

# Active Vision and Depth Estimation Toward a Peripersonal Space Encoding in a Humanoid Robot

Salomon Ramirez-Contla      Davide Marocco      Angelo Cangelosi  
School of Computing and Mathematics  
University of Plymouth

## 1. Introduction

Peripersonal space is defined as the space around a person's body, which is the space that defines the region of interactions between an agent and its environment. Estimating the distances at which objects are, in other words delimiting reachable (peripersonal) and not reachable (extrapersonal) space, is very important in order to properly interact with the environment and learn from it. Interestingly, behavioural and neurophysiological studies suggest that the brain encodes the peripersonal space differently from the extrapersonal space and that the coding of the former is achieved through the integration of different modalities (Farné et al., 1998).

This abstract presents a preliminary result of an ongoing work for developing an agent capable of learning a peripersonal space representation, the integration of different modalities and the use of active vision for interacting with objects in the peripersonal space.

Depth estimation in biological agents can be of great importance for a good performance in crucial tasks such as reaching, grasping or avoiding obstacles (Mon-Williams and Dijkerman, 1999). From literature it's known that monocular vision provides indirect cues for depth perception: motion parallax, accommodation effort, casted shadows by near objects and contrast to name some. Nevertheless, those cues can only be used in certain circumstances and in most of the cases the use of monocular depth indicators requires complex processing on the acquired image. For this reason, processes and algorithms that extract depth information from vision are widely focused on stereo images. That is, images of the same scene taken from two slightly different position (Reichelt et al., 2010). Besides the possible algorithms that can be applied to stereo images, vergence, an additional proprioceptive information is available to organisms endowed with two movable eyes. Vergence is the oculomotor adjustment needed to foveate the same point in space with both eyes. Recent studies show that in humans, vergence occurs well before the actual depth estimation (Wismeijer et al., 2008) and therefore it can be an important cue even in the absence of complex monoc-

ular cues or processing of stereo images. In this part of the work we study the possible relevance of vergence in the development of a peripersonal space representation.

## 2. Material and Methods

The experiment was carried out with a simulated version of an iCub humanoid platform (Tikhonoff et al., 2008). The task was to reach a red cube placed in front of the robot with the right hand. Only 5 DoF of the iCub arm were used. Two different conditions of the task were considered: using monocular or binocular vision. Tracking and foveating the object was achieved by an closed-loop pre-programmed controller that moves the head and the eyes of the robot so to locate the target's centroid in the centre of the right eye image, or in both eyes' images, respectively. In the binocular vision case, modulation of vergence was required.

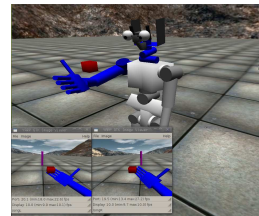


Figure 1: The simulated iCub performing the reaching task. Note the red target object and the blue painted arm.

A neural controller was used for moving the right arm of the simulated robot. Input for this controller was the proprioceptive information (pan and tilt joint positions from the head, and tilt, pan and vergence joints from the eyes) and pre-processed visual information. Pre-processing was colour based image segmentation: red was used in the case of the target object and blue for the arm of the robot (see figure 1). Image segmentation provided then two sets of data that were fed into the neural controller.

The controller was a feed forward, partially connected neural network with the following architecture: one input layer; one hidden layer *hA* which re-

ceives connections from visual input; an additional hidden layer  $hB$  which received connections from proprioceptive input and hidden layer  $hA$ ; an output layer which receives connections from the proprioceptive input and the two hidden layers  $hA$  and  $hB$ . This architecture was devised in order to analyse *unimodal* and *bimodal* contributions to depth perception, as described in (Farné et al., 1998), for the reaching task in a later stage of the study. The neural controller was trained using backpropagation (Rumelhart and McClelland, 1986). Training data consisted of 120 input/output pairs. Inputs corresponded to visual and proprioceptive data and desired outputs corresponded to a set of arm joints positions. For collecting the training data set the robot was pre-programmed to perform a reaching and grasping action followed by motor babbling. Half of the data was obtained using monocular vision and the rest using binocular vision because the controller was expected to generalise and perform well in both conditions.

### 3. Results

Data for this preliminary test was collected by placing the target object in 18 different positions. For each position the controller was activated in order to track and reach the object, starting from the arm in a home position along the body of the robot. Two test cases were used: *binocular* vision using vergence, and *monocular* vision both with the network continuously activated. Accuracy, direction of the arm respective to the target’s position and depth perception were measured for the 18 target positions. Accuracy was measured as the distance between the target and the palm of the hand. Direction was measured as the angle between a line from the head to the object and a line between the head and the hand in the horizontal plane. This measure gives an indication about the orientation of the arm with respect to the object. Depth perception was measured as the difference between the distance from the head and the target and the distance between the head and the hand. Results show that the robot is able to generalise well

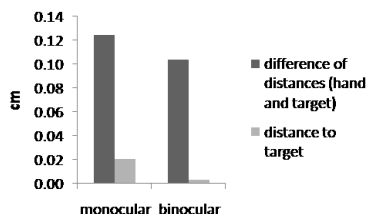


Figure 2: Comparison of the two visual systems’ performance. Distances are in cm.

from the training set and shows smooth transitions from different positions when continuously activated.

Analysis of the data shows that the effectiveness of the system to move the hand to the required direction in order to reach the target is very similar for both cases. However, we were more interested on depth accuracy rather than direction. On this matter of the data shows tendencies of binocular data being more accurate when reaching the target and generally better for depth estimation as it can be seen in figure 2.

### 4. Conclusion

Experiments indicated the effectiveness of the use of vergence for depth estimation in a reaching task in a simple active vision system implemented on the iCub simulator. These results make the use of these kind of depth estimation system suitable for our further development of peripersonal space representation. Additional studies will be carried out for investigations of the contribution of the different modalities used (proprioception and vision) as well as the implementation of a peripersonal space encoding that utilises future versions of this multimodal system.

### References

- Farné, A., Ládavas, E., Zeloni, G., and Pellegrino, G. (1998). Neuropsychological evidence of an integrated visuotactile representation of peripersonal space in humans. *Journal of Cognitive Neuroscience*, 10(5):581–589.
- Mon-Williams, M. and Dijkerman, H. (1999). The use of vergence information in the programming of prehension. *Experimental Brain Research*, 128(4):578–582.
- Reichelt, S., Häussler, R., Fütterer, G., and Leister, N. (2010). Depth cues in human visual perception and their realization in 3d displays. volume 7690, page 76900B. SPIE.
- Rumelhart, D. and McClelland, J. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition*. M.I.T. Press, Cambridge.
- Tikhanoff, V., Cangelosi, A., Fitzpatrick, P., Metta, G., Natale, L., and Nori, F. (2008). *Proceedings of IEEE Workshop on Performance Metrics for Intelligent Systems Workshop (PerMIS’08)*, chapter An open-source simulator for cognitive robotics research: The prototype of the iCub humanoid robot simulator. Washington, D.C.
- Wismeijer, D., van Ee, R., and Erkelens, C. (2008). Depth cues, rather than perceived depth, govern vergence. *Experimental Brain Research*, 184(1):61–70.