Trilobite, A Learning Robotic Creature Using DeepVision

Peter Beyls
University College Ghent, The School of Arts
Ghent, Belgium

e-mail: peter.beyls@hogent.be

Abstract

This paper reports on the most recent incarnation of Trilobite (2021), a bio-inspired robotic creature developing non-trivial behavior in response to environmental activity. Trilobite's objective is to cultivate a symbiotic relationship with its environment by maximizing behavioral complexity in relation to the complexity of movement observed in a given physical space. We take inspiration from trilobite, a creature of the Cambrian era endowed with near 360-degree vision. Trilobite implements similar functionality using two motorized pan-tilt cameras in combination with various computer-vision algorithms including DeepVision targeting face detection, as well as skin color classification. A reinforcement-learning algorithm helps Trilobite to search for people and activity in space efficiently. Trilobite generates large-scale projected images merging input from both camera-eyes with synchronized data visualization. In addition, faces in the image database are continuously analyzed looking for overlapping features as detailed in color histograms. A dynamic animation in 2D space emerges from the spatial organization of faces according to apparent similarity.

1. Introduction

At various occasions, my work took inspiration from examples of exceptional morphology or unusual behavior observed in nature. In particular, biological workspaces reveal abundant instances of behavioral processes displaying perplexing complexity. The present project is a case in point being motivated by trilobites, a species of great variety proliferating during the Cambrian era becoming extinct in the Permian era about 250 million years ago. Trilobites belong to the family of arthropods i.e. invertebrates with exoskeleton. Great diversity characterizes trilobites, e.g. sized from 10 mm to 675 mm and some 20000 different species have been described. Some trilobites featured exceptional vision while others were entirely blind [1]. The Opipeuter's eyes were so big it could perceive over 360 degrees - similar functionality is available using two motorized cameras in the robotic creature introduced in this paper.

1.1 Objectives

Trilobite proposes an artificial creature grounded in physical space. It is receptive of audio-visual input from the environment, including people moving in 3D space. Trilobite gradually learns about the dynamics of space and adapts its behavior accordingly. Then, Trilobite exhibits life-like behavior; it develops a behavioral agenda on the fly and appears as an autonomous creature rather than an automatic machine. So initially, the objective was to create a machine of unpredictable yet seemingly coherent behavior.

One underpinning opinion of much of my work I refer to as the principle of maximization of diversity [2]. Once a machine designed – virtually in software or grounded in hardware – we aim to maximize behavioral diversity, by activating implied degrees of freedom with random noise (software) or embedding a machine in an arbitrarily erratic physical environment (hardware). Trilobite acknowledges this principle in a diversity of ways, i.e. computer-vision runs a machine-learning algorithm aiming to optimize system input in relation to the complexity of audio-visual system output. So, one may view Trilobite as a robotic sensor-activator agent; a system of minimal intelligence in dialogue with the environment.

Trilobite evolves behavior by making decisions and talking actions based on changes in its context in real-time; perception is not static but develops sensitivity to variations over specific time intervals. The dynamics of these changes are my concern, so three main questions arise to be addressed in this introductory paper; (1) how to capture, analyze and interpret sequential camera images, (2) how to manage and coordinate physical articulation of the robotic structure and (3) how to evaluate global system behavior e.g. what is the relationship between observed system performance and the complexity of the aesthetic experience?

Trilobite is a complex reactive machine interacting with the environment; its behavior remains essentially opaque and people in space engage in spontaneous interaction. No one-to-one relationship exists between human and machine behavior. Trilobite is not just mechanistic responsive; it accumulates environmental data in a dynamic memory structure, which informs further truly life-like behavior. Some form of understanding develops in a motivated human interactor while trying to figure out the non-trivial relationship between machine dynamics and environmental stimulation, including human activity. Then, between utter predictability and perceived chaos lies a zone of interaction poetics [3].

While perception is local i.e. observing specific locations in 3D space, Trilobite only acts globally on the environment through image and sound. Memory and machine-learning play are imperative role in temporal complexity of visualization (large-scale projection) since current/recent images often merge with earlier observations.

1.2 Context

Some early vision-based art systems include David Rockeby's projects (1)

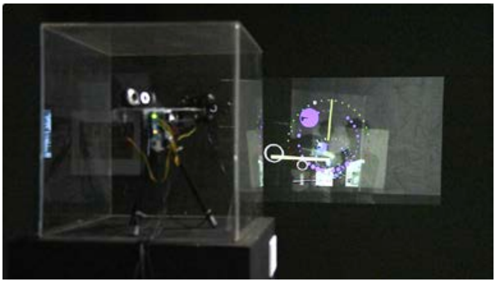

***Figure 1**: Trilobite 0.1 with radar function projection in background*

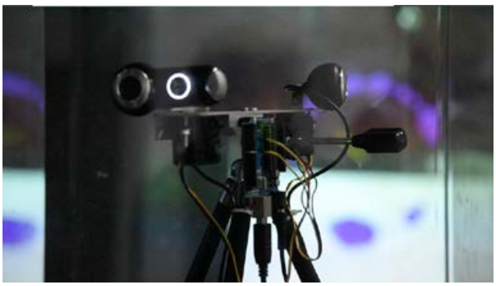

***Figure 2**: Trilobite 0.1 dual motorized cameras (detail)*

VNS, an early multiple camera-based system interpreting human gestures into complex musical patterns, effectively creating an invisible musical instrument and (2) real-time analysis of human activity in open public space with large-scale dynamic visualization [4] and (3) the highly sophisticated *Desire of Codes* installation by Seiko Mikami with motorized cameras observing gallery visitors, taking pictures and creating assemblies from collected material [5] – to name a few.

However, of most relevance to the current project is the earlier sensor-based cybernetics-inspired work of both Edward Ihnatowicz [6] and Gordon Pask [7]. Ihnatowicz designed The Senster (1970), a large computer-controlled hydraulic structure equipped with mechanical eyes (radar technology and robotic microphones) locating and interacting visitors in a large exhibition space. The Senster is a prime example of “the art of automated behavior” [8] a reactive system responding to human activity, however its behavior equally informed by earlier events reported and stored in computer memory. Then, negotiating real-time perception and delayed evaluation makes for complex deterministic yet often unpredictable behavior. Pask’s best know work is entitled The Colloquy of Mobiles, first shown at the celebrated show Cybernetic Serendipity, London 1968. Pask follows a totally biology-based design methodology; suspended electronic sculptures are referred to as male and female inter-actors whose behavior is further informed by human movement. Incidentally, the mechanical layout of The Senster was based on the structure of a lobster’s claw – form and content thus acknowledge a bio-inspired origin.

Contemporary art robotics work within the cyberpunk paradigm includes the wonderful interactive anthropomorphic machines by Chico MacMurtrie [9] and an ecosystem of repeatedly aggressive robots designed by Mark Pauline and his team at SRL [10]. Stelarc provides a new meaning to physically in performance by articulating an industrial robot through physiological signals extracted from his body in real-time [11]. More recent robotic art events include Robot Love (2018) stating: “The mission of Robot Love was to generate attention for humanity in the midst of advancing algorithms and AI. … There is room for fascination, also for conflict, for vulnerability and especially for love.” [12].

These examples remind us that art remains a dynamic system by means of the narrative embedded in art history in

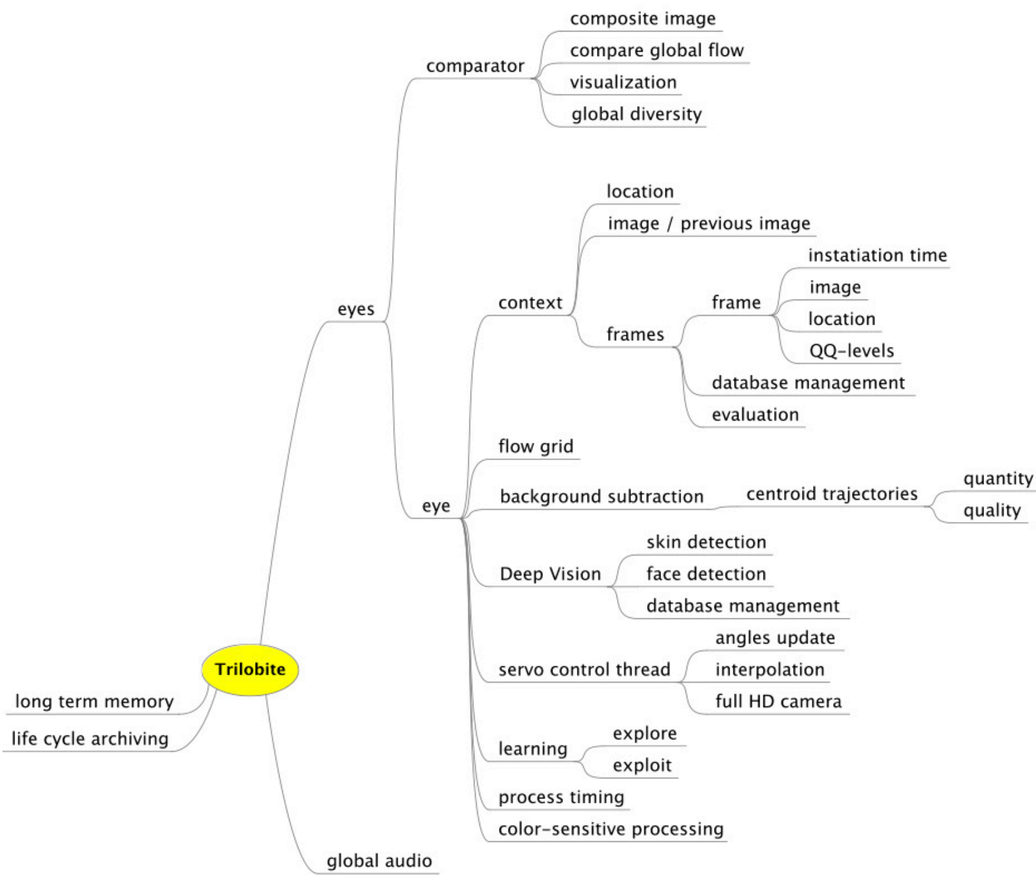

***Figure 3**: Global hierarchical structure of the Trilobite application*

addition to developing critical comment on the whole of human society. More specifically, according to French philosopher Nicholas Bourriaud: "art is an activity consisting in producing relationships with the world with the help of signs, forms, actions and objects" [13].

Instantiation and development of human-machine relationships is indeed key to the very essence interactive art.

1.3 Interaction

The works described above introduce the notion of 'interaction" – an often misapprehended keyword in today's digital culture. This study characterizes human-machine interaction as an evolving relationship between fairly independent participating partners belonging to different species confined in a common physical environment operating by mutual influence, affect and change. Even more specific: we are interested in symbiotic (Greek: living together) interaction where species engage in reciprocal action maximizing long-term mutual benefit. A classic example is the symbiotic bond between the anemone and the clownfish: the clownfish provides service to the anemone by scaring off potential predators and offering nutrition while the anemone offers protection and refuge to the fish. A clownfish' vibrant colors attract other fish looking for a meal, however the unwary would-be predators are then caught and eaten by the anemones.

Based on observations of fascinating morphologies and behaviors in biological and social workspaces, interaction becomes functional from the expression of a critical mass of simple functional building blocks both within a single agent and within natural biotopes. Organic perception and action might indeed be formalized as local interaction amongst simple cognition modules (referred to as 'agents') giving rise to a global emergent functionality [14]. Then, Trilobite – as a macroscopic agent – develops complex behavioral patterns from the expression of many algorithms spread out in its constituent software modules in relation to arbitrary stimulation from the environment. Trilobite optimizes its sensitivity to random environmental changes using a machine learning strategy as detailed below.

Trilobite further suggests the notion of art as a living interface [15]. Viewing art as process rather than product [16] implies an embodied dynamic multi-modal interactive experience. Then, in short, Trilobite views interaction as the exploration of an unpredictable space aiming to capture the complete active environmental options implied in that space. In addition, machine and non-machine agents coexist in a common environment, in a reciprocal relationship sharing resources at equal levels of authority. Both machine and implied spectator engage in a dynamic speculative association; Trilobite adapts and learns while engaged humans develop insight informing the level of aesthetic experience.

2. Implementation

2.1 Structure

Trilobite integrates custom designed hardware and software components. Hardware consists a two pan-tilt motorized Full HD resolution cameras driven by an Arduino board with servo shield. Pan motors span 180 degrees action while tilt motors cover 90 degrees.

Software (written in JAVA and C++) follows a hierarchical object-oriented programming paradigm as shown in figure 3. Vision comprises both local activity in a single eye and global action aiming integrated visualization from two independent vision inputs.

The Eye object contains various specific computer-vision modules (based on OpenCV, [17]), timed observation process control and learning. Vision aims to track changes in the environment using background subtraction (detection of moving objects) and global optical flow detection (estimation of the direction and amount of movement between two consecutive vision frames). In addition, grid flow detection computes local flow in a lower resolution grid superimposed on the camera image. Flow data updates to a 2-dimensional memory structure with activation (high values) or inhibition (low values) providing a fading memory of precise locations of activity in space. Finally, the Eye computes a hue color histogram informing about the diversity of the currently observed colors in space. Consequently, a color-tracking algorithm computes and visualizes major areas in space featuring a particular prominent color.

Trilobite's vision modules ultimately infer a more abstract deduction from actual sequential pixel-based camera images: observed changes are evaluated in terms of levels of quality and quantity. Using background-subtraction (figure 6), we compute the centroid of movement – the average location of change in the 2D image. Perceived consecutive images (about 20 frames/sec) then produce a sequence of centroids, a trajectory of xy-locations in 2D space. Formally, a trajectory is a series of vectors, units of

***Figure 4**: Studio setup: face recognition test and context visualization*

***Figure 5**: Dual camera visualization: merging two images*

particular size (amount of movement) and angle (direction of movement). Level of quality is then reflected in the diversity of quantized angles (6 degrees resolution) and sizes, the number of unique angles/sizes vs. the total number of angles/sizes. Level of quantity follows an adaptive algorithm, it computes the amount of change rather than its complexity. Quantity is then proportional to the area of change, the surface of the rectangular bounding box comprising a trajectory. Fluctuating Quality-Quantity levels (QQ-values) inform a reinforcement-learning algorithm to be addressed in a moment. Since Trilobite is motivated to interact with people, it features individual person detection and more focused face detection, including skin color detection. Faces are captured, including location in

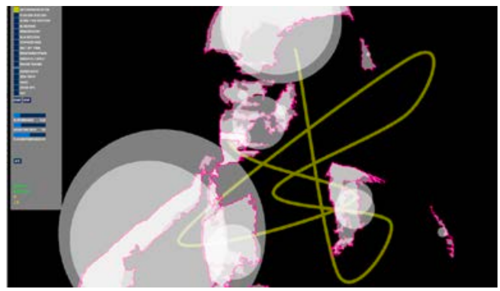

***Figure 6**: Background subtraction and movement tracking/visualization*

***Figure 7**: Skin color classifier animation test, snapshot with 100 images.*

3D space, and saved in a large online database of a deliberately limited size of 100. Following our diversity principle, the database continuously maximizes visual diversity, faces too similar are removed, a new face is only acquired when sufficiently dissimilar from all existing ones. Face detection is based on the

DeepVision library (2019) devised by Florian Bruggisser [18] and further extended to handle separation of pixels representing skin colors into new images. Then, following the HSB color model (hue, saturation and brightness) a histogram is computed for every color parameter, quantized into a 15–bin array, globally resulting in a 45 element list documenting skin color of a single person. Considering the absolute value of the respective difference of all histograms then provides an idea of the global skin-color multiplicity thus suggesting critical debate on the shared social impact of surveillance technology [20].

In addition, a classifier algorithm views the surfacing faces database as a complex generative system driven by the inherent features of the constituent images. Images reposition themselves according to local similarity, similar images tend to cluster and connect when within a given threshold distance - the yellow lines displayed in figure 7. An algorithm selects two random images and repositions one image at a distance to the other relative to mutual similarity creating dynamic groupings. However, all groups are temporary since clustering gradually builds up but also gradually disintegrates since a randomly selected image creates visual coherence at its new location and visual inconsistency at its previous location. Figure 7 shows a simulation using images taken from the high-quality FFHQ dataset initially developed for GAN-research [21].

2.2 Context and learning

Both Eye modules hold a single Context object holding the currently perceived camera image and its location (reflected in the respective angles of the pan-tilt servo motors) while also recording a private detailed observation history. A timing process commands the Eye to capture changes during a given time span, when finalized, a Frame is created, it documents its instantiation-time, the camera image, its location in 3D space and the resultant QQ-values. Eyes maintain a finite variable-size dynamic list of frames; existing frames only survive in relation to their age (instantiation-time), QQ-levels and pressure of a potentially higher-level nascent frame.

Our intention with reinforcement learning [19] is to gradually maximize behavioral complexity over time. Therefore, the eye needs to explore the environment, turning to inspection of random locations in 3D space. Exploration hopes to find areas of interesting behavior (revealed in high QQ-levels) through trial-and-error. In addition, a parallel competing process aims to exploit the information that has been learned so far. A simple learning procedure needs to balance two options – (1) random search or (2) making good use of accumulated data and revisiting a particular area expecting additional renewed activity. Typically, the ee-ratio (exploration-exploitation ratio) starts from zero (only exploration) and increases to about 0.5 when enough frames have been accumulated.

Figure 4 documents a studio experiment; a camera detects a human face in a painting on the wall (highlighted by the red square) and previously captured image frames held in the current context are displayed in a 2D image relative to their physical locations in 3D space. The projection of the eye's images has a major impact on the actual perception of Trilobite's actions – merging both eyes' images using linear interpolation of colors creates a single complex composite projected image as seen in figure 5.

3. Discussion and conclusion

Trilobite integrates a multiplicity of ideas: machine-mediated interaction, optimization of the interactive experience through machine learning and the notion of dynamic image database.

Trilobite incessantly searches for changes in the environment in terms of people's global movements and, more specifically, looking for human faces it has not encountered before. A database of all detected skin colors is maintained and organized according to similarity. An algorithm then manipulates the database aiming to maximize visual diversity of skin color.

This paper provides a short overview of Trilobite; a robotic installation integrating advanced computer-vision, reinforcement learning and complex visualization. Focus is on the relationship between the nature of machine behavior, physical movement in 3D space and how a human interactor/observer participates in this process. Trilobite aims to maximize the complexity of this relationship as to optimize the aesthetic experience. A reinforcement-learning algorithm takes note of the 3D position of highly dynamic locations in space – Trilobite then balances its behavior between exploration (random search for activity in space) and exploitation (revisiting successful previous locations). Background detection captures the amount and complexity (formalized as quality and quantity of spatiotemporal trajectories) of physical changes in space, which informs the learning algorithm.

Global dynamic behavior unfolds since Trilobite exists in a grounded, basically unpredictable environment including human interference. In sharp contrast with merely responsive systems, the notion of symbiotic interaction is suggested where humans and machine mutually coexist in a common biotope with equal authority. Initial studio experiments reveals fascination with system behavior: e.g. the coordination of eye movements in relation to external activity, the impression of the assimilation process of both eye's images into a single dynamic manifestation and the real-time sounds mapping variations in captured spatial activity to significant audio synthesis parameters. People appear to perceive the expression of life-like impulsive yet seemingly consistent machine behavior.

Full appreciation of the system's complexity definitely necessitates large-scale, long-term experiments in crowded public spaces.

References

[1] Gon, SM III. 2014 The Trilobite Eye, https://www.trilobites.info/eyes.htm (visited June 9, 2021)

[2] Ox, Jack. 2019. Peter Beyls and Algorithms in: *Coming Full Circle*, Verbeke Foundation, Belgium, pp. 81-116

[3] Penny, Simon. 2017 *Making Sense,* MIT Press, Cambridge, MA, p. 404

[4] Rockeby, David. 1998. The Construction of Experience, in: *Digital Illusion: Entertaining the Future with High Technology*, ACM Press

[5] Makami, Seiko. 2011 Desire of Codes, https://www.ntticc.or.jp/en/feature/2011/Desire_of_Codes/index.html (visited June 9, 2021)

[6] Ihnatowics, Edward 1977 On the Relevance of Mechanical Manipulation to the Process of Perception, *PAGE 38,* Bulletin of the Computer Arts Society, London, UK, pp. 6-7
[7] Pask, Gordon 1971 A comment, a case history and a plan, in: *Cybernetics, Art and Ideas*, JasiaReichardt (ed.) New York Graphic Society Books
[8] Penny, Simon, 2017 *Making Sense,* MIT Press, Cambridge, MA, p. 366
[9] McMurtrie, Chico http://amorphicrobotworks.org/ (visited January 1, 2021)
[10] Pauline, Mark https://srl.org/mark.html (visited January 1, 2021)
[11] Smith, Marquard, 2005 *Stelarc, The monograph,* MIT Press, Cambridge, MA
[12] *Robot Love*, 2018 TerraLannoo, Amsterdam, Holland
[13] Bourriaud, Nicholas 2002 *Relational Aesthetics*, Les Presses du Réel, Dijon, France
[14] Minsky, Marvin 1986 *The Society of Mind*, MIT Press, Cambridge, MA
[15] Beyls, Peter 2018 Art as a Living Interface, In: Marcus A., Wang W. (eds) *Design, User Experience, and Usability. Design Philosophy and Theory*. HCII 2019. Lecture Notes in Computer Science, Vol. 11583. Springer
[16] Kwastek, Katja 2015 *Aesthetics of Interaction in Digital Art*, MIT Press, Cambridge, MA
[17] Gollapudi, Sunila, 2019 Learn Computer Vision using OpenCV, Apress
[18] Brugisser, Florian 2019 https://github.com/cansik/deep-vision-processing
[19] Sutton, Richard and Barto, Andrew 2018 *Reinforcement Learning*, MIT Press, Cambridge, MA
[20] Facial Recognition Is Accurate, if You're a White Guy https://www.nytimes.com/2018/02/09/technology/facial-recognition-race-artificial-intelligence.html
[21] NVLabs Flicker Faces HQ Dataset (FFHQ) https://github.com/NVlabs/ffhq-dataset (visited July 9, 2021)