

Proceeding Paper Seeing like a Human: The Embodied Development of Active Vision in an Artificial General Intelligence System ⁺

Kai Liu^{1,2,*} and Changxin Sun¹

- ¹ Department of Educational Science, Bohai University, Jinzhou 121013, China; cxin2027@163.com
- ² Institute of Artificial General Intelligence, Bohai University, Jinzhou 121013, China
- * Correspondence: ccnulk@ccnu.edu.cn
- Presented at the 2023 Summit of the International Society for the Study of Information (IS4SI 2023), Beijing, China, 14–16 August 2023.

Abstract: In the field of computer vision, active vision systems based on Artificial General Intelligence have shown stronger adaptability compared to passive vision systems based on Special-purpose Artificial Intelligence when dealing with complex and general scenarios. The development of visual capabilities in intelligent agents follows stages similar to human infants, gradually progressing over time. It is also important to set appropriate goal systems for educational guidance. This paper presents preliminary research on the developmental overview of visual depth perception abilities using OpenNARS, an Artificial General Intelligence system equipped with visual perception and motion channels. This study reveals that the development of visual abilities is not based on a mathematical algorithm but rather an educational guidance of embodied experiences construction.

Keywords: artificial intelligence; active vision; artificial general intelligence

1. Introduction

In recent years, computer vision (CV) has made great progress thanks to the rapid development of deep learning in object recognition and image segmentation. However, such passive vision models are trained using Special-purpose Artificial Intelligence (SAI), which cannot escape from the intrinsic limitations of SAI, such as poor contextual understanding, lack of adaptability due to its offline experiences, and lack of flexible attention in a rigid target system. In contrast, active vision has emphasized the active nature of visual behavior from the beginning, and related work has made some progress. Unfortunately, the subjectivity and autonomy of active vision systems are still lacking, and active vision models are still essentially deformations of SAI systems. In this paper, we explore the developmental approaches of active vision based on sensation and motion and investigate the potential association between visual ability and subject experience, with an Artificial General Intelligence (AGI) system as the agent, under an active perspective, to adapt to its environment autonomously.

2. Active Vision Overview

2.1. Bionic Perspectives

From the bionic perspective, active vision is closely related to the complexity of a creature's visual ability. Frogs have relatively simple visual neural pathways, and the frog visual system is a common biological model. Although computer simulations of frog visual systems can be performed, the simulated system is hardly capable of handling the demands of complex visual tasks. Studies of higher organisms such as cats, rats, and humans, however, have struggled to achieve comprehensive and accurate simulations due to the complexity of the neural pathways. Such systems are mostly specialized vision systems that incorporate the biological properties of the eye. Although they can be used in different scenarios, their actual effectiveness is usually moderate.



Citation: Liu, K.; Sun, C. Seeing like a Human: The Embodied Development of Active Vision in an Artificial General Intelligence System. *Comput. Sci. Math. Forum* **2023**, *8*, 44. https://doi.org/10.3390/ cmsf2023008044

Academic Editors: Zhongzhi Shi and Wolfgang Hofkirchner

Published: 11 August 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/).

2.2. Computational Perspectives

In the computational model of active vision, deep reinforcement learning serves as the core technical framework of the model and undertakes important tasks such as prediction as well as recognition. Traditional 3D reconstruction techniques that were used in the past are gradually being replaced by SLAM (Simultaneous Localization and Mapping) techniques incorporating deep learning and point clouds [1]. While these active vision systems show impressive performance, they often lack interpretability and have lower credibility.

2.3. Developmental Perspectives

Researchers who advocate for a developmental perspective argue that computers are essentially blank slates and should learn through development and evolution, with the fusion of sensation and action being key [2]. Instead of solely relying on explicit visual representations for computation, it is more beneficial to enhance the system's inherent ability for perceptual-motion coordination. As active vision has evolved, it has become increasingly important to study the visual developmental issues in environmental adaptation, focusing on active agents equipped with visual sensors that can learn and grow [2,3].

3. Development of AGI Active Vision

For the visual agent, there is no theoretical difference between the visual nature of 2D and 3D environments. Therefore, in this study, we used the AGI system OpenNARS as the agent and developed a 2D grid-based virtual survival environment (as shown in Figure 1). In this virtual environment, OpenNARS can receive signals of different utilities from the environment and can also issue motion control commands to actively modify sensory signals. When the OpenNARS system is initially started, it does not have any pre-existing innate visual experiences. Through appropriate postnatal educational guidance, OpenNARS is aided in developing experiential concepts related to vision based on perceptual-motion foundations, gradually improving the system's visual capabilities [4,5].



Figure 1. This figure shows the virtual enviroment of training Active Vision of AGI system. Note that the "red triangle" is controlled by OpenNARS, "blue squares" are positive stimuli and "cyan squares" are negative stimuli.

The visible region received by OpenNARS spans three grid units in front of it, then five grid units in the further row, and so on. Since visual processing does not generate from a "God's perspective" but rather from a one-dimensional "Subjective perspective", placing objects at different distances can be used to test NARS's visual depth perception. This study emphasizes that, similar to the visual development of human infants, the active visual development of AGI systems also goes through multiple stages, such as "stationary agent, solid object moving back and forth", "agent moving back and forth, solid object stationary or moving", and "stationary agent, multicolored multiple objects occluding from the left and right". Experimental results show that the system's visual depth perception improves as the number of concept network nodes increases. The slow progress in the maturity of visual depth perception due to innate target settings significantly affects the system, demonstrating the unity of active vision's embodiment, constructiveness, and autonomy.

3.1. Visual Depth Perception of the "Object-Moving & Agent-Not Moving"

This is the first stage of visual depth development, characterized by a stationary agent and a solid object moving back and forth. As shown in Figure 2a, the intelligent agent remains in a fixed position in the environment, while the position of the block-shaped object changes either by moving left and right or moving forward and backward. As a result, the visual imaging of the object perceived by the intelligent agent varies in size due to the movement of the object. The intelligent agent learns to judge the distance between the object and itself based on these visual cues.



Figure 2. This figure shows the four stages of the development of Active Vision. (a) Scenario with active object with stationary agent; (b) active agent with stationary object; (c) left or right shading; (d) front or back shading.

3.2. Visual Depth Perception of the "Agent-Moving & Object-Not Moving"

This is the second stage of visual depth perception development, characterized by a stationary and monochromatic object and an agent moving back and forth. As shown in Figure 2b, the intelligent agent can move freely in the environment while the position of the blocking object remains fixed. Therefore, the visual presentation of the object to the intelligent agent changes in size according to the distance between the agent itself and the object. The intelligent agent learns to judge the distance between the object and itself from these variations.

3.3. Visual Depth Perception with Left or Right Shading

This is the third stage of visual depth perception development, characterized by a stationary agent and the presence of two multicolored objects that occlude each other from the left and right sides. As shown in Figure 2c, the intelligent agent is not able to move freely, while the objects form an occlusion relationship. The agent itself needs to learn to judge the depth perception in this scenario.

3.4. Visual Depth Perception with Front or Back Shading

This is the fourth stage of visual depth perception development, characterized by an agent's movement and the presence of two multicolored objects that occlude each other from the front and back. As shown in Figure 2d, the intelligent agent can move freely from left to right, while the object will form different angles of occlusion due to the agent's movement. The agent needs to learn to judge the depth perception in this scenario on its own.

4. Prospects

For diversiform cognitive systems, such as human or OpenNARS, they can "see" in a same way theoretically behind the different physical compositions. This experiment primarily focuses on the study of the active visual depth comprehension of intelligent agents and does not comprehensively explore the capability of active vision. The developmental patterns of visual elements, such as color and shape, will be gradually explored in subsequent studies.

Author Contributions: Conceptualization, K.L.; methodology, K.L.; software, C.S.; writing—original draft preparation, C.S.; writing—review and editing, K.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Kim, J.; Jeon, M.H.; Cho, Y.; Kim, A. Dark Synthetic Vision: Lightweight Active Vision to Navigate in the Dark. *IEEE Robot. Autom.* Lett. 2020, 6, 143–150. [CrossRef]
- 2. Bajcsy, R.; Aloimonos, Y.; Tsotsos, J.K. Revisiting active perception. Auton. Robot. 2018, 42, 177–196. [CrossRef] [PubMed]
- 3. Liu, K. The Seeing of Machine Infants: A New Approach to Interdisciplinary Theory of Active Vision. Available online: https://www.koushare.com/video/videodetail/54809 (accessed on 20 April 2022).
- Wang, P.; Hammer, P. Perception from an AGI perspective. In Proceedings of the Artificial General Intelligence: 11th International Conference, AGI 2018, Prague, Czech Republic, 22–25 August 2018.
- Liu, K. The Dual Paradigm Shift in the Two-way Integration of Artificial Intelligence and Education. Open Educ. Res. 2023, 29, 4–18.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.