Visual Perception in Intelligent Virtual Agents in Augmented / Virtual Reality

Punit Singh - 200083296
MSc Artificial Intelligence
Aston University

Abstract

This article discusses and reviews the research done in the design and development of virtual agents with a standalone integrated perception systems independent of the environmental and external factors — software and/or hardware — with a major focus on visual perception, and in that too, active visual perception rather than passive visual perception. Most of the agents, methods and techniques discussed in this paper are virtual or exist in a virtual environment. Previous research has been studied, critiqued, further thought upon and ultimately connected leading to a thought experiment that puts into perspective how giving power of sight to intelligent virtual agents affects their behaviour compared to agents that do not "see".
Keywords: ***Artificial Intelligence, Agent, Perception, Virtual, Vision***

1 Introduction

A picture is worth a thousand words makes much more sense when we compare the capabilities of Artificial Intelligence in different tasks. On one hand, we have AI and bots that can play chess and perform other strategic tasks very efficiently and one might even compare them to be on a human level while on the other hand we have AI which struggles to even differentiate a cat from a dog. Progress in AI has, since inception, always been uneven as noted by (Burton, 1993). It is quite observable that since the early stages of Artificial Intelligence, the progress of computer vision and perception is unimpressive because if we come to think of it, computers do not "see" anything in the literal sense. Even in the most complex Convolutional Networks, computers read images as arrays and matrices filled with numbers and that data is all that an AI can "see". Now, ultimately they are only computers —literal rocks that we are trying to teach how to think— and it should be obvious that they will not ever have the same perception as us humans do. Sure they can have much better senses in the future and can even have extremely complex and eerily similar imitations of human senses, but they can not be the same.

This article will focus on the existence of virtual agents in an augmented reality environment and the incorporation of perception systems using software and hardware methodologies. It will be more centered on visual perception rather than auditory or tactile perception because there is already limited research in virtual agent perception and even in that, a major part of the research is biased towards visual perception. Visual perception in virtual agents can be split into two types, active perception and passive perception. Although passive perception will be discussed later, the discussion will be rather short. For the sake of this article, I will limit the discussion to active perception systems.

Instead of calling these systems "active visual perception in intelligent virtual agents", I prefer calling them "Software Vision" which I understand as an umbrella term encompassing all the different types of mechanisms serving the common goal of providing visual perception to intelligent agents without external input to the agent via environment and/or hardware means.

1.1 Active Perception

(Bajcsy, 1988) defined active perception as a data acquisition problem in an intelligent way where systems proactively take steps to gather more information about their environment. The paper explores both, a top down approach based on a task or a query from a database and a bottom-up approach where the task is initialized without any database query and with the aim of exploring the environment. In the first step they gather either a geometric skeleton of the scene or a set of differentiating factors to be searched for followed by a search operation in the database enabled by a decision making algorithm which is a primary goal of the active vision systems. In Figure 1, we can see the field of view of the fish where because of the object occlusion, the fish can only see the fish on the left side while the fish on the right is hidden behind the spherical object. The center of the perspective is the eye position of the fish and it cannot see behind itself. Also, the fish cannot see the fish right in the front because the range of its vision does not reach that length.

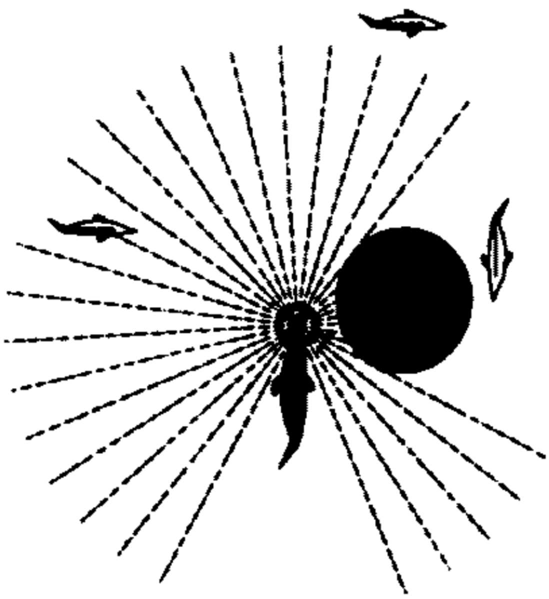

Figure 1: A representation of how active perception looks in effect. (Terzopoulos and Rabie, 1995)

2 Why Perception Matters

The argue point of this article is the same that (Burton, 1993) puts forward, that true knowledge can not exist without perceiving the environment around us independently and the same applies to Artificial Intelligence. The Molyneux problem (Davis, 1960) is a classic analogy to be given here; assume a visually impaired person, who has learned to distinguish by touch, different shapes, say, a cube and a sphere is suddenly provided with an ability to see, will the person be able to distinguish between the same sphere and the cube only by seeing both of them placed on a table in front of them and not touching them. The answer to this problem was given by (Held, 2011) when their study showed that the newly sighted subjects (previously blind people who had their vision restored by medical procedures) were not able to differentiate between different shapes based only on visual input. In the case of AI, it doesn't matter how better AI get at performing tasks, if it gets better only at performing tasks, we cannot call it truly intelligent. (Burton, 1993) gives a good example that a chess playing AI does not know that it is playing chess, for that matter, it does not know what chess is, just like a computer does not know that it is a computer. You have to program a computer for it to function properly and you have to train an AI for it to behave intelligently.

Human intelligence and artificial intelligence differ in the sense that humans do not receive input from the environment passively, we actively seek new sensory inputs form the environment and that is how we learn. (Kozma, 2007). Artificial Intelligence on the other hand is more goal oriented in the sense that if the goals that the AI is defined to accomplish, are being reached , the AI will not try to learn anything new or will not even explore the environment it is placed in and that might have something to do with the absence of perceptive "organs".

In support of the argument that AI performs better and more human-like when given perceptive capabilities, even if limited, the recent paper by OpenAI (Baker et al., 2019) shows quite interesting behaviour by the agents. In this paper, in the hide and seek task, seeking agents were given a line of sight perception and the hiding agents had to avoid being in the line of sight of the seekers. The agents continued to find new and efficient ways to win the rounds ranging from the seekers using ramps to go over walls and the hiders locking the ramps to the seekers finding bugs in the physics simulation of the environment to catapult themselves over the walls into the hiders' hideout. In a different environment, the agents discovered that they can "box-surf" because of a bug. The behaviour that these agents showed could be considered intelligent or at least more human because it involves actively seeking solutions to problems that the opposing team causes. But then again, these agents are also goal oriented and once their objective was being completed, they brute forced the same solution until met with another problem and did not try to improve upon existing solutions and did not even try to find different solutions.

3 Limits of Virtual Perception

When I say perception in virtual agents, I do not mean it in the biological sense where us humans design and develop complete organs for virtual agents in software because that will require a complete understanding of how our those organs and own brain works. What I really mean is, putting in place a system, hardware or software, for helping virtual agents experience the environment they exist in, instead of a wall of numbers. This article will focus on the software approaches to adding perception to virtual agents. Hardware approaches are those where we add sensors to the environment and integrate the inputs from the sensors with the agent actuators using an interface or a framework and software approaches are where we try to replicate the human/animal eyes using virtual sensors. This approach is called Animat Vision by (Rabie and Terzopoulos, 2001) and in the paper it was applied on complex agents having high mobility along with the

ability to control their eyes that had binocular vision. In another paper, (Rabie and Terzopoulos, 2000) applied the active vision to human like virtual character DI-Guy developed by Boston Dynamics Inc.
Although these systems appear to be life-like, there are a few limits that need to be respected. The researchers and the users have to keep in mind that these agents and their perceptive systems are completely virtual and are completely limited to the infrastructure set up by us humans. If the agent simulations are being ran on a computer processing at its maximum capacity, how so ever intelligent the agents might become, they cannot enter the physical world and upgrade the physical computer. In the same way that if the perceptive system is set up using some specialized hardware like LiDaR (Collis, 1970), the perceptive system of the agents will be limited by the limitations of LiDaR. Unless coupled with other hardware, the agent will only see a gray-scale environment without any colours and will not be able to see through fog and rain or even detect less reflective surfaces. For software perception models like in (Rabie and Terzopoulos, 2000), unless the agents learn to modify their own source code, they will be stuck with colour histogram methods and virtual eyes for vision that have limited functions, and when the agents learn to modify their own source code to add functionality, if that is not sentience, I don't know what is.

An interesting framework, proposed by (Zhang et al., 2018) also reinforces the concept that entities — agents, or objects — that exist in the real world, can exist in the virtual world by being scanned but entities that exist in the virtual world can not exist in the real world.

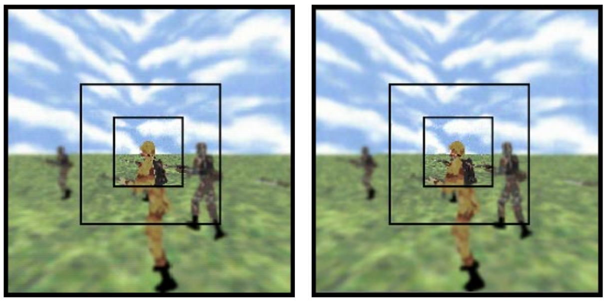

Figure 2: Binocular vision applied to the DI-Guy (Rabie and Terzopoulos, 2000)

Figure 3: Target of the DI-Guy as seen form the agents' perspective. (Rabie and Terzopoulos, 2000)

4 Related Literature

4.1 Animats

To begin with, (Meyer and Wilson, 1991) defined *animats*: machines in the form of animals placed in their natural habitats complete with actuators and complex muscle controls capable of independent movement in the environment and interacting with other animats present in the same environment. There has been great success in designing animats that can navigate complex environments. Animats are designed using simple rules of locomotion and possess superior sensorimotor control along with more information available than normal animals do like access to absolute position with respect to their environment etc. and as they state in the paper that in order to attain simple control mechanisms, the proprioceptive systems of these agents must be efficient.

4.2 Active Vision Systems

Animat fish were used by (Terzopoulos and Rabie, 1995) in an attempt to give them active vision without the use of any specialized hardware that might aid the fish in gathering information of their environment. Computer vision algorithms were applied on top of their virtual retinas. Some fish in the environment employed perception oracle as their vision system where they got the local, geometric, chromatic and every other form of information available to the rendering engine by directly integrating with the environment in a meta way.

They gave binocular vision to the animat fish such that the eyes were controlled by two "gaze angles," one for the horizontal rotation and one for the vertical rotation (θ, ϕ). The system was setup such that when the eyes are looking forward, the gaze angles are $0°$, $(\theta = \phi = 0°)$. The field of view also plays an important role when designing vision systems for agents. Figure 4 shows the binocular perspective of the artificial fish.

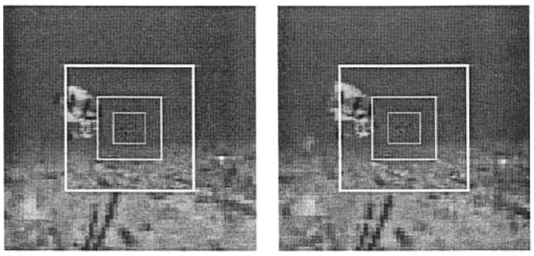

Figure 4: Binocular vision in animat fish. (Rabie and Terzopoulos, 2001)

Figure 2 shows the binocular perspective when applied to humanoid agent DI-Guy and Figure 3 shows how the DI-Guy agent sees the enemy.

The retinal imaging system uses colour modelling of certain objects that are to be required to be recognized by the animat, for example, a fish interested in mating will have models trained on colours similar to its own so that when it sees another animat fish similar to itself, it will start the mating ritual in an attempt to attract the other animat. The 2D retinas developed for animat fish made use of an improved colour histogram intersection method for recognizing colour blobs.

To build further upon this model, they (Rabie and Terzopoulos, 2001) added the stereo vision to the animats where they calculated the depth of the environment and the object in focus using the differences in the focal points of individual eye models.

Although virtual agents can use virtual sensors and actuators to perform tasks, there exist some limitations which are highlighted when the need for dynamic knowledge of the environment rises and to address that issue, (Conde and Thalmann, 2006) proposed a model of perception system for virtual agents that incorporated three techniques: active vision, proprioception and predictive vision in one framework. In order to make the agents self sustaining and autonomous, Reinforcement Learning algorithm, Q-Learning was integrated to the proprioception model. They realized this method and applied it on a virtual goalkeeper with an aim to prevent a football from entering the goal. The results showed that using this ALifeE framework, the goalkeeper successfully prevented the ball from entering the goal with input data only from its virtual "eyes".

The speed at which AI is improving, there could soon be an AI which has its own complete mind and is capable of self sustenance without any or at most, minimal human intervention (Minsky, 1988). As humans, the most humane decision that we can make to help a fully developed AI would be to give it its own world where it can thrive as an intelligent organism or even as an intelligent species. Imagine the concept of a parallel world like a parallel universe in physics theories, but here the parallel world in question is one that we have made for our intelligent agents where they can exist as we do in the physical world along with an intersection between the real and the virtual world where the virtual agents and humans can co-exist. (Zhang et al., 2018) proposed an Inverse Augmented Reality (IAR) framework where intelligent agents exist in a virtual world as humans in the physical world. The IAR framework is more agent centred instead of human centred in the sense that it focuses more on what real objects do the agents "see" instead of what virtual objects us humans see.

4.3 Passive Vision Systems

Among hardware systems that add visual perception to computers or virtual agents like cameras, LiDAR sensors, depth sensors or proximity sensors, there exists a perceptual oracle or oracle vision system that does not require any hardware. In perceptual oracle systems, agents in a virtual environment access information about the environment and the agents' locality by sending queries to the environment itself and get access to highly processed and privileged information which is not available to animats otherwise like if an animat is in a swarm and possesses perceptive oracle system, it can access the number of animats in a swarm or the absolute location of the animat with respect to the environment along with its relative position. This can cause a hive mind like situation where each animat is fed the same information and the whole herd is led to a unanimous goal rather than individualistic actions and reflexes.

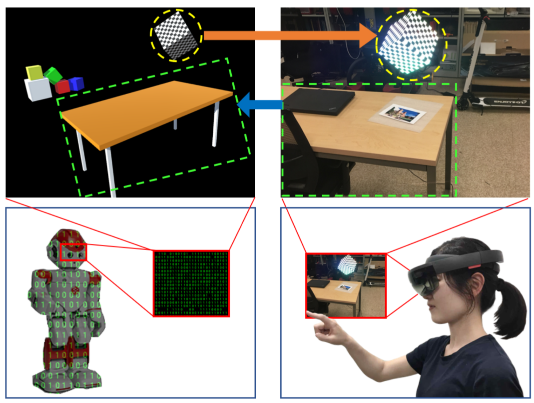

Figure 5: Representation of the Inverse Augmented Reality (Zhang et al., 2018)

5 Discussion and a Thought Experiment

Now that the different ideas put forward by other researchers have been laid out, in this section I would like to propose a thought experiment to put into perspective why perception, in general, let alone visual is important and why more research should be focused on integrating perception systems to intelligent virtual agents.

Imagine a virtual world applying the concept of either dual reality proposed by (Lifton and Paradiso, 2009) or one reality by (Roo and Hachet, 2017). Integrate into it artificial intelligence to make it an intelligent virtual environment as applied by (Luck and Aylett, 2000). Now, make the virtual world you have imagined evolve by using self learning mechanisms or evolutionary algorithms like the concept discussed by (Taylor, 2013). Put into the environment an animat, an intelligent animal like fish if the environment is surrealistic (underwater) or a humanoid animat and give it active perception like given to the DI-Guy in (Rabie and Terzopoulos, 2000). Now if we give the agent reproductive and evolutionary capabilities converting it into a multi agent environment (Steel et al., 2010), we get **multiple intelligent virtual agents in an intelligent evolving environment, each capable of independent locomotion and equipped with "software vision".** What outcome can be expected from this virtual world? Will the agents and species die out after a few million time steps or will they continue to evolve and develop to the point where they surpass human intelligence and become sentient, developing learning methodologies and behaviours not initially programmed into them? Will we be able to create a virtual world that resembles the real one as closely as possible limited only by our own understanding of physical laws of the real?

To further your intuition, Imagine the same world as in the first case, with the same agents and the same physics, only in this case, instead of giving the agents software vision systems, provide them with passive perception where the agents are fed information about their position and the surrounding directly by the environment. How will this world differ from the world with agents that have software vision and in what respects will both of these worlds differ from the real world.

This experiment seems realizable as most of its components have already been worked on separately. If successfully integrated, a whole new world of possibilities will open up and since all of this is happening in a virtual environment accessible by Head Mounted Devices (HMDs), it will open up a lot of possibilities for Human Computer Interaction research. Having access to an intelligent parallel world might also help us better understand our own world and because the simulation will be running on servers, the speed of the simulation will be corresponding to the clock speed of the computers running the simulation so we will be in control of the speed at which the simulation(s) run, speeding it up during the *uninteresting* times in the evolution of the world and slowing it down when we need to closely observe a particular behaviour or phenomena. Software vision possessing intelligent agents which learn dynamically can also aid in training military combat training although they will require visual as well as an auditory perception discussed by (Herrero and de Antonio, 2003).

6 conclusion

The possibility of the creation of intelligent virtual agents equipped with localised visual perception or "eyes" as software alternatives to the hardware based approaches might not have been possible in the early 2000's when there was more focus on the research, but development was limited by either the hardware capabilities of the time or because there was not enough commercial viability on augmented and virtual reality or both, but with web 3.0 emerging and 5G network being available to end users, corporations are focusing more on virtual and augmented reality and trying to incorporate into it Artificial Intelligence and Machine Learning methods. This article reviewed the research done in trying to give vision to virtual agents without the use of external hardware or software. We also discussed the concept of Inverse Augmented Reality where the world in existence is viewed from the perspective of a virtual agent and in doing so, the limits and scope of developing artificial vision were laid out and a few dots were connected which led to the formation of a thought experiment in which two intelligent virtual worlds were thought upon. It now remains a matter of time until such a system is materialized and once it does, what we have to see is, **will we learn from how the virtual world evolves or will the virtual world learn form us?**

References

R. G. Burton, *Natural and Artificial Minds*, 1st ed. SUNY Press, 1993.

R. Bajcsy, "Active perception," *Proceedings of the IEEE*, vol. 76, no. 8, pp. 966–1005, 1988.

D. Terzopoulos and T. F. Rabie, "Animat vision: Active vision in artificial animals," in *Proceedings of IEEE International Conference on Computer Vision.* IEEE, 1995, pp. 801–808.

J. W. Davis, "The molyneux problem," *Journal of the History of Ideas*, vol. 21, no. 3, pp. 392–408, 1960. [Online]. Available: http://www.jstor.org/stable/2708144

R. Held, "The newly sighted fail to match seen with felt," 2011.

R. Kozma, "Computational aspects of cognition and consciousness in intelligent devices," *IEEE Computational Intelligence Magazine*, vol. 2, no. 3, pp. 53–64, 2007.

B. Baker, I. Kanitscheider, T. Markov, Y. Wu, G. Powell, B. McGrew, and I. Mordatch, “Emergent tool use from multi-agent autocurricula,” *arXiv preprint arXiv:1909.07528*, 2019.

T. F. Rabie and D. Terzopoulos, “Modelling active vision systems for dynamic simulated environments,” in *IASTED International Symposium on Modelling and Simulation (MS'2001)*, 2001, pp. 34–41.

——, “Active perception in virtual humans,” in *Vision Interface*, vol. 2000, 2000.

R. Collis, “Lidar,” *Applied optics*, vol. 9, no. 8, pp. 1782–1788, 1970.

Z. Zhang, D. Weng, H. Jiang, Y. Liu, and Y. Wang, “Inverse augmented reality: A virtual agent's perspective,” in *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE, 2018, pp. 154–157.

J.-A. Meyer and S. W. Wilson, “From animals to animats,” *Cambridge MA*, 1991.

T. Conde and D. Thalmann, “An integrated perception for autonomous virtual agents: active and predictive perception,” *Computer Animation and Virtual Worlds*, vol. 17, no. 3-4, pp. 457–468, 2006.

M. Minsky, *Society of mind*. Simon and Schuster, 1988.

J. Lifton and J. A. Paradiso, “Dual reality: Merging the real and virtual,” in *International Conference on Facets of Virtual Environments*. Springer, 2009, pp. 12–28.

J. S. Roo and M. Hachet, “One reality: Augmenting how the physical world is experienced by combining multiple mixed reality modalities,” in *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, 2017, pp. 787–795.

M. Luck and R. Aylett, “Applying artificial intelligence to virtual reality: Intelligent virtual environments,” *Applied artificial intelligence*, vol. 14, no. 1, pp. 3–32, 2000.

T. Taylor, “Evolution in virtual worlds,” 2013.

T. Steel, D. Kuiper, and R. Wenkstern, “Virtual agent perception in multi-agent based simulation systems,” in *2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, vol. 2. IEEE, 2010, pp. 453–456.

P. Herrero and A. de Antonio, “Introducing human-like hearing perception in intelligent virtual agents,” in *Proceedings of the second international joint conference on Autonomous agents and multiagent systems*, 2003, pp. 733–740.