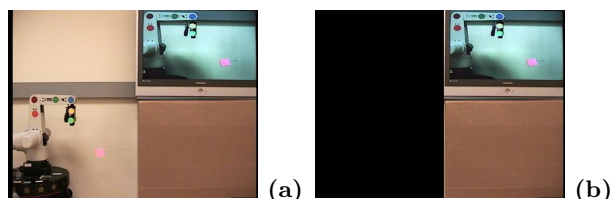Figure 2: Experimental setup for the robot experiments.

Figure 3: a) View from the robot's camera. b) During some experiments the robot is not allowed to see its own body (see text for details).

4. Self-Detection in the TV

This section describes how the robot can detect that the movements in the TV image are generated by its own motor commands (sub-problem 1 described in Section 1.). This problem is related to the general problems of self-detection and "self" versus "other" discrimination.

The novel method for self-detection described here makes the following assumptions. Let there be a set of visual features $F = \{f_1, f_2, \ldots, f_k\}$ that the robot can detect and track over time. Some of these features belong to the robot's body. Other features belong to the external environment and the objects in it. The robot can detect the positions of visual features and detect whether or not they are moving at any given point in time. The goal of the robot is to classify the set of features, F, into either "self" or "other." In other words, the robot must split the set of features into two subsets, F_{self} and F_{other}, such that $F = F_{self} \cup F_{other}$.

The self-detection problem described above is solved by splitting it into two sub-problems: 1) estimating the efferent-afferent delay of the robot; and 2) using the value of this delay, coupled with two probabilistic estimates, to classify the features as either "self" or "other."

The robot can estimate its efferent-afferent delay by measuring the elapsed time from the start of a motor command to the start of a visual movement

for some feature f_i (see Figure 4). A reliable estimate can be computed by executing multiple motor commands over and extended period of time.

Four different experiments lasting 45 minutes each were performed to estimate the value of the delay when the TV was not part of the scene. Figure 5 shows a histogram of the measured efferent-afferent delays during the first experiment; the results for the other three experiments are similar.
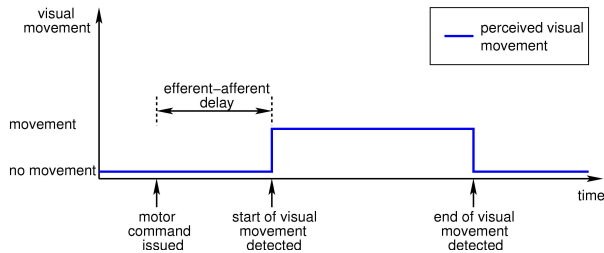

Figure 4: The efferent-afferent delay is defined as the time between the start of a motor command (efferent signal) and the start of visual movement (afferent signal). This delay can be learned from self-observation data.
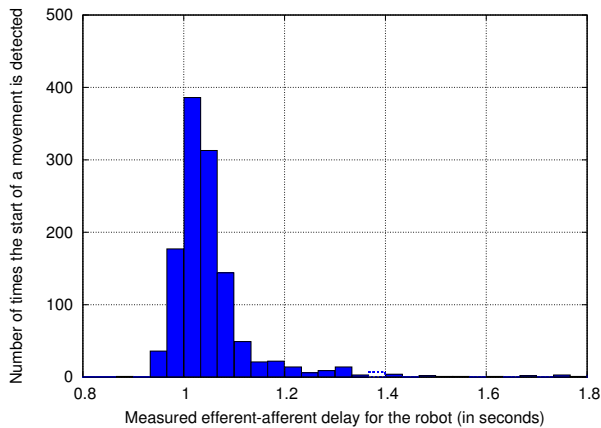

Figure 5: Histogram of the measured efferent-afferent delays for the robot during an experiment lasting 45 min.

The average delay value after the four experimets was estimated at 1.035 seconds. Once the value of the efferent-afferent delay is calculated the robot can identify which features have movements that are temporally contingent upon its own motor commands. This also includes the features in the TV image as they are displayed in real time. A movement is considered to be temporally contingent if its start occurs within $\pm 25\%$ of the average efferent-afferent delay following a motor command.

Up to this point, our method for self-detection is somewhat similar to the method described in (Michel et al., 2004). Their method, however, performs self-detection through movement detection at the pixel level and thus cannot keep a permanent track of which features belong to the robot's body. Our method performs the detection at the feature level (body markers) and also maintains a probabilistic

estimate across all features. The other novel modifications are described below.

For each feature the robot maintains two independent probabilistic estimates which jointly determine how likely it is for the feature to belong to the robot's body. The two probabilistic estimates are the *necessity index* and the *sufficiency index* described by Watson (1994). The necessity index measures whether the feature moves consistently after every motor command. The sufficiency index measures whether for every movement of the feature there is a corresponding motor command that preceded it. Figure 6 shows an example with three visual features. The formulas for the necessity and sufficiency indexes are given below.

$$\text{Necessity} \quad = \frac{\text{\# of temporally contingent movements}}{\text{\# of motor commands}}$$

$$\text{Sufficiency} = \frac{\text{\# of temporally contingent movements}}{\text{\# of observed movements for this feature}}$$
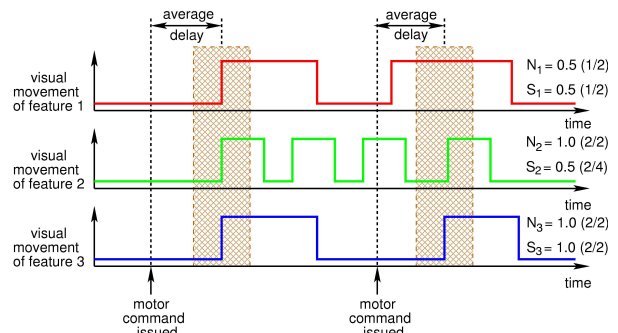

Figure 6: The figure shows the calculated values of the necessity (N_i) and sufficiency (S_i) indexes for three visual features. After two motor commands feature 1 is observed to move twice but only one of these movements is contingent upon the robot's motor commands. Thus, feature 1 has a necessity $N_1 = 0.5$ and a sufficiency index $S_1 = 0.5$. The movements of feature 2 are contingent upon both motor commands (thus $N_2=1.0$) but only two out of four movements are temporally contingent (thus $S_2=0.5$). All movements of feature 3 are contingent upon the robot's motor commands and thus $N_3 = S_3 = 1.0$. Based on these results, feature 3 can be classified as "self" and features 1 and 2 can be classified as "other".

Both of these indexes are updated over time as new evidence becomes available, i.e., after a new motor command is issued or after the feature is observed to move. The belief of the robot that f_i is part of its body at time t is given jointly by $N_i(t)$ and $S_i(t)$. If both are greater than some threshold value, α, then feature f_i is classified as "self," i.e.,

$$f_i \in \begin{cases} F_{self} & : \text{ iff } N_i(t) > \alpha \text{ and } S_i(t) > \alpha \\ F_{other} & : \text{ otherwise} \end{cases}$$

Ideally, both $N_i(t)$ and $S_i(t)$ should be 1. In practice, however, this is rarely the case as there is al-

ways some sensory noise that cannot be filtered out. Therefore, for all robot experiments the threshold value, α, was set to 0.75.

The above formula is valid only if all body markers of the robot start to move at the same time for every motor command, i.e., if all joints start to move together. If this condition is violated the necessity indexes must be preconditioned on the type of motor command, m, performed by the robot for each movement. Thus, feature f_i is classified as "self" if there exists at least one motor command, m, for which both N_i^m and S_i are above the threshold α after some time interval t. In other words,

$$f_i \in \begin{cases} F_{self} & : \quad \text{iff } \exists m : N_i^m(t) > \alpha \text{ and } S_i(t) > \alpha \\ F_{other} & : \quad \text{otherwise} \end{cases}$$

In order to detect the body markers in the TV image as "self" their N_i^m and S_i indexes should be updated only when these features are visible and not when they are outside the TV frame. The correspondence between individual body markers and their image in the TV is established using the nearest neighbor rule in color space.

To test this method for self-detection in the TV three experiments were conducted. During each one the robot performed 500 random motor commands while observing both its own body and its image in the TV (see Figure 3.a). Figure 7 shows the sufficiency indexes for the six TV markers plotted over time during one of three experiments (the results are typical for all three experiments). The plots for the necessity indexes are similar to this one when the individual motor commands are taken into account.
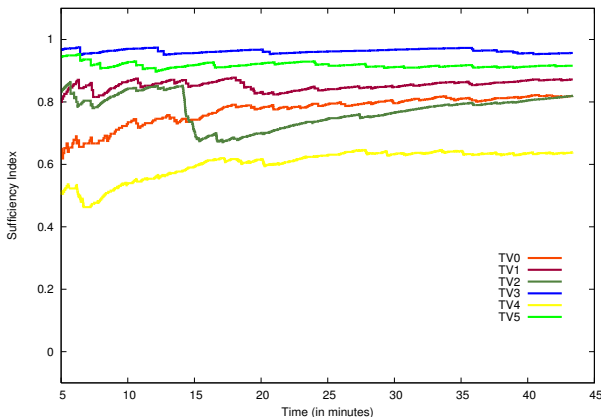

Figure 7: The sufficiency indexes calculated over time for the six TV markers. These results are calculated after taking the visibility of the markers into account.

In all three experiments all TV markers were correctly identified as "self." The only exception was the yellow marker whose position detection noise was larger than that for the other markers. The redundancy in the body markers (2 per rigid body in this case) ensures that the robot's video-guided behaviors described below can be performed even if some TV markers are incorrectly classified as "other."

5. Morphing the Body Schema

This section describes the method for calculating the similarity transformation (translation, rotation, and scale) between the position of the real robot arm and the image of the arm in the TV monitor, i.e., subproblem 2 described in Section 1. First, however, the representation for the robot's body is described.

Once the robot has identified which features belong to its body it can build a model for their most likely positions given the robot's joint vector, i.e., the robot can learn its own body schema. Due to space limitations this section provides only a brief description of the robot body schema model which is described in our previous work (Stoytchev, 2003).

The model is built around the concept of a *body icon* which is a pair of vectors $(\tilde{\mu}_i, \tilde{\beta}_i)$ representing the motor and sensory components of a specific joint configuration of the robot. The vector $\tilde{\mu}_i = \left[\tilde{\theta}_1^i, \tilde{\theta}_2^i, \ldots, \tilde{\theta}_M^i\right]$ represents a specific joint configuration. The vector $\tilde{\beta}_i = \left[\tilde{v}_{r_1}^i, \tilde{v}_{r_2}^i, \ldots, \tilde{v}_{r_N}^i\right]$ represents the coordinates in camera-centric coordinates of the robot's body markers for the given joint vector $\tilde{\mu}_i$. A large number of empirically learned body icons, $\left[(\tilde{\mu}_i, \tilde{\beta}_i), i = 1, \ldots, \mathcal{I}\right]$, is used. It is believed that the brain uses a similar representation encoded as a cortical map (Morasso and Sanguineti, 1995).

Figure 8 shows 500 observed positions for the red and the blue body markers (see Figure 2). These positions were recorded from real robot data while the robot was performing motor babbling when the TV was not part of the scene.
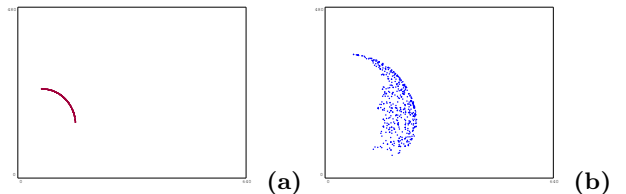

Figure 8: The visual components, $\tilde{v}_{r_i}$, in 500 body icons corresponding to: a) the red body marker; and b) the blue body marker (see Figure 2).

This body representation has several useful properties. Most notably, for an arbitrary joint vector μ the forward model $\beta = \beta(\tilde{\mu})$ can be approximated as

$$\beta^{approx}(\mu) \approx \sum_i \tilde{\beta}_i U_i(\mu) \qquad (1)$$

where U is a normalized Gaussian or softmax function (Morasso and Sanguineti, 1995). Formula 1 is used to approximate the position of the body markers when they are hidden from the robot's view as described below (also see Figure 9).

The similarity transformation is calculated using the method described by Umeyama (1991) which requires two sets of points. The first set consists of different positions of a specific body marker (the blue
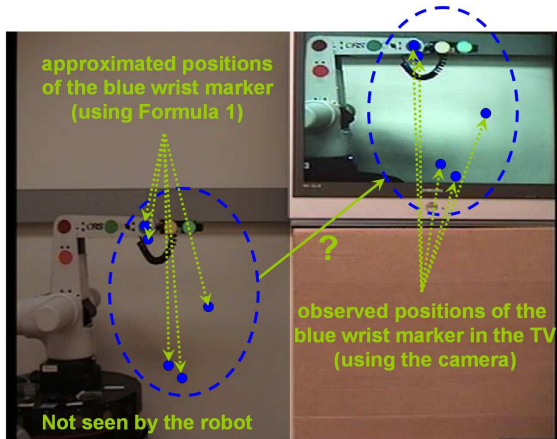
Figure 9: The robot calculates the similarity transformation (translation, rotation, and scale) between the position of its real arm and the image of its arm in the TV using two sets of points as shown in the figure. See text for more details.

marker shown in Figure 9 was used). These positions are estimated from proprioceptive information using Formula 1 as the body marker cannot be observed directly (see Figure 3.b). The second set consists of the corresponding positions of the same body marker but observed in the TV image. The robot gathers the two sets while performing motor babbling. If the wrist marker cannot be detected for some joint configuration (e.g., because it is out of the TV frame) the robot picks a new random joint vector and continues the motor babbling. For more details see (Stoytchev, 2007).

The next section describes 30 experiments (see Table 1) which were conducted to test this method for morphing the body schema. The calculated transformation parameters are used to extend the visual components of the body icons. After the extension the TV image can be used for video-guided behaviors as described in the next section.

The body schema representation can also be used for control of goal directed movements. The idea is to specify the movements in visual space but carry them out in motor space. The body schema representation allows that without the need for inverse kinematics transformations because the two spaces are combined into one in the form of body icons. For robots with multiple degrees of freedom several nested body schemas can be used. For more details see (Stoytchev, 2003, 2007).

The grasping behavior used in the next section was hand coded using the robot's body schema. During the reaching phase the robot moves its blue marker over the target. It then orients its wrist (using a second body schema for the wrist relative to the arm). Next, it lowers its arm by controlling the blue marker again. Finally, it closes its gripper.

6. Video-Guided Grasping

This section builds upon the previous two sections to achieve video-guided grasping behaviors, i.e., subproblem 3 described in Section 1.

Three experimental conditions were used to test the robot's abilities to perform video-guided grasping behaviors. For each condition 10 experiments were performed. The conditions differ by the rotation and zoom of the camera which affects the orientation and the scale of the TV image. The three test conditions are similar to the ones described by Iriki et al. (2001). In the first condition the TV image is approximately equal to the image of the real robot (see Figure 10.a). In the second condition, the camera is rotated by $50°$ and thus the TV image is rotated by $-50°$ (see Figure 10.b). In the third condition the camera is horizontal but its image is zoomed in (see Figure 10.c). The zoom factor is 1.6.

Because the setup was in 2D and stereo vision was not used there are only two meaningful translation parameters (t_x and t_y) and only one rotation parameter (θ_z). The scale factor was also estimated automatically. The transformation parameters were estimated correctly in all 30 trials (see Table 1).

Condition	Statistic	t_x	t_y	θ_z	Scale
Normal	Mean	336.8	234.9	$+0.23°$	0.932
(10 trials)	Stdev	18.5	17.2	$2.61°$	0.077
Rotated	Mean	258.9	425.0	$-49.84°$	0.985
(10 trials)	Stdev	17.5	11.6	$2.64°$	0.055
Zoomed in	Mean	206.8	80.5	$-1.16°$	1.606
(10 trials)	Stdev	16.5	24.3	$2.76°$	0.065

Table 1: Estimated parameters for the similarity transformation used to extend the robot's body schema.

For each of the 3 test conditions the robot was tested on the task of grasping an incentive object (pink object in Figure 10) which could only be seen in the TV image (see Figure 3.b). The grasping experiment was performed five times for each of the 3 test conditions after the transformation parameters have been estimated. The robot successfully grasped the incentive object in 5 out of 5 times in the normal condition; 4 out of 5 in the rotated condition; and 0 out of 5 in the scaled condition. In other words, the robot was able to successfully perform video-guided grasping in the normal and the rotated test conditions but not in the zoomed in test condition.

The reason why the robot failed in the scaled condition is due to the poor quality of the color tracking results at this high level of magnification. This result is counter-intuitive as one would expect just the opposite to be true. However, when the image of the robot's arm is really large the auto color calibration of the Sony Handycam (which could not be turned off) is affected even by the smallest movements of the robot. Shadows and other transient light effects are also magnified.

Figure 10: The figure shows three views from the robot's camera, one for each of the three experimental conditions. The image of the robot in the TV is: a) approximately the same size as the real robot; b) rotated by negative 50°; and c) scaled (zoomed in) by a factor of 1.6. During the video-guided grasping experiments, however, the robot cannot observe its own body (see Figure 3.b)
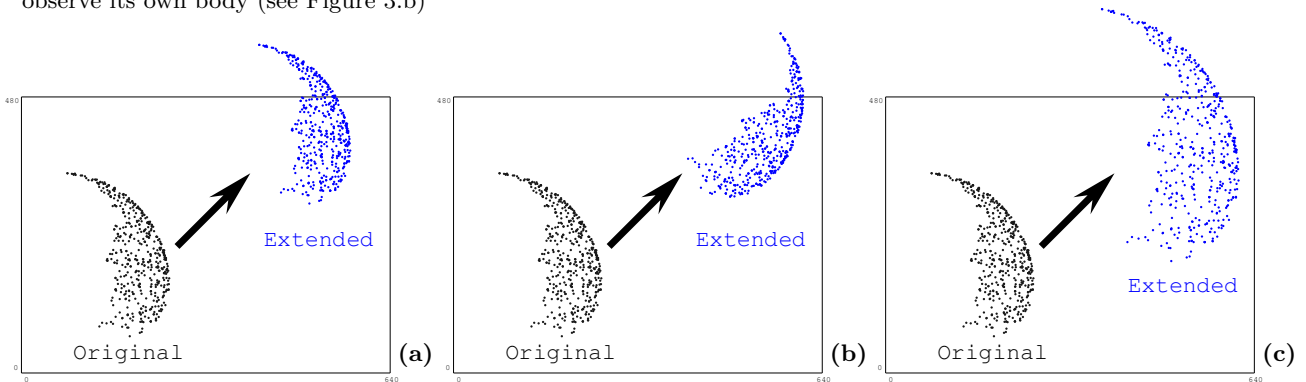

Figure 11: The figure shows the extended positions of the body icons (visual components for the blue wrist marker only) after the extension of the RBS. By comparing this figure with Figure 8 it is obvious that the visual components of the body icons are: (a) translated; (b) rotated and translated; and (c) scaled, rotated and translated relative to their original configuration. Furthermore, the new positions coincide with the positions in which the blue marker can be observed in the TV. Because the extended positions are no longer tied to the camera coordinates some of them may fall outside the 640×480 camera image.

Thus, it proved difficult to track the body markers in the zoomed in test condition using the current tracking method. Nevertheless, the transformation parameters for this test case were estimated correctly (see Figure 11.c) because the ten data points needed for the calculation are collected only when the wrist marker can be observed in the TV(which was possible for some body configurations but not others). The failure in the rotated test case was also due to poor color tracking results.

7. Conclusions

This paper described the components of a system which can be used by a robot to achieve video-guided behaviors. The paper showed how a robot can detect its self-image in a TV monitor and use that real-time video image to guide its own actions in order to reach an object that can only be observed through the TV image. To the best of our knowledge, this constitutes the first example of video-guided behaviors in robots reported in the robotics and AI literature.

To achieve this goal the paper introduced two novel methods. First, the paper introduced a novel method for feature-level self-detection in robots. This method maintains probabilistic estimates of necessity and sufficiency across visual features as to whether or not they belong to the robot's body. The methods was successfully used by the robot to detect its self-image in the TV.

Second, the paper introduced a novel method for extending the body schema of the robot which is triggered by the temporal contingency between the actions of the robot and the observed self-movements in the video image. Through extension of the body schema the robot can use the TV image to guide its arm movements in the same way as if it were observing its own arm directly.

Future work can build upon the principles described in this paper and extend the domains in which robots can use video-guided behaviors. For example, using a mouse to position a cursor on a computer monitor or using a rear view camera (or mirror) to back up a car are just two possible applications.

At this point it might be premature to try and draw conclusions from this study and apply them to analyze the ways in which the body schema of primates works. The experiments described by Iriki et al. (2001) served only as an inspiration to the robotics work. The specific representations described in this paper, while biologically plausible, are probably different from the representations used by the brain. What representations the brain uses is any body's guess at this point.

Nevertheless, robot studies can be used to model the main building blocks of the biological body schema. If these building blocks are treated as black boxes (i.e., if we focus more on what they do rather than how they do it) then roboticists can research the ways in which these black boxes interact with each other. This paper makes a small step in that direction.

Acknowledgments

References

Barth, J., Povinelli, D., and Cant, J. (2004). Bodily origins of self. In Beike, D., Lampinen, J. L., and Behrend, D. A., (Eds.), *The Self and Memory*, pages 11–43. Psychology Press, New York.

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

Berthoz, A. (2000). *The brain's sense of movement*. Harvard U. Press.

Epstein, R., Lanza, R. P., and Skinner, B. F. (1981). "self-awareness" in the pigeon. *Science*, 212:695–696.

Frederiks, J. (1969). Disorders of the body schema. In *Handbook of Clinical Neurology: Disorders of Speech Perception and Symbolic Behavior*, volume 4, pages 207–40.

Gallup, G. (1970). Chimpanzees: self-recognition. *Science*, 167(3914):86–7.

Graziano, M., Cooke, D., and Taylor, C. (2000). Coding the location of the arm by sight. *Science*, 290:1782–1786.

Graziano, M., Taylor, C., and Moore, T. (2002). Complex movements evoked by microstimulation of precentral cortex. *Neuron*.

Head, H. and Holmes, G. (1911). Sensory disturbance from cerebral lesions. *Brain*, 34:102–254.

Iriki, A. et al. (1996). Coding of modified body schema during tool use by macaque postcentral neurons. *Neuroreport*, 7:2325–2330.

Iriki, A. et al. (2001). Self-images in the video monitor coded by monkey intraparietal neurons. *Neuroscience Research*, 40:163–173.

Itakura, S. (1987). Mirror guided behavior in japanese monkeys (macaca fuscata fuscata). *Primates*, 28:149–161.

Maravita, A. and Iriki, A. (2004). Tools for the body (schema). *Trends in Cognitive Sciences*, 8(2):79–86.

Menzel, E. et al. (1985). Chimpanzee (pan troglodytes) spatial problem solving with the use of mirrors and televised equivalents of mirrors. *J. of Comparative Psychology*, 99:211–217.

Michel, P., Gold, K., and Scassellati, B. (2004). Motion-based robotic self-recognition. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Sendai, Japan.

Morasso, P. and Sanguineti, V. (1995). Self-organizing body-schema for motor planning. *Journal of Motor Behavior*, 27(1):52–66.

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

Nabeshima, C., Lungarella, M., and Kuniyoshi, Y. (2005). Timing-based model of body schema adaptation and its role in perception and tool use: A robot case study. In *Proc. of the 4-th Intl. Conference on Development and Learning (ICDL)*, pages 7– 12.

Ramachandran, V. and Blakeslee, S. (1998). *Phantoms in the Brain: Probing the Mysteries of the Human Mind*. William Morrow, NY.

Schultz, S. (2001). *Princeton Weekly Bulletin*, 90(26).

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

Stoytchev, A. (2007). *Robot Tool Behavior: A Developmental Approach to Autonomous Tool Use*. PhD thesis, College of Computing, Georgia Institute of Technology.

Umeyama, S. (1991). Least-squares estimation of transformation parameters between two point patterns. *IEEE PAMI*, 13(4):376–80.

Watson, J. S. (1994). Detection of self: The perfect algorithm. In Parker, S., Mitchell, R., and Boccia, M., (Eds.), *Self-Awareness in Animals and Humans: Developmental Perspectives*, pages 131–148. Cambridge University Press, Cambridge.

Yoshikawa, Y., Kawanishi, H., Asada, M., and Hosoda, K. (2002). Body scheme acquisition by cross map learning among tactile, image, and proprioceptive spaces. In *Proceedings of the Second International Workshop on Epigenetic Robotics*.