# Building Subjective Spatial Perception Based on Sensor Space Integration for Motion Generation

Yuichi Kobayashi and Eisuke Kurita
Tokyo University of Agriculture and Technology
Koganei, Tokyo, Japan

Manabu Gouko
Tohoku Gakuin University
Tagajo, Miyagi, Japan

## I. Introduction

To develop skills of autonomous robots, such as manipulating objects and navigation, it is important to construct spatial perception in a bottom-up way. In the case of humans, sensory organs are located suitably for their tasks. For example, tactile sensors on fingers provide rich information on how a human is contacting with an object. On the other hand, observation spaces of multimodal sensors do not always overlap each other in general (Fig.1), which has been discussed by sensor fusion [1] but not efficiently resolved. This paper proposes a motion generation based on multiple sensors having limited sensing range, with a discussion on the implementation issues on a humanoid robot with vision. To combine two observation spaces whose sensing areas do not overlap with each other, an extension of the observation space is proposed. To extend the observation space to the region out of the sensing range, diffusion-based learning [2] is introduced. Compared with reinforcement learning framework such as [3], [4], the proposed framework provides much more simple way given that continuity and smoothness of sensing and actuation holds. From the viewpoint of implementation of the proposed framework, getting depth information from visual information indicated in Fig.1 helps building three-dimensional perception of the peripersonal space. It will be also discussed how to detect the robot's body including depth information from the visual information under the assumption of avoiding embedded knowledge. We present a feature-based representation of the robot's body and discuss its applicability to depth perception.

## II. Integration of multiple sensor spaces by diffusion-based learning

As shown in Fig.2, observation space of two sensors do not overlap each other in the state space. The objective of a robot is to control itself from an initial point in an observation space to a target point in another observation space. Under the assumption that the system dynamics is unknown, the robot first learns the mapping from the control input to each observation vector by randomly moving around, which is restricted within the sensing range of each sensor. Basic idea of the proposed method is to extend the obtained mapping within the sensing range toward outside of the sensing range. The robot repeats moving out of the sensing range and then coming back to it. Using the history of control inputs during the sequential motion and the resultant observation vector obtained when coming back to the sensing range, the robot estimates the mapping from the control input to the observation vector. This extension of the sensing range can be regarded as construction of a 'virtual' observation space and the virtual observation space overlaps with the region of the other sensor space. When a task is given to the robot, it uses the information of virtual observation variable to generate a motion from the observation space (currently sensible) to the other observation space (see Fig.3).

## III. Simulation & Experiment

Suppose a two-link manipulator as indicated in Fig.4. The position of the end effector is detected by the camera, as far as the end effector is inside of the viewing range of the camera. Two proximity sensors are attached with the manipulator at the end effector. The direction of two sensors differs $\theta_s$. Each proximity sensor gives a value proportional to the distance to the floor and it has limitation of sensing range, that is, it does not give a value when the distance to the floor is over a

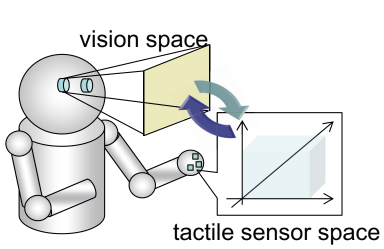


Fig. 1. Integration of sensor spaces

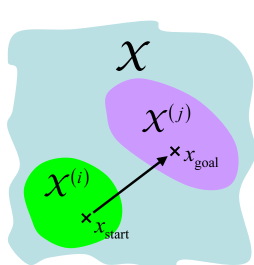


Fig. 2. Multiple sensor spaces

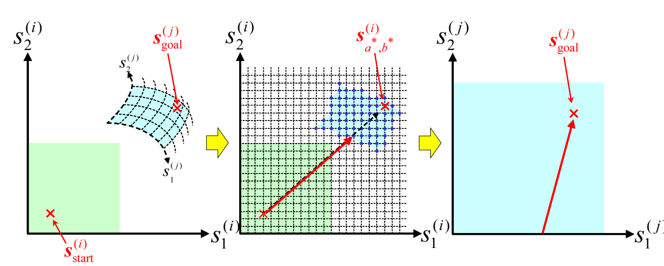


Fig. 3. Proposed motion generation

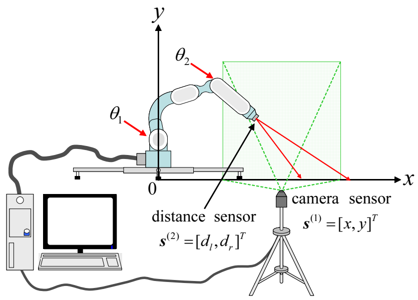


Fig. 4. Manipulator with camera and proximity sensor

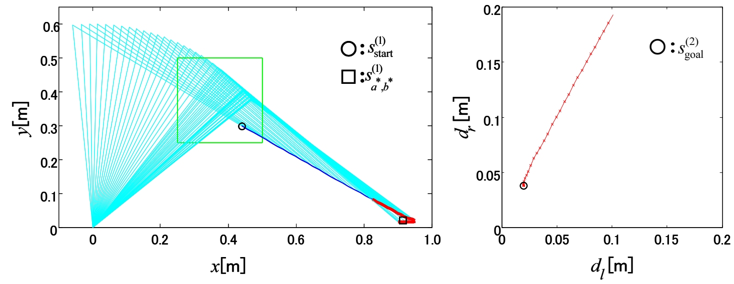


Fig. 5. Generated trajectory with proposed integration method

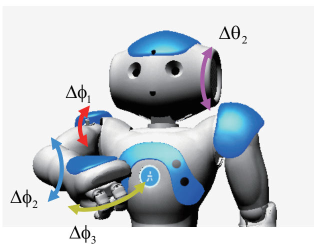


Fig. 6. Configuration of robot

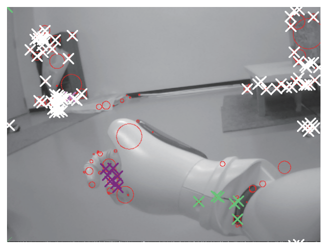

Fig. 7. Result of clustering

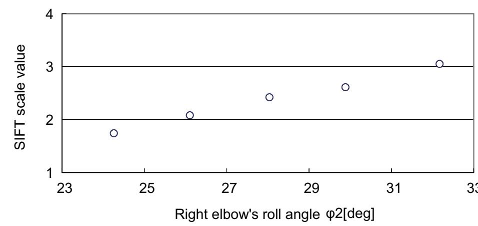


Fig. 8. Scaling parameter of SIFT feature

certain value. Trajectory realized by the proposed control is shown in Fig.5. The left hand shows the total trajectory in the camera coordinates and the right hand shows the trajectory in the proximity sensor coordinates. The robot could reach the target configuration through a region where both sensors can not obtain signals. Though the trajectory is almost straight in the first phase controlled with Jacobians of the camera, the last part of the trajectory in the second phase with Jacobians of the distance sensors is winding. This is caused by non-linear mapping between two observation variables.

An experiment was conducted using humanoid robot NAO. Three joint angles $\phi_1, \phi_2, \phi_3$ at its right arm were controlled (Fig.6). SIFT keypoints [5] were extracted from images obtained by the CMOS camera. Function approximation with RBF network was applied to the positions of the keypoints that succeeded to find matching. The keypoints were clustered using mean-shift algorithm with approximated Jacobians in the image coordinate. The result of clustering is shown in Fig.7. Fig.8 shows scale parameters of a keypoint at the end of the arm with the different elbow joint angles. The first result of clustering indicates that it is possible to extract the robot's body from the image without a specific knowledge based on a feature extraction and matching method. Moreover, keypoints can be separated using the approximated parameters, which expresses dependencies between motion in the image and change of the joint angles. The second result of the change of scale parameter indicates that it is possible to detect (change of) depth information in the image through the change of scale parameter of the SIFT features.

## IV. Conclusion

In this paper, a motion generation for robot that has multiple sensors with limited sensing range was proposed. The proposed method realized extrapolation of Jacobian where observation is not available, based on the diffusion-based learning. The proposed framework was verified by reaching motion toward the floor with a manipulator. The experiment with a humanoid robot showed that SIFT image features are available for body extraction and building spatial representation of the arm. The proposed framework can be extended to more general case with more sensors by considering how to select and combine observation variables, which will be one of the future works.

## Acknowledgment

This work was partly supported by Support Center for Advanced Telecommunications Technology Research and the Grant-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology, Japan (22700198), Grant-in-Aid for Young Scientists (B).

## References

[1] R. C. Luo, *et al.* , Multisensor fusion and integration: approaches, applications, and future research directions, IEEE Sensors Journal, 2, 2, 107-119, 2002.
[2] Z. Luo and M. Ito, Diffusion-based learning theory for organizing visuo-motor coordination, Biological Cybernetics, 79, 279-289, 1998.
[3] L. P. Kaelbling, *et al.* , Planning and acting in partially observable stochastic domains, Artificial Intelligence, 101, 99-134, 1998.
[4] D. Wierstra, *et al.* , Solving deep memory POMDPs with recurrent policy gradients, Proc. of ICANN, 2007.
[5] D. G. Lowe, Object Recognition from Local Scale-Invariant Features, Proc. of the Int. Cenf. on Computer Vision, 2, 1150-1157, 1999.