Developmental Learning of Integrating Visual Attention Shifts and Bimanual Object Grasping and Manipulation Tasks

Sungmoon Jeong and Minho Lee
School of Electronics Engineering,
Kyungpook National University,
1370 Sankyuk-Dong, Puk-Gu, Taegu 702-701, Korea
jeongsm@ee.knu.ac.kr; mholee@knu.ac.kr

Hiroaki Arie and Jun Tani*
Lab. for Behavior and Dynamic Cognition,
RIEKN Brain Science Institute,
2-1 Hirosawa, Wako-shi, Saitama, 3510198, Japan
arie@bdc.brain.riken.jp; tani@brain.riken.jp

Abstract—**In order to achieve visual-guided object manipulation tasks via learning by example, the current neuro-robotics study considers integration of two essential mechanisms of visual attention and arm/hand movement and their adaptive coordination. The present study proposes a new dynamic neural network model in which visual attention and motor behavior are associated with task specific manners by learning with self-organizing functional hierarchy required for the cognitive tasks. The top-down visual attention provides a goal-directed shift sequence in a visual scan path and it can guide a generation of a motor plan for hand movement during action by reinforcement and inhibition learning. The proposed model can automatically generate the corresponding goal-directed actions with regards to the current sensory states including visual stimuli and body postures. The experiments show that developmental learning from basic actions to combinational ones can achieve certain generalizations in learning by which some novel behaviors without prior learning can be successfully generated.**

Keywords-component; shift sequence in visual scan path; action generator; object manipulation task

I. INTRODUCTION

HUMANS control gaze shifts and fixations (visual attention) proactively to gather visual information for guiding movements, which is highly related to a specified task [1]. This visual attention can improve reach accuracy by providing visual feedback on the hand position to guide the hand to a target [2] and guiding the hand even when the hand is not visible [3]. Also the visual attention can effortlessly detect (location) and recognize (identification) an interesting area or object within natural or cluttered scenes through the selective attention mechanism with various visual features such as color, orientation, scale and symmetry [1], [4], [5]. To achieve the visual-guided object manipulation tasks, the visual attention needs to switch to a specified task in time with hand movement.

Andrew proposed a rule based reinforcement learning by visual attention and short-term memory with similar experiments to teach the visual attention shifts for goal-directed behaviors [6]. They defined the states of environments (world and sensory states) then taught the policy, which consists of “if-then-else” tree statements, to the model to avoid the obstacle in the driving way. It is important to find the close connection of visual attention shifts with action sequences such as hand/arm control, but it is complex because it involves the visual guidance of both the eyes and hands [7]. In contrast to this model, this study’s model can automatically generate the actions with association of visual attention and motor behavior.

The current study examines how a set of object manipulation action can be learned by acquiring adequate visual attention shifts in a specific brain model. The visual attention shifts are realized by a top-down visual attention model which consists of a top-down visual attention command generator and a top-down visual saliency map (SM) combined with bottom-up SM models. Another essential idea is to utilize a functional hierarchy and to integrate visual attention and behavioral generation by employing a new dynamic neural network model so-called the multiple timescale recurrent neural network (MTRNN) [8], [9]. It has been shown that a certain functional hierarchy can develop through learning of complex behaviors by utilizing timescale differences of neural activities set in the network model. More specifically, it was shown that a set of behavior primitives (reusable movement segments) are acquired in the fast dynamic networks in a lower level while the sequencing of these primitives takes place in the slow dynamics in a higher level. In the current model, the function for a top-down visual attention command generator is considered to be acquired in the slow dynamic networks along with the sequencing of the behavior primitives. These two different roles of dynamic networks can enforce the developmental learning for complex actions by learned behavior primitives. The output command of the visual shift from slow dynamic networks is sent to a hard-wired gaze system to achieve precise gaze control. To verify the proposed model performance, a humanoid robot experiment was conducted, which is similar to previous human experiment setting [1], [10]-[12]. They presented eye–hand coordination when subjects moved color blocks from a pickup area and placed them in a desired location. Subjects attended a block before picking it up and a desired location before placing the block. In the current experiment the robot learns to perform similar visual attention shifts followed by acquired bimanual motion patterns; the robot attends to an object to pick it up and then to another destination object to place the object on.

The paper examines how a set of object manipulation actions combined with an active vision can be learned by

978-1-4244-6901-7/10/$26.00 ©2010 IEEE

introducing variances in object positions. Then, how learning of these basic actions and attention shifts can be generalized to achieve novel goal-directed actions which are composed by combining the basic actions is examined. This paper is organized as follows. Section 2 presents the problem description, the brain model of this study and the proposed new dynamic model. Section 3 presents the experimental results. Discussion and conclusions follow in Section 4.

II. Integrative Model of Visual Attention & Action

A. Biological Background

Fig. 1 shows the overall schematic of the two pathways of visual attention and arm motor movement to achieve the visual-guided tasks such as grasping and transporting an object to a desired location [13]-[15].

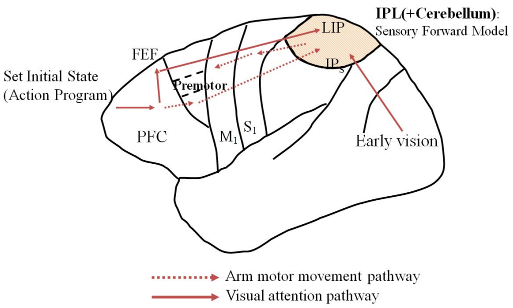


Figure 1. Action programs generating by using vision and action pathway.

From event related potential (ERP) and Magnetoencephalography (MEG) studies it has been shown that, in attention tasks, prefrontal cortex (PFC) is activated earlier than the parietal cortex, followed by the extrastriate and the striate cortices [13]. Therefore it is theorized that an action program for each task is set in PFC in terms of an initial state. The dynamics starts at the PFC with an initial state, which proceeds along two pathways for top-down visual attention and arm motor movement. The top-down visual attention is not directly modulated in frontal eye filed (FEF), but instead through intraparietal sulcus (IP_S), which is part of the parietal lobe [14]. The FEF may send top-down modulated signals to lateral intraparietal area (LIP). Neurons in LIP flexibly code color information, when color indicates a task-relevant location for an eye movement as saccades [15]. In the arm movement pathway proceeds from PFC, the premotor and inferior parietal lobe (IPL) to predict change of arm posture in time with receiving visual related inputs from V1. By predicting posture in terms of target joint angles in the next step, necessary motor torques to achieve a target angle are computed in M1.

B. System Overview

Fig. 2 describes the overall information flow in the proposed model. When receiving a desired action as the input in the slow dynamic networks of MTRNN, the network predicts how the arm posture in terms of proprioception m_t and visual attention command v_t change in time. Here, the visual attention command represents 4 object categories (red, green, blue and default preference color) to be attended. The visual attention system receives a visual attention command from the MTRNN and the retina image from the robot's vision. Then, matching is made between the top-down visual attention of an object specified by a visual attention command and a specific color preferable SM generated from a retina image. By selecting the visual object to be looked at through this matching process, the robot generates a camera head movement to achieve the gaze to a selected object. The resultant camera head position s_t is sent back to the input part of the MTRNN as representing visual perception of relative position for objects. The prediction of an arm posture m_{t+1} is sent to the robot as a target joint angle of the arms at each time step, which results in actual movements of the arms.

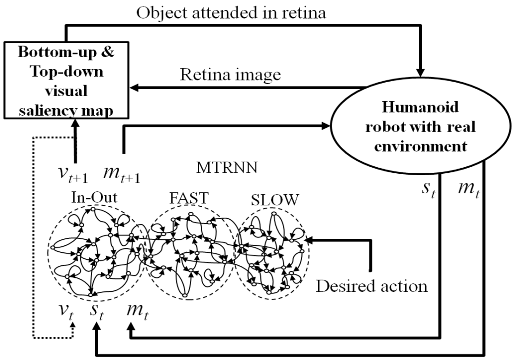

Object attended in retina
Bottom-up & Top-down visual saliency map
Retina image
Humanoid robot with real environment
v_{t+1} m_{t+1}
MTRNN
In-Out
FAST
SLOW
s_t m_t
Desired action
v_t s_t m_t

Figure 2. The architecture of the proposed model. In-Out: input-output context units, FAST: fast context unit, SLOW: slow context unit, m_{t+1}: predicted proprioception value at time step t, v_{t+1}: predicted top-down visual attention command at time step t

C. Bottom-up & Top-down Saliency Map for Attention Shifts

The dynamic neural network generates a goal-directed visual attention command to a vision system with a top-down visual SM to find a specified object and complete the task. Fig. 3 shows the architecture of a bottom-up and top-down visual SM model using reinforcement and inhibition networks [5], [16]. Itti et al. and Lee et al. used primitive features such as intensity, orientation, color and symmetry information to construct a bottom-up SM model [17], [18]. This study used only color features to construct a bottom-up SM because this experiment just consider a colorful object to move from one place to the other place in time. A localized area was regarded as that with the highest intensity values in the bottom-up SM as most salient regions to be analyzed for object identification. In the top-down manner, the human visual system determines salient locations through perceptive processing such as understanding and recognition considering task-dependent processing. In the course of detecting an object to achieve the object manipulation task, both the bottom-up and top-down processing work together for attention of a specified object region in an input scene. The Fuzzy adaptive resonance theory (ART) network together with the bottom-up SM model was used to implement a top-down visual SM model [5], [16]. The Fuzzy ART network learns and memorizes the characteristics of uninteresting areas and/or interesting areas selected by the bottom-up SM model. After training the reinforcement and inhibition networks by interaction with a human supervisor, the MTRNN generates goal-directed visual attention sequences through time with a test mode. Then, a localized area selected by a bottom-up SM model is tested for matching how much the selected area meets the visual characteristics of an object for a specific task generated by the MTRNN. The matching process is realized by comparing a vigilance value (ρ) between the selected color characteristics and the memorized color information. The model can focus on a specified colorful object, while it does not focus on a salient area with unwanted area.

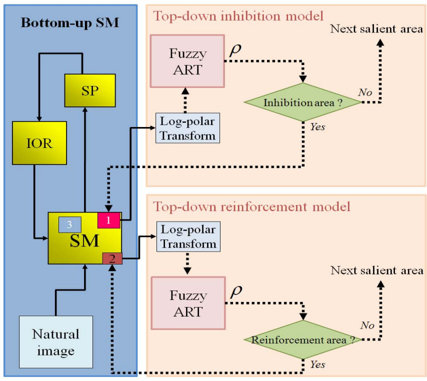


Figure 3. The architecture of a bottom-up and top-down SM model using reinforcement and inhibition networks. SP: salient point, IOR: inhibition of return. Square block 1 is an uninteresting area, but block 2 is an interesting area. Sold line: bottop-up attention, Dash line: top-down saliency

D. MTRNN for Behavior Generation Related with Top-down Visual Attention Command and Arm Movement

The MTRNN, which is a type of the continuous time RNN (CTRNN), was used to generate the behavior sequences [19], [20], in which neurons have different time scales therefore the MTRNN has the functional hierarchy characteristic.

1) Forward generation for Behavior Generation

The MTRNN has three groups of neural units in the present study, namely input-output units (116), fast context units (70) and slow context units (30). Among the input units, the first 64 units (i=1-64) correspond to the proprioceptive input (arm joint angle), the next 36 units (i=65-100) correspond to the visual input (neck joint angle) and the last 16 units (i=101-116) correspond to the visual attention command, respectively. The 14 dimensional inputs, which consist of 8 joint angles for two arms with 4 degrees of freedom in each arm, 2 joint angles for neck and 4 dimensional visual attention command, were thus transformed into 116 dimensional sparsely encoded vectors by a topology preserving map (TPM) with 3x10^6 training epochs [21]. This transformation reduces the redundancy of the input trajectories for units. The size of the TPMs is 64 (8 × 8) for proprioception, 36 (6 × 6) for the visual input and 16 (4 × 4) for the visual attention command. Fig. 4 shows the architecture of the MTRNN model with TPMs. The fast context units are connected with the input-output units of which synaptic weights are determined through learning by examples. The activation of these units is calculated by Eq. (1)

$$\tau_i(du_{i,t}/dt) = -u{+}_{i,t} \quad \sum_j w_{ij} x_{j,t} \qquad (1)$$

where $u_{i,t}$ is the membrane potential of each i-th neural unit at time step t and $x_{j,t}$ is the neural state of the i-th unit, and w_{ij} is synaptic weight from the j-th unit to the i-th unit. The time constant τ is defined as the decay rate of a unit's membrane potential. This decay rate might be considered to correspond to an integrating time window of the neurons, in the sense that the decay rate indicates the degree to which the earlier history of synaptic inputs affects the current state. Context units were divided into two units such as fast and slow units based on the value of time constant τ. The fast context units with small time constant (τ=4) whose activity changed quickly, whereas the slow context unit with a large time constant (τ=20) whose activity, in contrast, changed much more slowly. Among the input-output units, units corresponding to the proprioception and visual attention commands were not connected to each other. In addition, input units were also not directly connected to slow context units. Neurons in the CTRNN are modeled by a firing rate model, in which the activity of each unit constitutes an average firing rate over units of neurons. Continuous time characteristics of the model neurons are described by Eq. (2).

$$u_{i,t+1} = (1-1/{+}\tau_i)u_{i,t} \quad (1/\tau_i)(\sum_{j\in N} w_{ij} x_{j,t}) \qquad (2)$$

Actual updating of $u_{i,t}$ is computed according to Eq. (2), which is the numerical approximation of Eq. (1). The activation of the i-th unit at time t is determined by the following Eq. (3)

$$y_{i,t} = \begin{cases} \dfrac{\exp(u_{i,t})}{\sum_{j\in Z}\exp(u_{j,t})} & if\ i \in Z \\ f(u_{i,t}) & otherwise \end{cases} \qquad (3)$$

where Z is a set of output units corresponding to the proprioception or the vision sense. Softmax activation is applied only to output units except for the context units. Activation values of the context units are calculated by a conventional unipolar sigmoid function f.

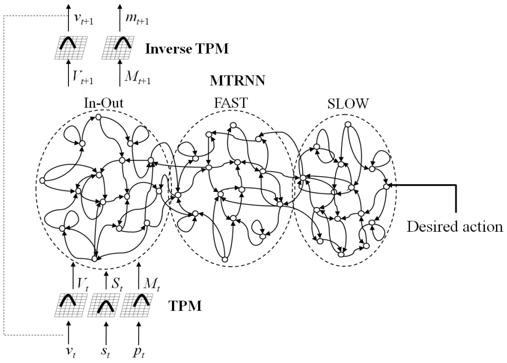


Figure 4. Architecture of MTRNN model with TPMs

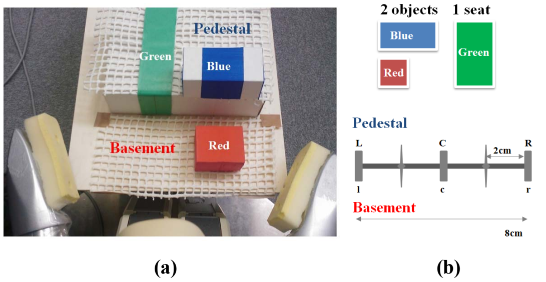
2) Additional training of novel sequences

During additional action training after the basic training for learning the primitive behaviors, only the connections of the slow dynamic units were allowed to change. The other weights were unconnected with the slow dynamic units fixed at values from the basic training. Initial states of slow dynamic units were set to different values from those of the basic action patterns. Slow dynamic units remembered sequences of primitive behaviors which were remembered by fast dynamics units. Therefore it was shown how the model can generate novel actions with object positions by using primitive actions.

III. Experiment and Results

A. *Task Design*

A small humanoid robot named HOAP3 was used in the role of a physical body interacting with actual environments. A table was set in front of the robot where a fixed pedestal attached with a green sheet was placed as shown in Fig. 5 (a).



(a) (b)

Figure 5. Workbench for robot experiments. (a) real environment of workbench, (b) the specification of work bench

The robot was supposed to displace two objects, 6x8x6 cm^3 red object and 6x10x6 cm^3 blue object between the basement and the green sheet on the pedestal. It was also supposed that the robot put one object on the other object which was placed on the basement. These sorts of level differences were introduced in the robot workspace because of the limited manipulation capability of the robot with multiple objects on a flat workspace. The position of the object in the basement as well as that of the green sheet on the pedestal can be varied with 8 cm range from left to right. In the current convention an object located in the first, the second, the third and the forth level denote that the object is on the basement, on an object placed on the basement, on the green sheet attached on the pedestal and on an object placed on the pedestal, respectively. It is also noted that the exact robot arm posture for holding these two objects are different because of the size difference.

The robot was initially trained for 3 types of basic displacement actions with all possible combinations for object position variations of left, center and right in the source and the destination. After the training of the basic action sets, the tests were conducted by a regeneration of them. Then, the experiment was further conducted for the additional learning where the robot was trained additionally for two more types of actions which share some primitive actions that appeared in the basic actions but combined in different ways in those actions. Only one teaching sequences was used to train for each task which is learned by back-propagation through time (BPTT) algorithm with 5x10^3 training epochs [8]. In this additional training, not all possible combinations of the position variations were completed but a part of them were trained. Also, only the synaptic weights in the slow dynamics part were modified with an expectation that a transfer of the skills preserved in the fast dynamics part could take place by means of generalization.

Table I shows the experimental conditions of the tasks. The basic action I is to displace a blue object placed on a green sheet, the basic action II is to move a blue object placed on the pedestal onto a red one on the basement and the basic action III is to move a red object located on the basement onto a blue one located on the pedestal. Training of each basic action accompanied with the visual attention shifts was repeated for 9 different positions under physical guidance of the arm movement trajectories by the experimenter. The additional action IV for the generalization test is to displace a red object located on a green sheet onto a blue object located on the basement. Although this action seems to share the similar motor profile with the basic action II, there is a slight difference in the arm posture for grasping different objects in the same position. This action was trained for 6 out of the 9 possible object position variations and the remaining 3 unlearned position variations were used for the generalization test. The additional action V is a sequential combination of the basic action II and action I in which the blue object located on the pedestal is moved onto the red object located on the basement and then moved back on to the green sheet located on the pedestal. This action was trained for 8 out of 27 object position variations and the remaining 19 unlearned position variations were used for generalization test. The time sequences of basic actions I, II, III and additional action IV are the same but the additional action V has 1.4 times of sequences of basic actions as shown by experimental trials.

TABLE I. Experimental conditions for robot tasks

Action Task	Origin : Object color	Destination : Object (seat) color	Transport direction (position level)
Basic I	Blue	Green	Down (2) to up (3)
Basic II	Blue	Red	Up (3) to down (2)
Basic III	Red	Blue	Down (1) to up (4)
Additional IV	Red	Blue	Up (3) to down (2)
Additional V	Blue	Red and then green	Up (3) to down (1) and then up (3)

B. *Results*

Each action was tested with every possible combination of object positions (left [L], center [C] and right [R]) in an origin and in a destination. If the goal-directed task is that the robot places the source object located in left basement to place on right destination located in right pedestal, this was called "action of [LR]". Performance was scored in terms of a success rate across all trials for each task. It was considered that a trial was successful if the object was successfully moved to the desired destination within the range of 2 cm.

1) *Basic actions I, II and III*

The robot could efficiently reproduce the whole learned basic actions through interacting with the environment. Whole basic actions are simultaneously generated by one network and this network had 0.003631 learning error between the teaching and output sequences, which is calculated by Kullback-Leibler divergence [8]. Additionally several trials were conducted for each action by placing the target object in arbitral points between the left and the right of trained positions. It turned out that the robot can perform the tasks successfully with more than 95 percentage success rate. This indicates that the robot achieved position generalization for the object to be manipulated via learning. Fig. 6 shows the result of the basic action II of [CC]. Here it can be observed that teaching signals are reconstructed in the generation only with minimal errors. It is not important to measure an error by the difference between the teaching signals and the generated signal, but within this experiment it was only considered whether an object was successfully moved to the desired destination within the range of 2 cm. For the analysis of the dynamic neural activities of an action, the PCA with the trajectories of the fast and slow dynamic units was applied. It can be seen that the profiles of fast dynamic units after the PCA contains more complex patterns as compared to those in the slow units as shown in Fig. 6. This might be because that the activities in the fast dynamics units are self-organized such that they are responsible for reconstructing details of sensory-motor profiles by utilizing their fast dynamics as have been shown in the prior studies [8].

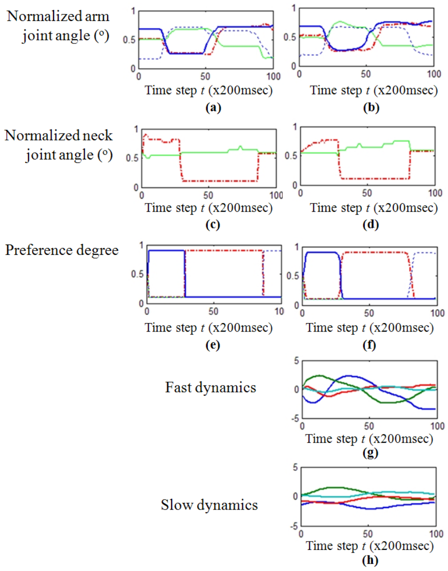


Figure 6. Examples of robot trials in basic action II of [CC]. (a), (c) and (e) are encoded arm joint angle (proprioception), encoded neck joint angle (vision sensation information), preference degree (visual attention command) of teaching signal, (b), (d), (f), (g) and (h) are actual sensory feedback value in physical environmnet during robot action. The activation profiles for the fast dynamics units (g) and the Slow dynamic units (h) using PCA with 1st PC axis to 4th PC axis, respectively. In proprioception, 4 values out of 8 normalized motor joint angles were plotted. In vision sense, bold dash-dot line represents normalized x-axis of neck joint angle and solid line represent normalized y-axis of neck joint angle. In visual attention command, bold, solid, bold dash-dot line and dash line are blue, green, red and default color effect.

Figs. 7 (a) and (b) represent the fast and slow dynamics activations, respectively, by PCA using 1st PC and 3rd PC to show the robot actions according to different object positions. Three different types of trajectories represent different object position cases. As shown in Fig. 7 (a), basic action I and III are similar patterns of the fast dynamics because these actions are similar in moving the objects in an upward direction. On the other hand, their slow dynamics patterns are different because their visual attention shift sequences are different. Finally, it can be observed that there are slight shifts for the trajectories within a limited range depending on the object positions for both of the fast and the slow dynamic units. This means that the internal neural activities successfully represent the action categories regardless of object position variance in each action.

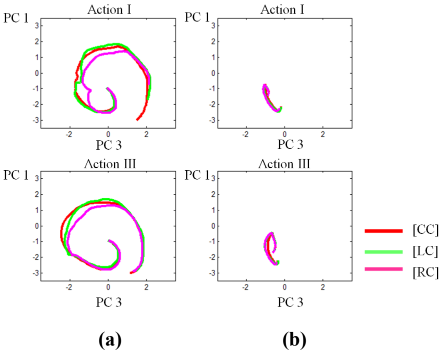


Figure 7. The trajectories of basic actions I and III by the fast and Slow dynamic units with PCA. (a) Projection results for fast dynamics unit activation with 1st PC axis and 3rd PC axis. (b) Projection results for slow dynamics unit activation with 1st PC axis and 3rd PC axis. [CC], [LC] and [RC] represent destination areas are equaly located in center but basement object is located in center, left and right, respectively.

2) *Additional action IV*

It was expected that the action IV could be learned with position generalization even with a partial set of training examples for position variances because this action is similar to the basic action II of previously acquired but with a target object of different size and color. It was found that the fast dynamics unit activation profiles are mostly similar between the two actions, those for the slow dynamics are different. This is because the slow dynamics part is responsible for generating different visual attention sequences. Also, it was found that all 3 untrained object position cases can be successfully generated along with 6 trained with 0.003689 learning error.

3) *Additional action V*

It was expected that this action could be learned with combination of two basic actions with novel object positions, which are composed by the basic action II and I of previously acquired. It was found that 15 untrained object position cases in additional training sessions can be successfully generated along with 8 trained cases with 0.003678 learning error. It was observed that 4 unsuccessful cases resulted from just slight position error in placing an object in the destination. Also, it was found that the visual attention command and hand postures were successfully generated by network considering top-down visual attention shifts for achieving the task. Even though the visual attention sequences skip to see the blue one in the mediate stage between basic action II and I, additional action V is successfully completed after learning the primitive actions.

IV. DISCUSSION AND CONCLUSION

A. *Different roles of two level networks with time constant*

Additional experiments showed that with each level adequate time constant in the network model is essential in successful developmental learning. This is because the two levels play different functional roles such as storing primitive behaviors in the lower level with fast dynamics and combining them in sequences guided by visual attention shifts in the higher level with the slow dynamics. By these means, new combinations of prior trained primitive behaviors can be achieved by modifying the connectivity only in the slow dynamic networks. Therefore, it is argued that the time constant difference can enforce segregation of functional roles in hierarchy via self-organization and that this gains generalization through developmental learning.

B. Future studies

Five kinds of future studies to construct autonomously operating robot considering active environments should be considered; (1) Adding the object recognition function, by utilizing texture, depth and appearance of object, to generate the complex top-down attention using high level visual cognition for achieving the high-level object manipulation task. (2) Introducing more diversity and complexity of actions rather than the current simple ones such as just placing objects. For example, a study should consider how robots can acquire skills for tool usage actions which would require more complex spatio-temporal association between visual attention and behavior generations. (3) Introducing a reinforcement learning paradigm in acquiring attention shift skills as inspired by Andrew's model. (4) Both additional action IV and V used partial set of training examples. The number of training examples in additional actions IV and V is an important factor to verify the generalization capability of dynamic networks that this experiment does not have. Therefore, future studies should look at the effect of the number of training examples for additional actions. (5) Also, future studies will include the investigation of the relationships between the learning models and literature on human's eye-hand coordination.

C. Concluding remarks

For achieving the visual-guided objects manipulation tasks in neuro-robotics via learning by examples, it was proposed that a new dynamic neural networks model in which visual attention shift and motor behaviors are associated in task specific manners by learning with self-organizing functional hierarchy required for a visual cognitive task. The proposed model can generate the goal-directed actions, such as basic and additional actions, for humanoid robot with regarding to the current sensory states including visual stimuli and body postures through interacting with real environments. Additionally, this has shown that the robot could successfully generate each basic action with placing a target object in arbitral object positions between the trained positions. This experiment has shown that there are the two kinds of generalization performances: (1) unknown object position and size can be successfully generated along with previously trained cases. (2) Novel action based on two basic actions can be generated by using previously trained primitive actions not only for the trained object position but also untrained cases.

ACKNOWLEDGMENT

This research was supported by the Original Technology Research Program for Brain Science (2010-0018845) (50%) and the Converging Research Center Program (50%) through the National Research Foundation (NRF) of Korea funded by the Ministry of Education, Science and Technology (2009-0082262).

REFERENCES

[1] R. S. Johansson, G. Westling, A. Ba¨ ckstro¨m, and J. R. Flanagan, "Eye-Hand Coordination in Object Manipulation," *J. of neuroscience*, vol. 21, pp. 6917-6932, 2001.

[2] M. C. Bowman, and R. S. Johannson, "Eye-hand coordination in a sequential target contract task," *Brain Res*, vol. 195, pp. 273-283, 2009.

[3] C. Prablanc, M. Desmurget, and H. Grea, "Neural control of on-line guidance of hand-reaching movements," *Prog. Brain Res.* vol. 142, pp. 155–170, 2003.

[4] M. B. Berkinblit, O. I. Fookson, B. Smetanin, S. V. Adamovich, and H. Poizner , "The interaction of visual and proprioceptive inputs in pointing to actual and remembered targets," *Exp. Brain Res.*, vol. 107, pp. 326–330, 1995.

[5] S. Jeong, S.-W. Ban, and M. Lee, "Stereo saliency map considering affective factors and selective motion analysis in a dynamic environment," *Neural Networks*, vol. 21, pp. 1420-1430, 2008.

[6] A. K. McCallum, "Learning to Use Selective Attention and Short-Term Memory in Sequential Tasks," Proceedings of the 4th *International Conference on Simulation of Adaptive Behavior*, pp. 315-324, 1996.

[7] J. D. Crawford, W. P. Medendorp, and J. J. Marotta, "Spatial Transformations for Eye-Hand Coordination," *J. Neurophysiol*, vol. 92, pp. 10-19, 2004.

[8] Y. Yamashita, and J. Tani, "Emergence of functional hierarchy in a multiple timescale neural network model: A humanoid robot experiment," *PLoS Biol.*, vol. 4, no. 11, pp. 1-18, 2008.

[9] R. Nishimoto, J. Tani, "Development of hierarchical structures for actions and motor imagery: a constructivist view from synthetic neuro-robotics study," *Psychological Research*, vol. 73, pp. 545-558, 2009.

[10] D. H. Ballard, M. M. Hayhoe, F. Li, S. D. Whitehead, J. P. Frisby, J. G. Taylor, and R. B. Fisher, "Hand-Eye Coordination during Sequential Tasks [and Discussion]," *Philosophical Transactions: Biological Sciences Royal Society*, vol. 337, pp. 331-339, 1992.

[11] D. H. Ballard, M. M. Hayhoe, and J. B. Pelz, "Memory representations in natural tasks," *J. of Cognitive Neuroscience*, vol. 7, pp. 66–80, 1995.

[12] J. B. Smeets, M. M. Hayhoe, and D. H. Ballard, "Goal-directed arm movements change eye-head coordination," *Exp Brain Res*, vol. 109, pp. 434–440, 1996.

[13] T. Grent-'t-Jong, and M.G. Woldorff, "Timing and sequence of brain activity in topdown control of visual–spatial attention," *PLoS Biol.*, vol. 5, e12, 2007.

[14] H. Tsubomi, T. Ikeda, T. Hanakawa, N. Hirose, H. Fukuyama, and N. Osaka, "Connectivity and signal intensity in the parieto-occipital cortex predicts top-down attentional effect in visual masking: An fMRI study based on individual differences," *NeuroImage*, vol. 45, pp. 587-597, 2009.

[15] M. Corbetta, and G. L. Shulman, "Control of goal-directed and stimulus-driven attention in the brain," *Nature reviews: neuroscience*, vol. 3, pp. 201-215, 2002.

[16] S. B. Choi, B. S. Jung, S. -W. Ban, H. Niitsuma, and M. Lee, "Biologically motivated vergence control system using human-like selective attention model," *Neurocomputing*, vol. 69, pp. 537-558, 2006.

[17] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transaction on Pattern Analysis and Machine Intelligence,* vol. 20, no. 11, pp. 1254-1259, 1998.

[18] S. J. Park, K. H. An, and M. Lee, "Saliency map model with adaptive masking based on independent component analysis," *Neurocomputing*, vol. 49, pp. 417-422, 2002.

[19] R. J. Williams, and D. Zipser, "A learning algorithm for continually running fully recurrent neural networks," *Neural Computation*, vol. 1, no. 2, pp. 270–280, 1989.

[20] K. Doya, and S. Yoshizawa, "Memorizing oscillatory patterns in the analog neuron network," *in proceedings of international joint conference on neural networks*, vol. I, pp. 27–32, 1989.

[21] J. Saarinen, and T. Kohonen, "Self-organized formation of colour maps in a model cortex," *Perception*, vol. 14, no. 6, pp. 711–719, 1985.

T. Kohonen, "Self-Organizing Maps," *Springer Series in Information Sciences,* vol. 30, 2001.