Proceedings of the 5th World Congress on Intelligent Control
and Automation, June 15-19, 2004, Hangzhou, P.R. China

第五届全球智能控制与自动化大会
2004 年 6 月 15 - 19 日，中国杭州

Developmentally Cognitive Robot Vision*

Tao Jiang and Fuchun Sun
State Key Lab. of Intelligent Technology and Systems
Tsinghua University
Beijing, China
{jt & sfc}@s1000e.cs.tsinghua.edu.cn

Xingzhou Jiang
Department of Weapon Engineering
Navy Engineering University
Wuhan, China

Abstract **– A generic methodology for designing developmental cognitive robot vision systems was proposed to add development capability to cognitive vision systems. Four issues were discussed: (1) human-like architecture suitable for different visual purposes, (2) efficient memory organization, forming, and indexing mechanism proper for emergence of machine imagery and experience-based learning, (3) life-like development model for open-ended cognitive improvement and perception development, and (4) engineering methodology to treat the robot intelligence embodiment as a life-like process combining evolution, development and learning mechanisms. Their possible solutions were also provided: (1) a so-called dual-visual system and dual-memory architecture framework, (2) an "interaction procedure recording" theory plus a new ecological neural evolution approach to solve the interaction and binding problem between dual visual systems, (3) a development model based on a novel hybrid intelligent system theory embedding memory and communication as its inner components, and (4) a four-phase robot life-span development schema for behavior engineering.**

Index Terms –Developmental cognition, cognitive agent, robot vision.

I. Introduction

Conventional robot vision systems can seldom provide opportunities for later development, e.g. adaptation to new tasks, after its designing phase has finished. Few do those issues are severely considered in current ongoing researches on cognitive vision systems (e.g. in the CogVis project, see [1]). The new approach proposed in this paper, named as developmentally cognitive robot vision (DCRV), tries to save robot vision systems designing from the dilemma (e.g. the preprogrammed way which pays more attention on task designing, environment symbolic based modeling and low-level human intuition/experience/knowledge based visual perception embodiment), instead by a computational development model which is inspired by some principles of human visual process and its cognitive development. The goal of this approach is not to design a robot vision system that can immediately be directly used into certain tasks, but instead to provide it development opportunities to let it become more and more competent for kinds of tasks gradually during training and application. Thanks to the added development mechanism, the novel kind of vision systems may become more competent in wider range of tasks, e.g. visual servo, navigation, exploration, tour-guidance and so on such services, by experiencing different development means as evolution, self-organization, structure development, supervised training, unsupervised leaning, communication based learning and co-evolution in one or several tasks context after it is designed or born.

A. Foundation of DCRV

Definition A DCRV system is a 3D active robotic vision system with cognitive capability and development mechanism.

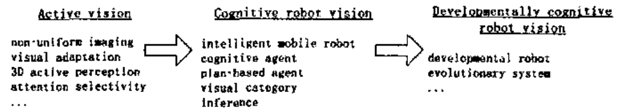

Fig. 1. The concept of DCRV is extension to cognitive vision system [1] by adding it development mechanisms and modifying its inner architecture for autonomous mobile robot applications. The cognitive robot vision system in turn is based on 3D active robotic vision [2].

Figure 1 illustrates the emergence of the concept about developmentally cognitive robot vision (DCRV). A DCRV system is an extension of a cognitive vision system by both adding developmental mechanisms into it and carefully redesigning the inner perceptual and cognitive architecture. The DCRV approach proposed in this paper mainly is based on such concepts: 3D active vision [2], cognitive agent for mobile robots (cognitive agent such as the belief-desire-intention one called as BDI agent [3] and the plan-based control agent [4], inference models such as qualitative inference [5]), and development mechanism (combining evolutionary, self-organization, and learning).

The motivation of a cognitive robot vision system is to provide semantic description and pragmatic evaluation from the robot's visual experience and then using them to benefit its action planning and decision-making. The planning and decision-making includes two parts: one is to control the action of the vision system; the other is to control the rest action systems of the agent. Two major elements of it are cognitive architecture and inference mechanism. Capabilities like visual categorization [6] and perceptual grounding [7] should be embodied into it. The aim of the active vision system is to provide visually perceptual representations (e.g. 3D map of its explored surrounding, visual feature set, etc.) for conceptual level manipulation by other later cognitive processes in the cognitive robot vision system. The perceptual

* This work is supported by the National Excellent Doctoral Dissertation Foundation (Grant No: 200041) of China.

0-7803-8273-0/04/$20.00 ©2004 IEEE

representation is provided by an active way, e.g. firstly by non-uniform imaging and adaptation coding, then modulated by attention-driven active perception.

A DCRV system owns a human-like basic architecture, and endures a life-like four-phase development phases including the pre-designing, postnatal development, training, application phases defined in the approach proposed in this work. It adopts different corresponding development mechanisms, e.g. evolution, pre-natal development, postnatal development, individual centered learning and cultural/communication based learning.

B. Relative research fields to the DCRV approach

Fundamental ideas of the DCRV approach and its corresponding methodology are encouraged by many meaningful works in an interdisciplinary manner: human vision research (esp. the dichotomy of "ventral stream" and "dorsal system" [8], visual action, visual attention [9] and active vision), developmental cognitive science (the Piaget school [10] and cognitive neuroscience), computation intelligence (evolutionary computing, artificial neural networks, fuzzy inference system, and hybrid intelligence system, i.e. HIS [11]), general systems theory and artificial life (selected-self organization theory [12], the "POEtic machine" owning an open-ended development potential in three levels--evolution, development, and learning [13]), artificial intelligent (BDI agent, plan-based agent, qualitative inference), developmental robots (e.g. DAY [14]), and empirical behavior engineering (e.g. robot shaping [15]).

C. Issues about DCRV discussed in this paper

In this paper, only the development mechanism and its related issues with the cognitive structure and the vision system architecture are discussed. These issues are considered as of most importance: the visual system *architecture* design, the *memory* organization and its indexing mechanism, the *development* process modelling, and the engineering *embodiment* methodology. The architecture of a robot vision system provides a physical frame to make "developmental cognition" possible. The memory indexing mechanism trying to interplay semantic/symbolic memory and episodic/implicit memory efficiently can creatively enable machine imagery and categorization possible, as well as may improve the efficiency and speed up cognition, e.g. accelerate the decision making in cases expected. The development mechanism can make the system become a more vital machine to learning on its experience including through communication. With the help of development mechanism, cognitive processes can make machine intelligence *emerge* through interacting with its external environment, other machines, and human beings.

II. System Architecture

To access those upper goals in this approach, a human-like and life-like methodology is adopted. It means a DCRV system has a human-like architecture and human-like perceptual, cognitive functionality, while owning adaptive potential in a life-like way. The systematical framework is required to provide comprehensive architecture, competent visual perceptual and cognitive functionalities, valid development mechanism based on experiences, and a practical engineering solution.

The architecture of a DCRV system is going to meet these means.

Be competent in different task contexts It means the visual system can continently process information for different vision-based behavior. Moreover, it should be adaptive in task context shifting. For example, the same vision system should be valid in two main categories of actions: the vision-based environment/self perception and vision-based motion servo.

Practically memory embedment The architecture design should provide efficient solutions to embed a memory system for required quite complex perceptual, cognitive process.

Embody perception and cognition The architecture should make visual perception and cognition work efficiently. So, a visual perception and cognition derived structure is necessary.

Work as a whole To a certain extent, modularity is the only efficient way in engineering. All the subsystems should interact, coordinate in a harmonious way. It requires the prefect designing of other modulation subsystems. They must integrate perceptual systems, cognitive systems, action systems, and memory systems coherently together.

A. Dual vision system

Keeping the basic idea in mind that there exists distinction of visual perception and visuomotor systems in visual cortex [8], the basic consideration for DCRV is also a similar dual-visual systems model as shown in figure 2.

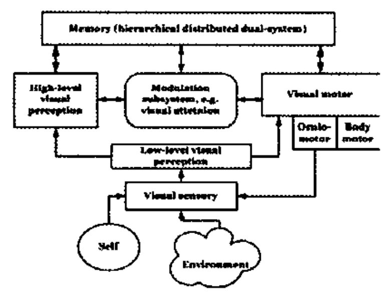

Fig. 2. The vision system consists of dual visual systems.

As shown in figure 2, like similar low-level visual information processing, a common low-level perception provides basic input for both the visual perception and visuo-motor systems. Interaction and coordination of the two visual systems in visual actions is on the shoulder of the modulation subsystem.

Visual attention selection and other purposive control of the visual system are mainly carried in the modulation subsystem interlinked with the memory system. Basic active vision mechanism, such as saliency representation, saccade pre-generation is performed by their collective actions. Other mechanisms as saccadic control, smooth pursuit, and gaze stabilization, vergence control is indeed executed by the oculomotor control system linked with the visuo-motor system under the command from the modulation subsystem.

B. Dual memory system

The memory system in a DCRV system has such properties: variety of contents, hierarchically distributed organization, multiple formation mechanism, indexing mechanism among implicit information and explicit information. Moreover, its formation is coupled with perception/cognition and development.

In visual information processing, information is manipulated into different levels from the implicit form to the explicit form. Like the human memory organization, a dual-memory schema is used in order to equip an efficient hierarchically distributed memory into the dual-visual system, and combine it with higher-level cognitive processes that are mainly based on the function of semantic memory. Figure 3 illustrates the basic frame of the visual memory system and its relation to the visual system.

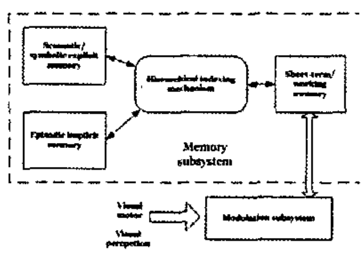

Fig. 3. Dual memory in the DRCV system.

Considering the HIS based implementation and the requirement of machine imagery/creativity, special bi-directional indexing mechanism is introduced into the memory organization. It lets the cognition processes to interpret semantic information (always coming from external sources, e.g. from communicating action) with/as information acquired from real experiences (e.g. from visual sensors with only early processing), which is always represented in implicit form.

Different machine learning methods are used in memory formation according to the source type, e.g. Hebbian leaning SOM (self-organized map) for implicit memory formation.

C. Solving the interaction problem

An *interaction procedure recording* theory is proposed to simplify the interacting dependence and the development of such relationship when designing coupled systems. Figure 4 illustrates how an "interaction procedure" may be retrieved (invoked) from memory to coordinate (modulate) the information processing behavior of the dual visual systems.

These procedures can be fulfilled and improved during life span. Some fundamental procedures are innate in the pre-designing phase. They can be used by the "invoke" process with the help of modulation subsystem. New "interaction procedure" emerging in latter life-span is coarsely recorded into the working memory, then can be monitored by the modulation subsystem, and "saved" into the "procedure memory". Alternatively, external or inexperienced "procedures" can enter into the "procedure memory" easily and then be "invoked" to modulate the two visual systems.

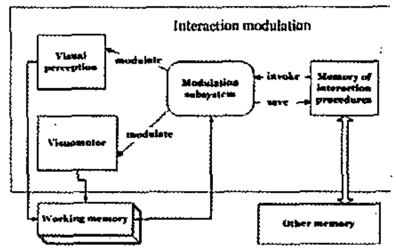

Fig. 4. Interaction procedure recording between dual visual systems.

To obtain a competent architecture for the visual system, basic ideas come from human visual systems. In order to get those fundamental interaction procedures, and improve them or make new interaction patterns emerging possibly in latter phases, a set of ecologically inspired evolutionary computation and neural evolution algorithms is proposed.

Based on the harmony nature of ecosystems, the ecologically inspired evolutionary computation (EEC) algorithm treats different coupled modes among subsystems co-existing in a whole complex system in a symbiosis, parasitism, or a combination of these styles according their different interdependence, while keeping the value of a whole system harmony function in a limited domain.

Different to general EC, EEC has such properties:

(1) Co-existing of more classes of species, which have different intelligence levels and vary population distribution, and interplay styles.

(2) Interplaying among different classes of species (symbiosis-like, parasitism-like, or combination of them) influencing the individual fitness, the population density and size.

(3) Adding a necessary energy function and its effect on each population to describe interdependence between species.

(4) The interplay mechanism between different classes adopts game-like models, [16].

(5) A harmoniousness of the ecosystem function innovated to evaluate the harmoniousness of the whole system and also as a necessary condition of the termination criterion.

Based on it, a novel ENE (EEC inspired neural evolution) approach is then proposed.

Multiple CNNs (constructive neural network, [17]) are used in one ENE. Each represents a subsystem. Adopting neural evolution method (NE, [18]) and the EEC algorithm in neural genome coding, two goals reach. One is the proper structure for each subsystem, e.g. the visual perception subsystem. It is achieved mainly due to the constructive process of CNN, and controlled by EEC. Another is the proper multi-interacting relationship among systems, also defined by EEC. So, by such steps of designing basic architecture as a human-like mode, of introducing "interaction procedure recording" theory on the life-span development framework, of implementing them with novel techniques as EEC and ENE (on the base of CNN, NE, EC), the architecture and its inner interaction or binding mechanism and its later development all seem reachable.

III. Development Mechanism

A. *Life-like development model*

Cognitive, perceptual capability should be treated as a development process in real robot vision applications, since the environment may be dynamic, the task context may shift. A life-like development model plays the central role in the DCRV system designing.

Development mechanisms design There are three levels of development means such as evolving, ontogenetic developing, and learning. All of them should be involved for obtaining more comprehensive development potential.

Developing along life span It means in different phases different development attention should play. For example, in postnatal phase perceptual development plays the central role by comparing to cognitive development.

Shaping basic cognitive elements Building the computational model of developmental cognitive processes involves basic cognitive elements identification, their coordination, and their improvement, i.e. the development mechanism designing.

Intact development Another issue is how to make the development process to apply different class of development means (e.g. postnatal development, imitation learning, vicarious learning, supervised training, and communication based evolving) into a smoothly integrated frame.

Embodiment Development model including basic perceptual and cognitive elements must be tightly coupled with memory designing, architecture designing.

B. *Development model based on a novel HIS theory*

Beyond cognitive development or improvement, the expected development model involves the phylogeny (evolution), ontogeny (individual development on the biological meaning), and epigenesist (learning on the stand of contortionists and logicians) behaviors. Moreover, it tries to combine them coherently. More importantly, these issues should be given under the background of robot vision system implementation, i.e. the architecture design, improvement and its visual intelligence (i.e. visual perception and visual cognition) development. Figure 4 gives the life-like development model.

A novel HIS theory is proposed to abridge the gap between development modeling and engineering embodiment.

Different to conventional HIS, [11], by adding roles of memory and (symbolic) communication, this approach can challenge more difficult tasks, e.g. amplify the forward knowledge extraction and backward knowledge mapping with the dual-memory retrieve and forming mechanisms, to study the communication ethological emergence, development by using the communication engine and the memory mechanism esp. again by its bidirectional indexing capability. Such a HIS theory approach is used to indirect designing each subsystem in the visual system and try to let them "as a whole", as once R. Brooks advocated [19].

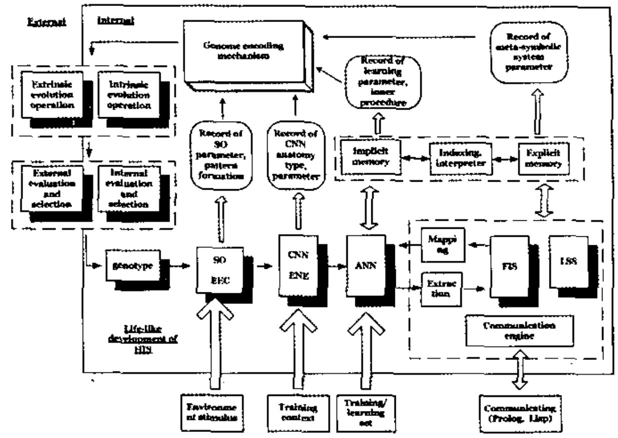

Fig. 5. The development model for DCRV is built based on the HIS theory.

As illustrated in figure 5, one genome records one kind of compressed symbolic description of four major classes of information that defines the development process of visual systems. They are: a. parameters set of EEC, b. parameters set of SO (self-organization) functions, c. type and parameters set of learning algorithms of CNNs, ANNs, ENE, and d. type and some meta-parameters of the FIS/LLS.

During system life span development, firstly, each genome is interpreted by a four-phase smooth embodiment process.

After enduring structure-to-function shaping and improvement based on its "anatomy structure and function", a "genome encoding mechanism" may update the previous genome using the information collected from different sources such as "record of SO parameter, CNN learning parameters", and so forth.

The development framework provides both intrinsic and extrinsic power for evaluating (i.e. fitness evaluation), evolutionary operating, and selecting the updated genome. By adding an internal evaluation and operation mechanism, the robot visual system may own an opportunity for self-adaptation. However, it may cause some risk in cases of real application. So such internal evaluation mechanism should be carefully designed and controlled during its validation. With the external mechanism, the robot will have another chance to experience and embody intelligence acquired by other visual systems or robots, i.e. by adding new fragment of novel SO parameters, CNN type, etc.

Other implicit knowledge and explicit rules may be either embodied in a pre-programming way or developed, obtained by communication-based learning or implicit learning process.

The extraction and mapping potential among vary stages in the HIS system and the special bi-directionally indexing and hierarchically forming mechanism in the embedded memory subsystem, make the latter way, i.e. communication-based learning or implicit learning, be possible and more efficient, since certain level of machine imaginary capability may be supported by the special memory system combined with the extraction and mapping process of the HIS.

IV. Engineering Embodiment

The most critical issue in engineering embodiment is how to *across the gap* between development modeling and engineering implementation. It implies a proper embodiment explanation for the development model is necessary, and must be designed on available current-level techniques. Meanwhile, since during different period of life span different development means may work dominantly, the engineering process should *smoothly merge* them along the machine life span.

As like human being, a robot vision system has four phases in its life span as mentioned in the introduction section. As shown in figure 6, the genome records a symbolic description of the four-phase development mechanism.

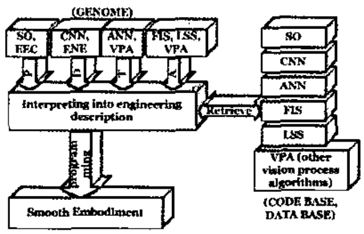

Fig. 6. The development schema is recommended to adopt in direct the process of engineering embodiment.

To embody it in engineering implementation, a group of corresponding explanation or description is needed. An engineering description is like a collection of methods, code that interpret each genotype "stored" in the genome.

Different development means apply in different phases, implicated by the development model based on HIS, i.e. *pre-designing phase* (Code base, data base, EC, EEC), *postnatal development phase* (CNN, ENE, SO), *training phase*, *application phase* (ANN learning, FIS, LSS communication).

Another critical issue is how to link smoothly the four embody phases together. Figure 6 illustrates the basic idea in the development schema. Different fragments of the genome are sequentially "interpreted" in related development phases, i.e. at the sequence from pre-designing, to postnatal development, to training, and to application.

V. Conclusion

The novel DCRV approach proposed in this work is going to provide a robot vision system such opportunities to process visual information in a more natural way, and improve its perceptual, cognitive intelligence during its life span combining several means: pre-born innate pre-programming, postnatal structure constructive developing, training by others esp. human operators, imitation learning from human and other machines, communication based ethological learning, co-development among machines and people. The proposed methodology is illustrated by these issues in this paper: the dual-visual system and dual-memory architecture, the "interaction procedure recording" theory, the development model based on a novel HIS theory, one class of novel approaches for natural intelligence-inspired algorithms as EEC, ENE, and the four-phase schema for engineering embodiment. They give reasonable solutions to implement DCRV systems in our future work. It shortens the distance between designing more generic, experience-based developmental robot vision systems and the practical engineering, although only using current available technology.

The methodology for DCRV systems designing is generic to other intelligent perceptual/cognitive systems development or simulation. It may also provide a generic framework for designing intelligent agents which are required to percept, to be cognitive in dynamic environment, to work in multiple contexts, and most importantly to develop in a life-like way: evolution, development, individual centered learning and cultural/communication based learning.

References

[1] http://cogvis.nada.kth.se/.

[2] Thierry Viéville, "A Few Steps toward 3D Active Vision," *Springer Series in Information Science*, 1996.

[3] Anand S. Rao and Michael P. Georgeff, "Modeling Rational Agents within a BDI-Architecture," *Proc. of the Second International Conference on Principles of Knowledge Representation and Reasoning*, CA, 1991.

[4] Mitchael Beetzf, "Plan-Based Control of Robotic Agents: Improving the Capabilities of Autonomous Robots," *Springer LNAI 2554*, 2002.

[5] A. G. Cohn, D. R. Magee, A. Galata, D. C. Hogg, S. M. Hazarika, "Towards and Architecture for Cogntive Vision using Qualitative Spatio-Temporal Representations and Abduction," *Proc. Spatial Cognition 2002*, Tutzing Castle, Lake Starnberg, May 2002.

[6] Nick Barnesa, Zhi-Qiang Liu, "Embodied Categorisation for Vision-Guided Mobile Robots," *Pattern Recognition*, vol. 37, pp. 299–312, 2004.

[7] Steels, L., "The origins of intelligence," *Proceedings of the Carlo Erba Foundation Conference on Artificial Life*, Fondazione Carlo Erba. Milano, 1996.

[8] Shaun P. Vecera, "Dissociating 'What' and 'How' in Visual Form Agnosia: a Computational Investigation," *Neuropsychologia*, 2001.

[9] David LaBerge, "Defining Awareness by the Triangular Circuit Of Attention," *PSYCHE*, vol. 7, June, 1998.

[10] Jean Piaget, "The Origins of Intelligence in Children," International Universities Press, 1952.

[11] Ajith Abraham and Baikunth Nath, "Hybrid Intelligent Systems Design: a Review of a Decade Research," *IEEE tans. SMC*, 2001

[12] Luis Mateus Rocha, "Selected Self-Organization and the Semiotics of Evolutionary Systems," in *Evolutionary Systems: Biological Perspectives on Selection and Self-organization*, S. Salthe, G. Van de Vijver, and M.Delpos (Eds.), Kluwer Academic Publishers, pp. 341-358, 1998

[13] Christof Teuscher, "On the State of the Art of POEtic Machines," *Technical Report 01/375, Swiss Federal Institute of Technology Lausanne, CH-1015*, November, 2001.

[14] http://www.cse.msu.edu/~weng/research/LM.html.

[15] Marco Dorigo and Marco Colombetti, "Robot Shaping: an Experiment in Behavior Enginnering," MIT press, 1998.

[16] F. Renaud and T. de Meeus, "A Simple Model of Host-Parasite Evolutionary Relationships," *J. theor. Biol*, vol. 152, pp. 319-327, 1991

[17] Rajesh Praekh, Jihoon Yang, and Vasant Honavar, "Constructive Neural-Network Learning Algorithms for Pattern Classification," *IEEE Transaction on Neural Networks*, vol. 11, No.2, pp. 436-451, 2000.

[18] Xin Yao, "Evolving Artificial Neural Network," *Proceeding of the IEEE*, vol. 87, No. 9, pp. 1423-1447, 1999.

[19] Rodney A. Brooks, "Steps toward Living Machines," in *Evolutionary Robotics, LNCS 2217*, Springer, 2002.