Interfacing Niha With Nao (Nina) By Implementing Sensory-Motor And Perceptual Memory Units

Author

Syed Kamran Ahmed

Regn Number

00000170346

Supervisor

Dr. Yasar Ayaz

ROBOTICS AND INTELLIGENT MACHINES ENGINEERING

SCHOOL OF MECHANICAL & MANUFACTURING ENGINEERING

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY

ISLAMABAD

AUGUST, 2020

Interfacing NiHA with NAO (NiNA) by implementing sensory-motor and perceptual memory units.

Author

SYED KAMRAN AHMED

Regn Number

SMME – RIME 2016 - 170346

A thesis submitted in partial fulfillment of the requirements for the degree of

MS Robotics and Intelligent Machine Engineering

Thesis Supervisor:

DR. YASAR AYAZ

Thesis Supervisor's Signature: ______________________________________

ROBOTICS AND INTELLIGENT MACHINE ENGINEERING

SCHOOL OF MECHANICAL & MANUFACTURING ENGINEERING

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY, ISLAMABAD

AUGUST, 2020

Declaration

I certify that this research work titled "*Interfacing NiHA with NAO (NiNA) by implementing sensory-motor and perceptual memory units.*" is my own work. The work has not been presented elsewhere for assessment. The material that has been used from other sources it has been properly acknowledged/referred.

Signature of Student

SYED KAMRAN AHMED

2016-NUST-Ms-RIME-170346

Thesis Acceptance Certificate

It is certified that the final copy of MS Thesis written by *Syed Kamran Ahmed* (Registration No. 00000170346), of Department of Robotics and Intelligent Machine Engineering (RIME) has been vetted by undersigned, found complete in all respects as per NUST statutes / regulations, is free from plagiarism, errors and mistakes and is accepted as a partial fulfilment for award of MS Degree. It is further certified that necessary amendments as pointed out by GEC members of the scholar have also been incorporated in this dissertation.

Signature: ___ _

Name of Supervisor: Dr. Yasar Ayaz

Date: __________________

Signature (HOD): __________________

Date: __________________

Signature (Principal): __________________

Date: __________________

Plagiarism Certificate (Turnitin Report)

This thesis has been checked for Plagiarism. Turnitin report endorsed by Supervisor is attached.

Signature of Student

SYED KAMRAN AHMED

Registration Number

SMME – RIME 2016 - 170346

Signature of Supervisor

Copyright Statement

Acknowledgments

I am thankful to my Creator Allah Subhana-Watala who guided me throughout this work at every step and for every new thought which He set up in my mind to improve it. Indeed I could have done nothing without His priceless help and guidance. Whosoever helped me throughout the course of my thesis, whether my parents or any other individual was His will, so indeed none be worthy of praise but Allah.

I am profusely thankful to my beloved parents who raised me when I was not capable of walking and continued to support me throughout every department of my life.

I would also like to express special thanks to my supervisor Dr. Yasar Ayaz for his help throughout my thesis and also for AI and Mobile Robotics courses which he has taught me. I can safely say that I haven't learned any other engineering subject in such depth than the ones which he has taught.

I would also like to pay special thanks to Dr. Wajahat Qazi for his tremendous support and guidance. Each time I got stuck in something, he came up with the solution. Without his help, I wouldn’t have been able to complete my thesis. I appreciate his patience and guidance throughout the whole thesis.

I would also like to thank AP. Sara Baber Sial, Mr. Hamza Asif and Syed Tanweer Shah Bukhari for being on my thesis guidance and evaluation committee, and express my special thanks to Maaz Mohsion for his help. I am also thankful to Saran Khaliq and Mehmood Hussain for their support and cooperation.

Finally, I would like to express my gratitude to all the individuals who have rendered valuable assistance to my study.

Dedicated to my exceptional parents and adored siblings whom tremendous support and cooperation led me to this wonderful accomplishment.

Abstract

Cognitive architectures are the key foundation of any socially interactive robots in a human-robot interactive environment. Nature-inspired Humanoid Cognitive Computing Platform for Self-aware and Conscious Agent (NiHA), published in 2018, is a cognitive agent that was implemented using Quantum Bio-Inspired Cognitive Agent (QuBIC) and simulated iCub robot. This research intends to integrate NiHA with the NAO robot, replacing simulated iCub in NiHA. NAO will enable NiHA to socially interact with humans using its existing cognitive capabilities. In this research, a cognitive interface has been developed by re-implementing sensory-motor memory, visual perceptual associative memory, and procedural memory block of NiHA. Keeping in view due to the cognitive and computational complexity of NiHA, only socio-communication skills will be interfaced and evaluated. Current research implements perceptual associative memory along with modules required to get sensory data from NAO robot, which is named NiNA. NiNA encodes visual perception in NiHA’s Knowledge Representation Scheme in the form of perceptual signals. These signals will be transmitted to other existing cognitive modules of NiHA which can generate a response. These responses will be utilized by the newly implemented version of procedural memory to generate actions that will involve speech with related body language.

Key Words: Cognitive Architectures, Humanoid NAO, Human-Robot Interaction (HRI), Visual-Spatial Relationship.

Table of Contents

List of Figures

List of Tables

CHAPTER 1: INTRODUCTION

The research work in this dissertation has been presented in two parts. The first part is related to the current era of robotic development, the use of robots in every aspect of life is increasing dramatically, the growth of robotic arms and other robotic machinery in the industry has already increased drastically. Now the introduction of service robots in daily human life as personal robots has increased the human-robot interaction and opened a completely new domain of socio-cognitive robotics, which requires a complex robotics architecture to evolve further. Their software architectures and the introduction of robots to a human lifestyle strengthened human-robot society's collaborative existence. To accomplish the desire of creating or be part of such technological advancement, this research work has implemented sensory-motor memory; visual Perceptual Associative Memory (PAM), and procedural memory of NiHA [1], a cognitive architecture on NAO [2], a humanoid robot. The second part includes visual Spatial Relations such as left, and the top can provide fine queries to locate object location as well as its relation with other objects within an image.

1.1 Background and Scope

Robotics includes a broad variety of disciplines, one of which is Human-Robot Interaction (HRI). HRI to its other counterparts is comparatively young, but it has received a lot of attention in recent years. Because of the collective growth of complex robotic models; their software architectures and the exposure of robots in the human lifestyle to make interactive society more possible for human robots. The next generation of companion robots or rather a human-robot working environment will need to satisfy certain social requirements for collaborative robots, somehow, similar to the famous laws of robotics envisaged by Isaac Asimov [3].

Robots and people have long been co-workers, but we never work together. This can change with the growth of collaborative robotics. Collaborative robots are not placed behind glass or in the cages, as opposed to traditional industrial robots. They are instead built to be safe and dexterous. According to Jim Lawton and Daniel Huber [4], a roboticist from Rethink Robotics, optimizing robot software and AI will be the key in the near future, for making robots collaborative. Many others including Mr. Huber are working on "socially conscious" robots so that they can better understand what robots can do to react appropriately according to a social setting. If a robot cannot distinguish someone and something from another or does not understand when someone asks them to stop it, it is ineffective for the crowded place of work. Robots must be designed to carry out a wide variety of tasks to integrate robotic systems into real-world environments and evolve constantly for the modifications in working conditions.

As all environments and task situations cannot be modeled, standard end-user programming will not be able to adapt to any new tasks [5]. Instead, it must provide a robot with the advanced capacity to autonomously exist, and through its user to learn new tasks and new working conditions. In everyday circumstances, there is a tremendous increase in robots that can communicate securely with individuals. These robots must be in a position to anticipate the impacts of their actions and their outcomes on the individuals around them. For that purpose, we need to combine two major streams, the first, physical structure designed specifically for an unconstrained environment to communicate and coexist. Second, architecture to make use of knowledge-base and the need to obtain information.

1.1.1 Robot Cognition

So, we can say cognitive architecture is the intersection of robotics and cognition. This means cognition can be described and manifested through action as the capacity to perceive, learn, and reason [6]. It remains disputed and challenging to produce such a machine. To analyze the gap between the state of art and challenged demands the BICA Challenge is introduced [7]. There will be identified special challenges and barriers; a method to resolve them and overall functional criteria for success.

A hopeful prospect in regards to creating such a machine is the BICA Challenge. The BICA Challenge is the challenge of constructing an overall, real-life calculating equivalent of the human mind using a Biologically Influenced Cognitive Architecture (BICA) approach. To overcome this, we must understand how natural intelligent systems are evolving their cognitive, metacognitive, and learning functions at a computational level. Three primary criteria are included in BICA challenge that smart agents must have:

1) Compatible with human beings and their usefulness in human society
2) Auto-sustainability

3) Human-extending

First, it must have the standard of human beings in general intelligence, communication, and training capabilities, and generally helpful as human experts and workers. On the other hand, an agent must be honest, clear, and trustworthy as judged by humans. Second, it must be able to take care, with progress, production, and demand in mind, of its continuity and growth in a manner that guarantees an open-ended progress scenario and social integration. Third, the agents must be efficient carriers of human nature, the spirit, human ideals, and human society and even human minds in a distant future. It is expected that the solution to these challenges will lead to a groundbreaking for intelligent agents incorporated into human society. This enhancement allows people to solve many human civilization issues and ultimately makes our planet a safer place for us.

In the last few decades, there was a lot of development in the evolution of robots. With the exponentially increasing knowledge and an unlimited changing environment, it will not be possible to program and develop efficient robots that can survive in a human-robot environment in the absence of either environment perception or communication skills. Those skills can be according to the social norms of society coupled with the verbal and non-verbal communications [8]. To present verbal or non-verbal information a robot must be created with the appropriate sensory structure. Such structure may include motor perceptual memory to sense (receive) data from the environment, visual PAM to detect objects from the visual feed, working memory to create an environment and detected object's relation, and procedural memory to act either as a response to something detected or initiate a new task altogether.

Several researchers are involved in the quest to build human-like machines able to mimic consciousness to make them more efficient and applicable for versatile environments. To support these technologically advanced artifacts, emerging lots of different models and architectures. Some scholars are trying to develop artificially intelligent models that can have human-level intelligence or at least they have a minimal level of consciousness. A question arises, can a mechanical system have consciousness? Does it depend on the material of the brain? Will we ever be able to achieve artificial consciousness? [9]

As we progress to achieve the human-robot coexisting community, it's imperative without the basic knowledge of how humans perceive its surroundings and create a mental image, and how well a robot can understand its environment. Will we be able to mimic humans in this aspect of creating a perception of its environment? There are many degrees of environment perception from a minimal level of the pre-given map to high-level cognitive architectures to support the perception of a conscious agent. Retrospectively, there have been efforts to create robots in human form having the ability to interact and respond to dialog cues engaging in imitation and emotion recognition. According to Alekseander and Dunmall, machine architecture must observe certain Axioms (Depiction, Imagination, Attention, Volition, and Emotion) to possess minimal consciousness [10]. There are many cognitive architectures based on these Axioms, working on versatile frameworks and implementing diverse techniques and theories to accomplish artificial consciousness. Some models and techniques to implement them are discussed further.

There have been many architectures based on metaphysical theories of consciousness for the understanding of the processes and states involved in cognition. Furthermore, these theories have provided the essential elements for the formation of the working theory of mind/brain and the problem to incorporate consciousness along with emerging cognitive architectures [11]. Moreover, to build such an agent, an ontology for comparative cognition has been presented by IDA [12], according to which perception and procedural memory are essential parts of cognitive architecture and to act together with sensors, sensory-motor memory is compulsory. There are cognitive architectures that might not have those modules prominently but the functionalities of these modules exist in there, like Three Layered Cognitive Architecture of Karlsrube Humanoid Robot [13].

Towards accomplishing such cognitive abilities this study on NiHA will be a stepping stone by reimplementing sensory-motor memory for getting sensory data from different sensors of NAO [2] and forwarding it to Depiction, for perceiving and experiencing the environment to the formulation of awareness. Afterward getting responses from NiHA consciousness and generating actions from procedural memory.

1.2 Motivation

This and other concepts used in the introduction motivated me to work in the field of cognitive robotic. The use of robots in daily lifestyle has increased a huge market for service robots all around the world including Pakistan. Every major institute and industry is moving towards Industry 4.0 that is growing tremendously in Pakistan.

1.2.1 RISE Lab & NAO

RISE Lab is a leading robotics lab in Pakistan, working and researching in the areas of therapy robotics and machine intelligence with a special focus on design, control, and motion planning for robotic systems including mobile robots as shown in (Figure 1.1) humanoid robots Like NAO & Pepper, multi-legged robots, intelligent bionics, and robotic manipulators, etc. There are many amazing projects like Therapy for autistic child, Robo Cup etc

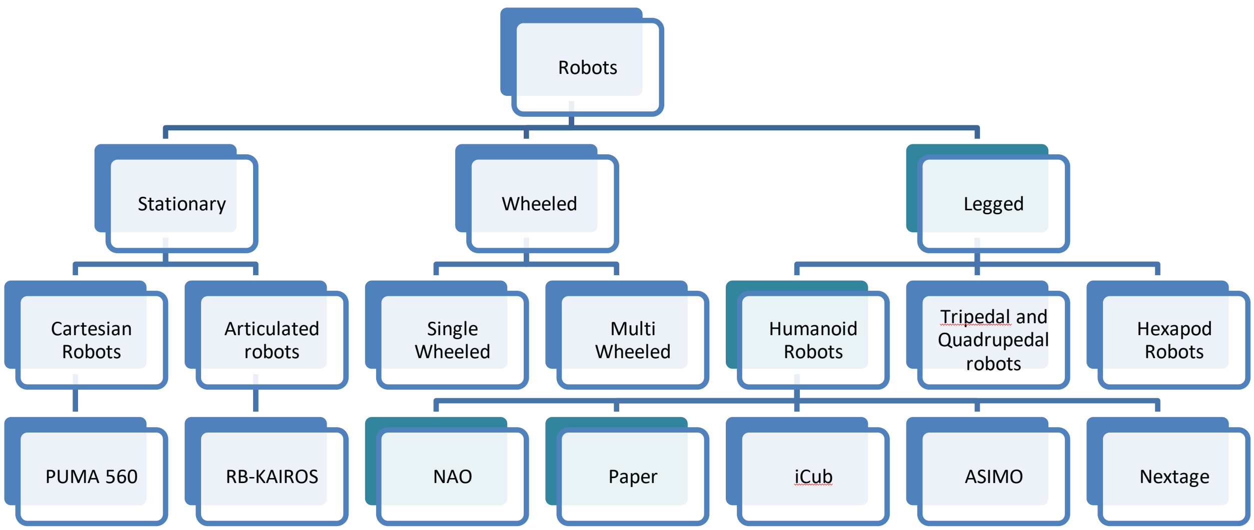


Figure 1-0-1: Mobile Robots & Types

1.2.2 Cognitive Architectures and NAO

This gave the inspiration to work on the NAO humanoid robot. Creating an interface between NAO and cognitive architecture. That can enable NAO to work autonomously in a social environment, creates memory, understand natural language, and act accordingly. In other way behave more humanly to remove barriers between human-robot interactions.

Cognitive architectures are too complex to be built all at once. But spatial relations of objects in an image and the importance of these in perceiving the world influence me the most. How humans perceive their surroundings and objects placed in it. Creating relations between different objects and infers based on them.

1.3 Problem Statement

An agent is required to perceive and understand the environment to have social interaction with their co-workers. This ability involves the incorporation of cognitive capabilities on top of existing control structures. The current robots have limited skills to cognition which involve: visual perception, coordinated actions, and co-occurrence of objects in an environment. Hence are unable to comprehend and efficiently interact with the environment. The problem that will be addressed in this research is to incorporate existing robot (NAO) with cognitive architecture to mimic the human-inspired skills. The artifacts of architecture involve sensory, working, perceptual, procedural, semantic, and episodic memories. These cognitive constructs will allow an agent to have visual perception, and object co-occurrence to understand and effectively interact with the environment.

1.4 Research Objectives

Our focus in this study is the bottom-up attention which is also called a stimulus-driven approach in which information comes to memory through sensory organs. This knowledge is then influenced by perceptual associative memory to generate symbolic information. This is enhanced by working memory to identify spatial relationships from visual feed and propagate visual depiction of the environment. Producing action through procedural memory.

CHAPTER 2: LITERATURE REVIEW

In this section, we will be discussing cognitive architectures and their types of different cognitive modules in comparison to NiHA. Some implementations of cognitive modules on humanoid robots including NAO. In the end, analyzing the literature review.

2.1 Cognitive Architectures

Conscious experiences are related to sensory stimuli (Input), awareness, thoughts, and perceptions. In addition, the interpreters, memories, language, and automatism are other elements affecting the emergence of consciousness. Cognitive architecture studies often describe different skills, properties, and parameters, including understanding, decision-making, awareness, estimation, preparation, performing, communicating, studying, setting of goals, adaptability, generality, self-sufficiency, problem-solving, re-workout, meta-learning, etc.

While these classification principles may be implemented, they are too fine-grained to refer to a common architecture. A more general category of architectures may also be focused on the method of representation and retrieval of information they carry out. There are three primary paradigms established: Symbolic, emergent, and Hybrid [14]. As shown in (Figure 2.1)

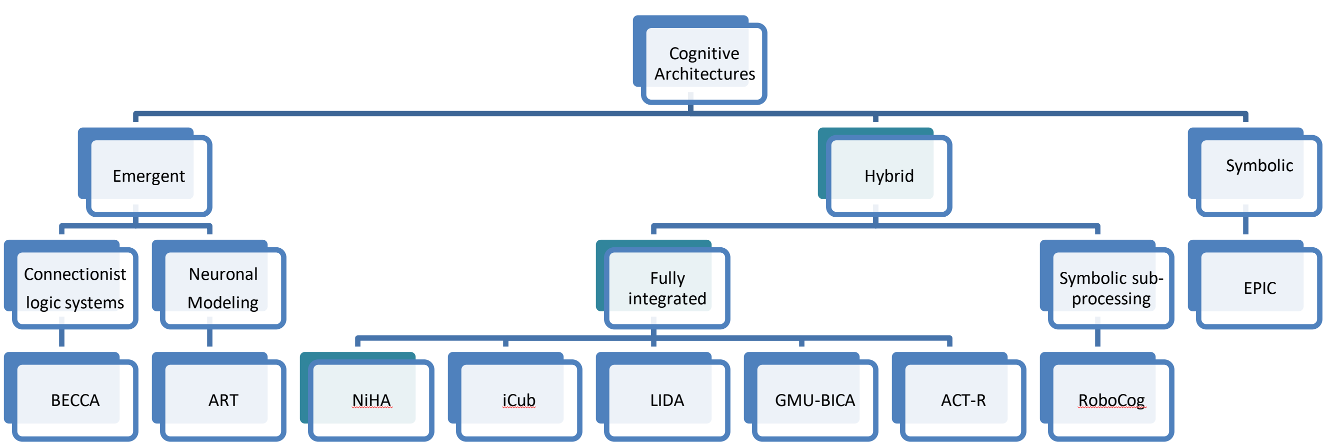


Figure 2-0-1: Three Major Paradigm of Cognition

Symbolic structures reflect ideas that can be managed using a predefined series of instructions using symbols. These guidelines might well be implemented as if they were laws for the symbols describing world-facts. While symbolic structures are superior in architecture and analysis, they are less capable of resolving the complexity and solidity needed to cope with an evolving world and for perceptional processing.

The emergent addresses adaptability and learning problems by developing massively parallel models that are similar to neural network data streaming, where the distribution of signal from input nodes is interpreted. The resulting method, however, also lacks clarity since information is no longer a collection of abstract entities and is spread over the network. For these purposes, in a typical context, logical inference becomes a challenge in evolving architectures, if not impossible.

Hybrid architectures aim to merge conceptual and evolving elements. The key focus of our research studies is on these structures. The fully integrated architecture incorporates various concepts using a range of techniques. Such architectures are seen as a collection of integrated, conflicting, and cooperative modules that do not confine individual modules to a peculiar theory.

2.2 Mind-Body Cognitive robotics

In the last few decades, there has been a lot of development in the evolution of robots. To support these technologically advanced artifacts there are various emerging models and architectures. Some scholars are trying to developing Artificially Intelligent models that can have human-level Intelligence or at least they have a minimal level of consciousness. The questions arise, can a mechanical system have consciousness? Does it depend on the material of the brain? Will we ever be able to achieve artificial consciousness? [9] To develop an understanding of different Cognitive Architecture and why we need them, more importantly how they work, these are the theories that serve the foundation of different models and architectures.

1. Theory of Dualism [11].
    a. Interactionist dualism
    b. Psychophysical parallelism
2. Property Dualism
    a. Epiphenomenalism

3. Theory of monism
    a. Idealism
    b. Physicalism [15].
4. Global Workspace Theory [16, 17] as show in Figure 2.2

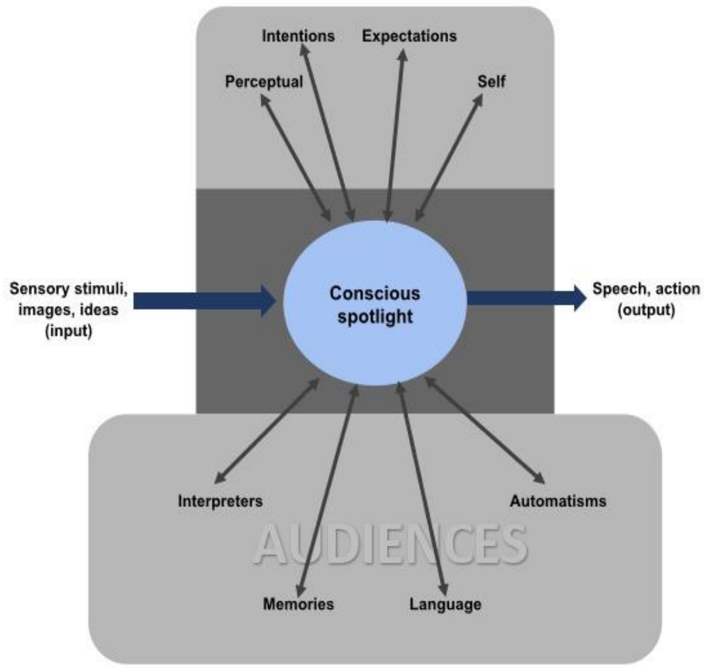


Figure 2-0-2: Global Workspace Theory

Disjoined scientific groups may speak different languages and follow separate, specific objectives in order to achieve a conscious agent. A widely promoted public discussion of the overall BICA Problem will play an integrative role in this situation.

2.3 BICA Road Map

A cognitive architecture is a theoretical model for the creation of intelligent agents and we call it "biologically inspired", as it attempts to replicate functional properties of the human mind. The BICA Challenge [18] can be described as the challenge of constructing a digital version of the human mind for real-life and general use. To address this, we need to understand how natural intelligent systems build their cognitive, metacognitive, and learning functions at a computational level. A solution to the problem will be an intelligent agent (a cognitive architecture) that implements a conceptual representation of the basic human mind and can, therefore, be viewed by humans as a human mind. Can it learn from humans as an apprentice; can become a valuable member of a team as a partner; and more importantly, can become a part of human rights society. The BICA Challenge, with its ties to financial, legal, ethical, political, technical, technological, and other aspects, is emerging as a modern all-scientific mainstream challenge of our time that requires a multi-national system.

In essence, the BICA Challenge can be described as the task of making virtual agents recognized as useful participants by social structure and viewed on an equal footing with human members of the community, initially within restricted territories, teams, and environments. Significant investments and rapid improvement in the sector can be anticipated at the point when the problem is overcome, resulting in rapid growth. There are currently a variety of research projects across the globe that address the BICA Problem directly or indirectly. However, despite impressive achievements and increasing interest in BICA, there are still large gaps separating various strategies from each other and solutions sought in biology.

There are various purposed architectures with the intent of implementing distinct features from the above-mentioned theories, some are summarized here.

2.3.1 LIDA by Stan Franklin

Stan Franklin has developed a comparative cognition ontology, where he uses a functional paradigm to describe various steps to build a basic model of cognitive architecture [12]. He describes how percept receives attention and contact between episodic Memory, also describes how procedural memory and enhancement learning can be used-The IDA cognition mode [19] he used is a sort of theory of everything, including perception, emotions and feelings, various types of memory, attention, atomization, philosophic zombies, reasoning, and so forth.

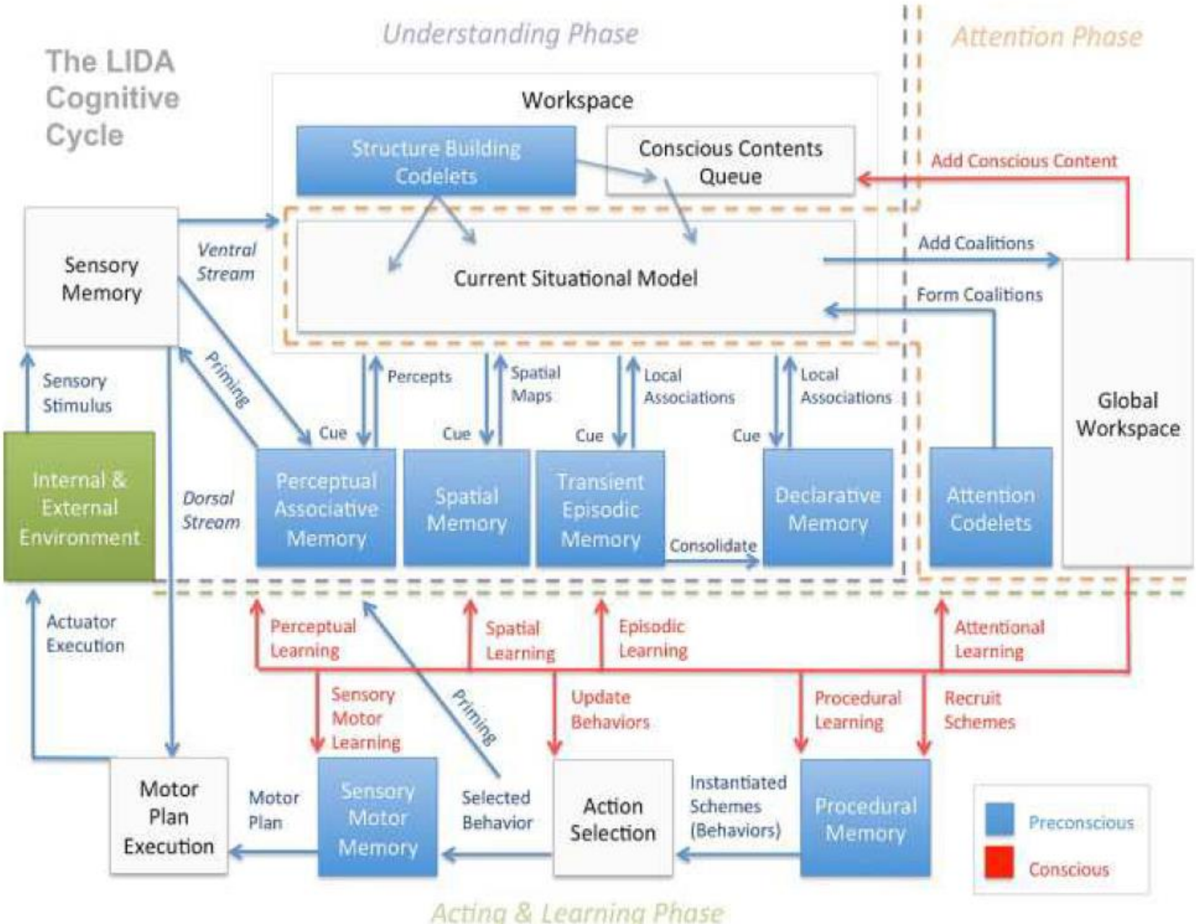


Figure 2-0-3: LIDA

2.3.2 CLARION

In some important respects, CLARION [20] varies greatly from most current cognitive architectures. CLARION is a hybrid approach by (a) combining connectionist and symbolic representations, (b) combining psychological processes implicit and explicit, and (c) combining cognition (in the narrow sense) with other psychological processes. Ultimately, CLARION is a cognitive architecture that is modularly organized and consists of many functional subsystems. This also has a dual representational structure, with representations both implicit and explicit. CLARION has succeeded in identifying several psychological processes through a range of mission domains based on its integrated modules structure.

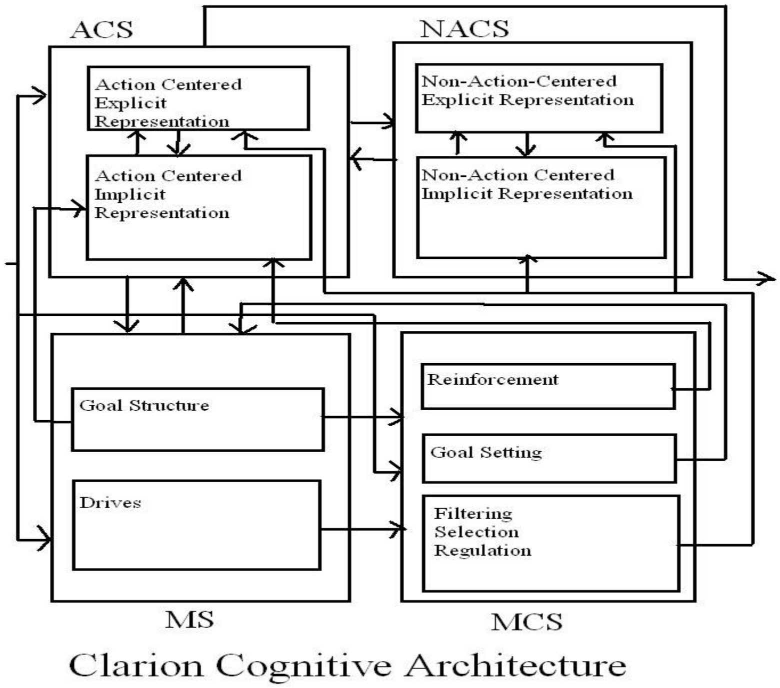


Figure 2-0-4:CLARION

2.3.3 ACT-R

It is a well-known cognitive architecture, resulting from the incremental development of Anderson's model of human cognition [21], which originates from his model of Human Associative Memory [22]. The key principle is the separation of two forms of information semantic and functional with adequate activation, the machine is only conscious of information.

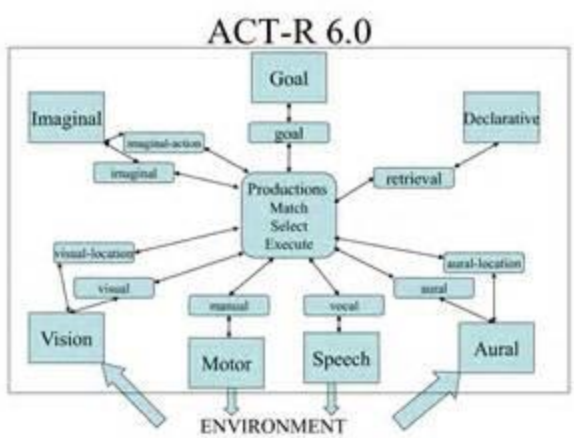


Figure 2-0-5: ACT-R

2.3.4 iCub

After a detailed survey of Symbolic, emerging and hybrid cognitive architecture, an analysis of human neurons phylogeny and ontogeny, and a review of design principles for development Systems were carried out, the iCub cognitive architecture was created. The design of iCub architecture is especially influenced by two architectures surveyed: Shanáan's Global Cognitive Workspace Architecture and the Dynamic Neural Field Architecture of Erlhagen and Bicho [23]. The cognitive architecture of the iCub focuses on self-design. Development requires evolution as a basis; in other words, ontogenesis requires some initial phylogenetic structure to be based on.

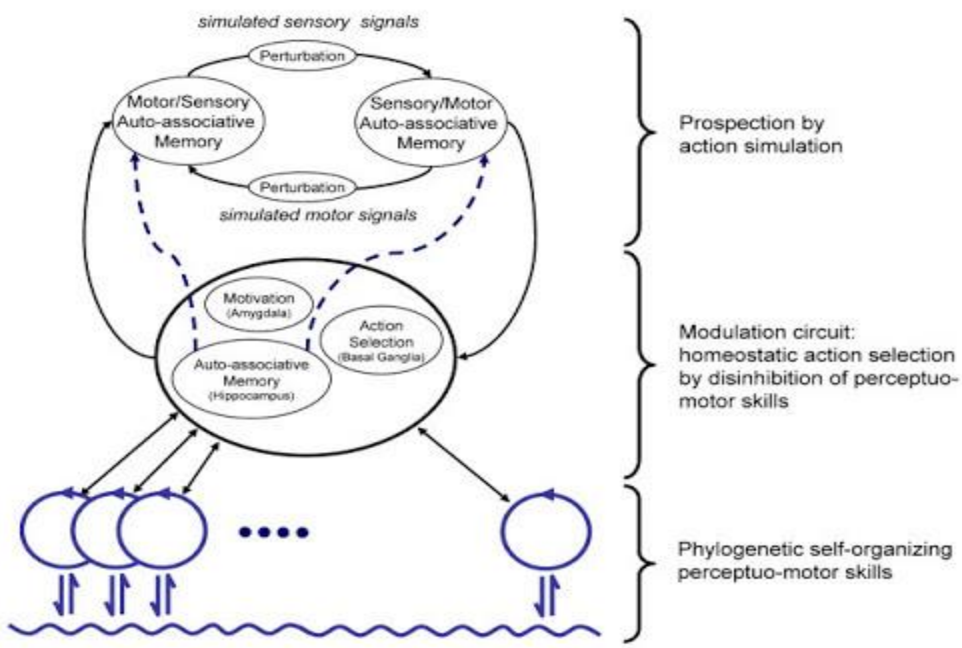


Figure 2-0-6: iCub Architecture

The cognitive architecture of iCub consists of a multifunctional, competing, and cooperating distributed perceptual-motor circuits, a modulation circuit that effects homeostatic action selection with disinhibition of perceptual-motor circuits, and a perceptual-action simulation system to anticipate effects. The modulation circuits comprise three components: Automotive memory, neural field-based action dynamic selection, and hippocampus-based motivation, basal ganglia, and amygdala respectively, while the advancement circuit contains combined hetero associative memories of the engine sensor and sensor motors.

2.3.5 GMU-BICA

There is an emotional intelligence candidate paradigm defined for the combination of theoretical, modeling, and experimental methods. The paradigm consists of three new elements that allow emotional processing to be represented: an emotional state, an evaluation, and a moral schema. These components are merged into the inadequate cognitive foundational map that reflects the principles of emotional evaluations. The system measures the results in two new experimental paradigms that demonstrate basic features of human social processing, such as the presence of recurring positions interpreted subjectively by individual virtual agents. The results are tested. Implications refer to heterogeneous teams of human robots [24].

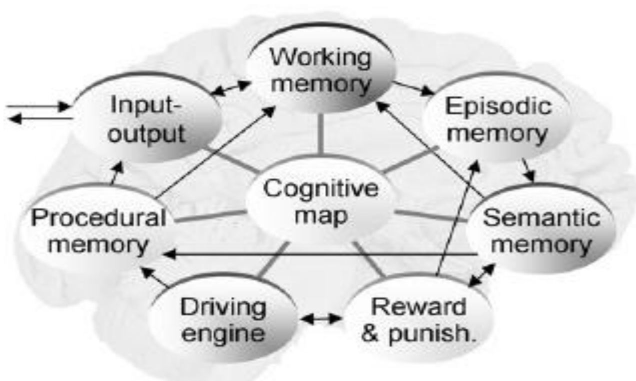


Figure 2-0-7: GMU-BICA

2.3.6 A Cognitive Architecture for a Humanoid Robot: A First Approach

Talking about humanoid robots we think of a future where humans and robots co-exist. Humanoid robots taking part in our daily life and for that purpose Catherina Burghart and his fellow researcher's present research on “A Cognitive Architecture for a Humanoid Robot: A First Approach” [13] This Architecture tried to create a complete symbolic world model of the robot environment using sensor data and to make a plan on a symbolic level. It has a 3 layered Architecture adapted to the requirements of a humanoid robot.

a. Top Level
b. Mid-Level
c. Low Level

This architecture supports interaction between components on a parallel based. In other words, each level acting independently. The core advantage is fast reaction time to external events, an explicit integration of robot goals in the planning Layer by using a Global knowledge database, and a modular design approach. Also, the Dialogue manager helps improving responses. Major components are; -

i. Perceptual Components
ii. Dialogue Manager
iii. Task-Oriented Components
iv. Global and Active Models
v. Learning Components

2.4 NiHA a cognitive agent

NiHA is an ongoing work in the development of self-aware artificial general intelligence [1]. The architecture is based on the QuBIC for machine consciousness. Hitherto, numerous cognitive characteristics have been introduced for making NiHA simpler to execute. Such characteristics comprise of imaginations [25], dreams, personal semantics, psycho-psychological based motivations, and ethics.

NiHA: The primary objective of this research was to look at the NiHA system limitations. To achieve this we have used the semantic memory model of the human brain. This framework consisted of short-term memory, working memory, long-term memory, and other low-level human cognitive parts segments. Cognitive Science Toolkit (CST)-[26] has been used as a guide for our work. We are also exploring our cognitive model, and how we are applying it. Our cognitive architecture prescribes the following kinds of memories.

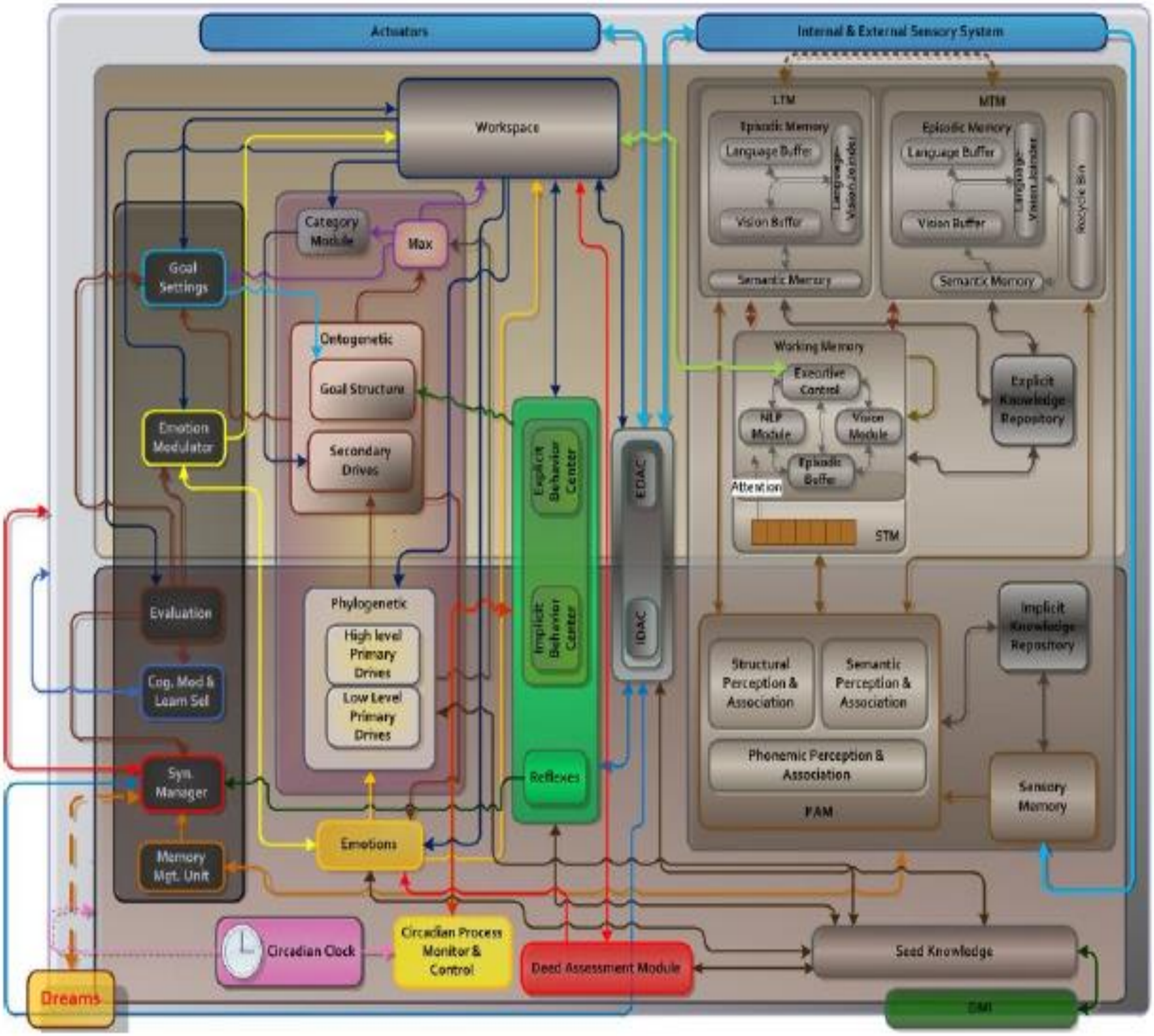


Figure 2-0-8: NiHA

2.4.1 Implementation of NiHA on iCub

The conceptual processes of NiHA are QuBIC-based. The thought processes were currently linked to an iCub robot simulator [71]. The iCub is built on a desktop mounted on an iCreate robot base and is fitted with the Microsoft Kinect controller. The iCub roBot's software platform (YARP) in C++ and YARP.NET is a YARP / YARP system wrapper of C#-code obtained after the YARP+ C++ application has been optimized using SWIG.

2.5 Cognitive Architectures Analysis

Each model of course has its strengths and disadvantages. For starters, every symbolic architecture needs a lot of work in creating an initial knowledge base, but the architecture is usable once it is completed. Emergent architectures are more quickly developed but have to be educated to generate usable behavior. Moreover, with the subsequent acquisition of new habits, their current experience will deteriorate. There are usually no constraints on how to pursue hybridization and numerous prospects. In addition to symbols, the structures may be categorized as single or multifunctional, heterogeneous or homogeneous, concerning the graininess of hybridization, the mixture of symbolic and sub-symbolic elements as well as the forms of memory and learning. The hybrid architecture does not, however, discuss directly what are considered symbolic and sub-symbolic components, and why they should be merged.

The fact that writers seldom discuss the types of representations used on their structures makes our analysis more complicated. Just a few of them consider these integrations to be important and address extensively, namely, ACT-R, CLARION, CogPrime, and GMU-BICA as shown in Table 2.1. In the absence of such fine detail in certain papers, symbolic or sub-symbolic components cannot be defined for all the systems studied and we, therefore, concentrate on representation and processing like sensory modules, memory implementation, attention, etc.

2.5.1 Sensory Modules Implementations

There are different types of sensory modules implementation by cognitive architectures as shown in Table 2.1. Like NiHA's visual sensors implementation use both simulation and physical Sensor. Hearing and Touch sensors are implemented only through physical Sensors. And data input by simulation. Whereas some only use physical sensors like iCub, and others use combinations of simulation or physical sensors. There are multiple architectures like GMU-BICA that only uses simulations.

Cognitive Architectures	Vision	Hearing	Touch	Smell	Data input	Other sensors
NiHA	S & P	P	P	NI	S	NI
LIDA	S	NI	NI	NI	NI	P
ACT-R	S & P	S	NI	NI	NI	NI
BECCA	S & P	NI	NI	NI	P	NI
CLARION	S	NI	NI	NI	NI	P
Epic	S & P	S	S	NI	P	NI
GMU-BICA	S	NI	NI	NI	NI	NI
SOAR	S & P	NI	NI	NI	P	P
iCub	P	NI	P	NI	P	NI
RoboCog	P	P	P	NI	NI	NI

Table 1: Sensory Modules Implemented in different Cognitive Architecture

Not Implemented	NI	Simulation & Physical Sensors	S & P
Simulation	S	Physical Sensors	P

Table 2: Index for table 1

2.5.2 Memory Units in Cognitive Architectures

Memory is an integral part of the cognitive proposed scheme irrespective of whether it is used to research human consciousness or to overcome problems in engineering. Almost all the architectures included in this evaluation have the memory systems that store intermediate analysis details, allow learning and adapting to the evolving environment. But despite their practical similarities, such memory systems architecture varies significantly and relies on research priorities and technical constraints, such as biological plausibility and engineering constraints as shown in Table 2.2.

Yet despite their practical similarities, such memory systems architecture varies significantly and relies on research priorities and technical constraints, such as biological plausibility and engineering constraints.

Memory is characterized in terms of its length (short-term and long-term) and form (semantic, declarative, procedural, etc.) in the cognitive architecture theory, but it is not generally applied as distinct information stores. In psychology, this view of memory is dominant, but its relevance for engineering is disputed by some since it does not present a comprehensive explanation of different processes of memory. However, most architectures can distinguish between different forms of memory, although naming conventions can vary. Architectures designed for planning and problem solving, for example, include short- and long-term memory retrieval structures that do not use cognitive science terms.

Cognitive Architecture		**NiHA**	**LIDA**	**ACT-R**	**BECCA**	**CLARION**	**Epic**	**GMU-BICA**	**iCub**	**RoboCog**
Sensory memory		[1]	[27, 28]	[29]	×	[30]	[31]	×	×	×
Working Memory		[1]	[27]	[32]	[33]	[30]	[34]	[35]	[36]	[37]
Long-Term Memory	**Semantic**	[1]	[27]	[38]	[33]	[39]	[40]	[35]	×	[41]
	Episodic	[1]	[27]	[38]	[33]	[42]	×	[35]	[43]	×
	Procedural	[1]	[27]	[44]	[33]	[45]	[40]	[35]	[43]	[41]
Global Memory		×	×	×	×	×	×	×	×	[37]

Table 3: Memory Units in Cognitive Architectures

- **Sensory Memory**: The goal of sensory memory is to store and pre-process the incoming sensory data before passing it to other memory structures.
- **Working memory**: Working memory (WM) is a system for the temporary storing of current job-related information. It is important for cognitive skills such as concentration, thinking, and understanding, so it has been incorporated for each cognitive architecture.
- **Long-term memory**: Long-term memory (LTM) retains for a very long period of time a lot of information. It is usually split into two categories on the basis of their representation of information.
    - ***Declarative memory***: contains explicit knowledge and further subdivided into
        - semantic Memory
        - episodic memory
    - ***Procedural memory***: continuing implicit knowledge (like motor skills and routine behaviors)
- **Global Memory:** Global memory is used by those frameworks which do not have distinct representations for various forms of knowledge to preserve all information in the system using a common framework.

2.5.3 Visual Processing

Most architectures in all visual processing processes involve robotic analysis, biologically influenced, and biomimetic design. They are physically organized. Although, visual processing is a vast area and it is a costly undertaking to transform visual experience into functional unstructured environments in order to collect all visual information. The perceptual, psychological, or philosophic dimensions of the design of the intellect of humans of much greater significance, whereas the specifics of technological application are either lacking or absent. Most architectures in all visual processing processes involve robotic analysis, biologically influenced, and biomimetic design. They are physically organized as shown in Table 2.3.

- **Features**: Initial vision typically requires edge identification and variance calculation.
- **Proto-objects**: Those functions are then clustered into blobs with identical characteristics.
- **Objects**: Features are then converted into claimant objects with centroid coordinates.

- **Object models**: Used for off-line machine learning methods.
- **Object labels:** We then categorize and mark candidate objects.
- **Spatial relations**: Displays semantic picture knowledge. The distribution of the space may not only indicate the location of the object but also define structural details among objects.

Cognitive Architecture	NiHA	LIDA	ACT-R	BECCA	CLARION	Icub	RoboCog
Features	[1]	[46]	[47, 48]	[49]	✗	✗	[50]
Proto-objects	[1]	[46]	[47, 48]	[49]	✗	[51]	[50]
Objects	[1]	[46]	[47, 48]	✗	✗	[51]	[50]
Object models	[1]	[46]	[47, 48]	✗	✗	✗	[50]
Object labels	[1]	[46]	[47, 48]	✗	[52]	[51]	[50]
Spatial relations	[1]	[46]	[47, 48]	[49]	[52]	[51]	[50]

Table 4 : Visual Processing in Cognitive Architectures

2.5.3.1 Spatial relations

Spatial relationships such as left, right, on, riding, eating, and wearing can give you the ability to locate the object in a picture or to connect to other objects. The use of image-based forecasts to enhance the depiction of problems and the relationship between cognition and action. Three distinct representations support spatial and visual cognition as shown in Figure 2.9 [53]

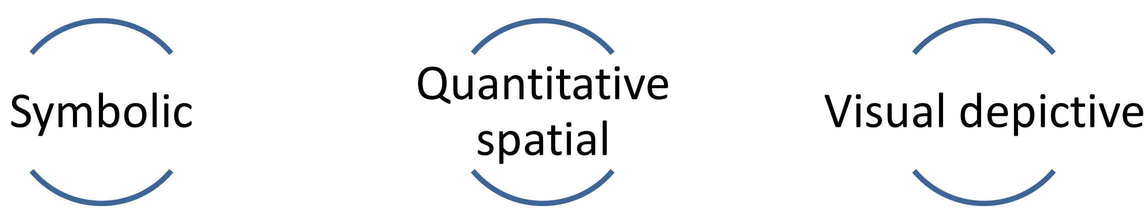


Figure 2-0-9 : Components of Spatial Relations

- Symbolic:
    - Object identities
    - Qualitative spatial and visual properties
    - Non-perceptual information
- Quantitative spatial:
    - Object labels
    - 3D Spatial Properties(explicit)
        - General shape
        - Location
        - Orientation
    - 3D Spatial Properties(implicit)
        - Size
        - Topology
        - Direction
        - Distance
- visual depictive
    - Object labels
    - 2D Visual Properties(explicit)
        - Shape
        - Texture
        - Empty space
    - 2D Spatial Properties (implicit)
        - Location
        - Size
        - Topology
        - Direction

This approach helps the system to gain visual reasoning precision and reliability while retaining control of neural core processes. The processing comes from a seamless mix of multiple representations, which leverages each system's basic efficiency and capabilities.

2.5.4 Humanoid Robot & Cognitive Architectures

Different computational systems have introduced their cognitive abilities with humanoid robots. In the mental purpose of each respective system, which is to benefit from the visual experience and natural speech expression, cognitive components like Memory System, activity behavior, concentration, collection of actions are used. In addition, it is also possible to learn and display certain feelings and actions as shown in Table 2.

Cognitive Modules	**NAO**			**iCub**		**Maggie**
	CAIO [37]	***DAIM [38]***	***Project ROSE [39]***	***iCub Cognitive Architecture [35]***	***SAM [40]***	***Maggie Architecture [41]***
Sensors and Sensory Memory	• Audio sensors • NAO Camera • Multimodal Perception	• Audio Activity • Visual Activity	• Audio Activity • Visual Activity	• Vision • Touch [4]	• Vision • Audio • Touch • Proprioception	• Vision • Audio • Tactile Touch • Proprioception • Other
Cognitive Memory	• Episodic • Semantic • Procedural	• Perception	• Perception • Episodic • Semantic • Procedural	• Perception [4] • Working Memory [26] • Procedural [27] • Episodic [27]	• Perception	• Long Term Memory • Working Memory
Action behaviors	• Emotional Appraisal • Planning and Scheduling • Emotional Multimoda	• Reactive behavior • Verbal Communicate • Sheared Space	• Reasoner • Symbolic Task Planner	• Gaze control • Compute optical flow	• Verbal Communicate	• Emotional Supervisory System • Drives

	l Action Renderer					
Attention	• Deliberation	• Virtually access characteristics	• Human Head pose • Weighted attention map	• Stabilize gaze • Detection of biological motion	• Human Moment	• Visual Attention
Actuators Motion	• Physical Actuators • Communicative Actuators	• Move Delay • Move Type • Swiping its arm	• Move Base(DMP) • Open Gripper • Move Arm	• Hand Motion • Head Motion [4,27]	•	• Hand Motion • Head Motion
Learning	• Emotions	• DAIM Learning Module • Game Logic	• Natural Language Communication	•	• Human Face • Human Action	• Natural Language Communication

Table 5 : Humanoid Robot & Cognitive Architectures

2.5.5 Object Detection

Deep learning includes procedures like identification of objects with an image, video, or web camera feed in the sub-discipline known as 'Object detection'. Object detection is amongst the most complex and crucial aspects of computer vision, often used in people's daily lives, Object detection is capable of providing useful information for semantic image and video interpretation and is linked to many implementations including image classification, security and surveillance, social cognition analysis, autonomous vehicles, facial recognition, visual perception, and so on, with the aim of locating the instances of semantics objects of a certain type. The distinction between object detection algorithms and classification algorithms is that we seek to draw a boundary box for the object in the picture in detection algorithms. Often, in a case of object detection you may not draw just one bounding box, there might be other bounding boxes that display the various objects of interest in the image and you do not know how many previously.

Object detection has previously been implemented using basic techniques to suit the prototype. In this manner, target objects are cropped and features created using different descriptors such as HOG and SIFT. The method then used a sliding window in the picture and compared each area with the object function vector database. Enhanced algorithms, such as SVM classifiers, have been used to eliminate the use of these databases. Because objects of different sizes are being used, people have been using different window sizes and picture image sizes. These complex pipelines partially resolved the issue of object detection but still had several downsides. The pipelines were highly time-consuming in terms of computation, and the hand-designed features incorporating techniques such as HOG and SIFT were not very precise and accurate.

With the emergence of fundamental machine-vision processing, the efficiency of object detectors has been significantly enhanced. The rapid growth of more efficient deep learning frameworks helps to learn semantic, high-level, deeper features, which are designed to overcome problems in conventional architectures. Researchers have started to look at deep-learning approaches to solve object detection problems with the groundbreaking discovery of these algorithms.

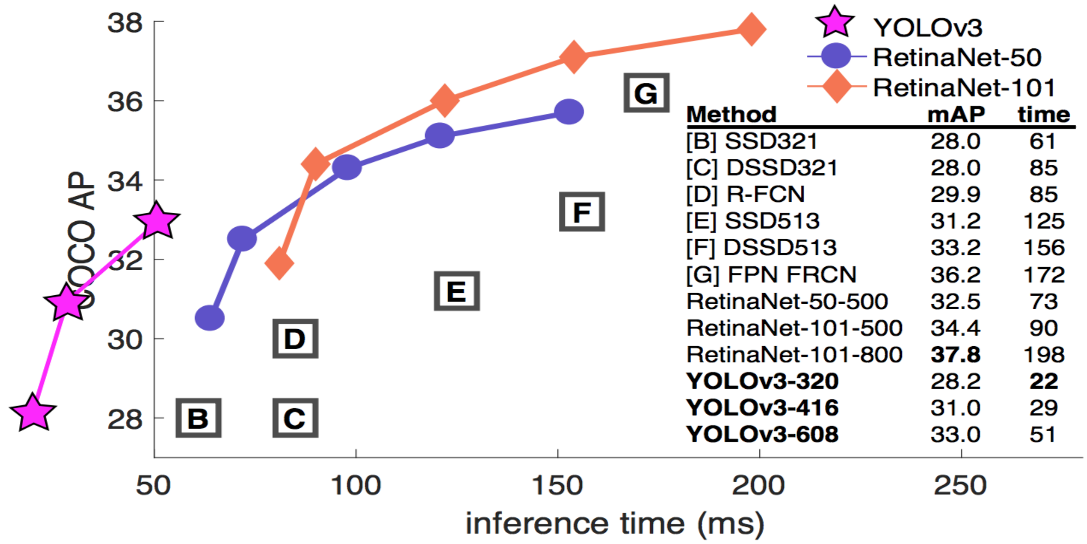


Figure 2-0-10: YOLO v3 Performance

So far, too many techniques have been created for object detection but YOLO has outperformed all as shown in Figure 2.10. The YOLOv3 pair on the COCO Dataset of Microsoft were done in tandem with the versions RetinaNet and SSD, which indicates the power of the layout to match the boxes to objects. However, if the IOU level decreases, the model fails to correctly match boxes with objects. Redmon and Farhadi are claiming the model does not fit well in the 0.5 to 0.95 IOU metric average AP, but does well on a 0.5 IOU threshold metric. This functions best for small objects than for large objects.

2.5.6 Analysis

After the detailed analysis of the research that has been done cognitive architecture, NiHA seems to improve many aspects of cognitive agents in terms of Sensory modules & Memory implementation including some other features as attention selection, emotions, goal setting, seed knowledge, etc. Furth help is available and supporting our motivation for industry 4.0 in Pakistan.

In the current scenario of visual perception and **object spatial relation** a requirement of cognitive architecture in humanoid robots has been established, which can:

- *Interact in a human-social environment*
- *Recognize objects*
- *Identify objects co-occurrence*

CHAPTER 3: METHODOLOGY PROPOSED

NiHA is based on past works involving the ' Quantum & Bio-inspired Intelligent & Consciousness Architecture (QuBIC)' unified theory (see Figure 2) and an agent named Juhi. [15]. NiHA cognitive infrastructure is designed in layers (see Figure 2). The first layer is made up of physical components, the second layer of conceptual processes. The sensors and actuators are solid components.

Therefore, the behavioral system includes unconscious and conscious stages. The unconscious layer includes a number of cognitive units that work together to control unintended and pre-designed tasks that are essential elements for self-regulation and the optimal output of the agent. Awareness, concentration, intention and cooperative actions are responsible for the components of the aware network. The conscious framework helps to control the mind and to define the system functionally.

In order to establish a matric sensory level of perception which leads to the symbolic level of semantic representation, the current method is being proposed. The following workflow representation is given by NiNA as shown in Figure 3.1, a mixture of NiHA and NAO. First Enabling NAO to provide a visual stream for Sensory-motor Memory that will prepare data for perceptual associative memory to process.

3.1 NINA's Work Flow

- ***Sensory Motor-Memory:*** Sensory Memory retains sensory signals from both external and internal receptors. In the implicit information archive, sensed contents are stored. Sensory mechanisms implement fundamental sensory material philters
- ***Perceptual Associative Memory:*** The interaction between subjects is focused on identification, interpretation, and assessment of the perceptual associative memory (PAM). Information is encoded in structural and semantic analyses for further analysis through collaboration between STM and WM
    - Object Detection
    - Object Environment Relation

- ***Working Memory:*** PAM transmits this information to a working memory through encoding in the form of percepts.
  - Correlation of Objects
  - Visual Scene Descriptions
  - Scene Graph Creation
- **Procedural Memory:** Receive commands in the form codlets from WM and select required actions to perform and generate values for NAO movement, completing NiNA's workflow.

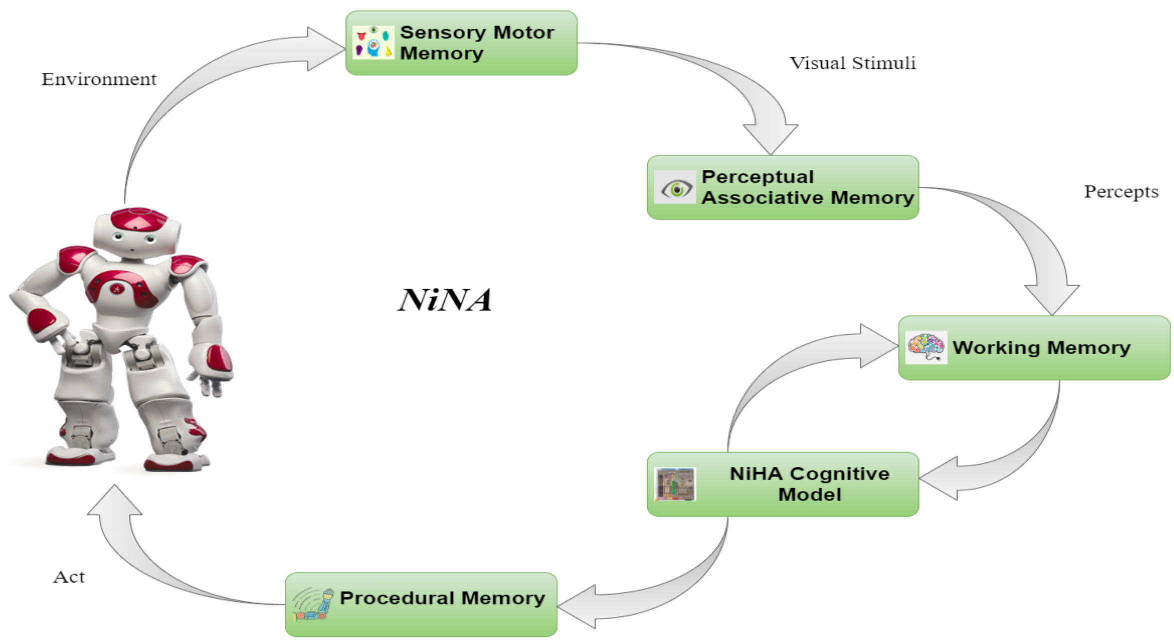


Figure 3-0-1: NiNA Workflow

3.2 Sensory Motor Memory

Sensory Memory is responsible for the storage of sensory inputs coming from external and internal sensors. Sensed contents are collected in the implicit knowledge repository. Sensory processes apply basic filters to the sensory contents. The filters include those responsible for resizing the incoming image stream, encoding the stream as required, and applying part-of-speech-tagger on the lingual contents. This sensory information is then transferred to various unconscious units.

Sensory memory has been the raw collection of information from visual, auditory, touch, and other sensory modalities within a time period that typically snaps over anything like 50-500ms [54]. Typically this memory contains un-interpreted data that is used for the initial phase of perception. There have been at least two categories of sensory memory, the classic memory, powerful visual pattern stimulus, and the echoic memory, retaining auditory stimuli, although other memories may also exist for other senses, not so commonly studied.

Sensory memory is a pre-processing module that Receives visual feed from NAO "**ALVideoDevice**" and converts it using the *OpenCV frame* and adjusts the resolution according to the YOLO v3 from 640 x 480 pixels into 416 x 416 for better performance?

Next, create a *Queue* of Images from which it takes a single frame based on (FIFO) to either forward it to PAM or discard. Before sending it to PAM it checks if PAM is free or not. IF **PAM is free** then send to PAM else check the image frame **Timestamp** if 20s buffer is up it discard the image frame. And take the next image frame from the queue.

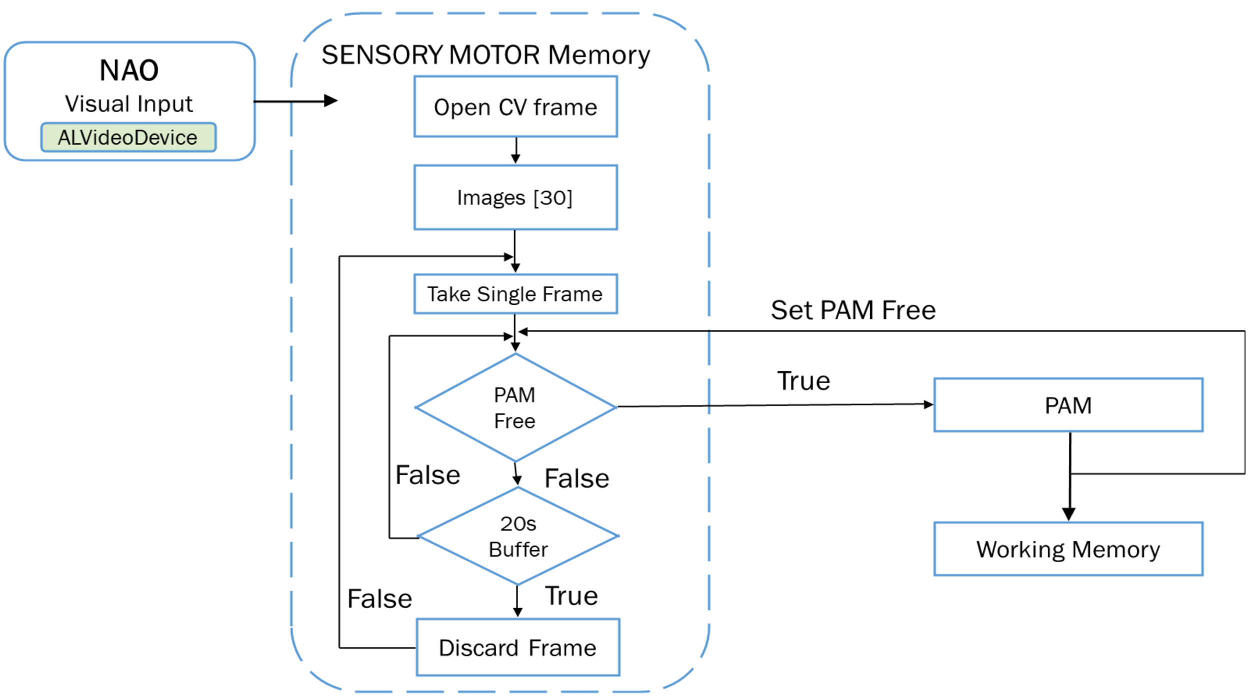


Figure 3-0-2: Sensory Motor Memory

If the **20s buffer** is not up then it checks again if **PAM is free** or not. This continues until either PAM is free or the time limit expires. As soon as PAM completed it processing **PAM** sends percepts to Working memory and set status free as shown in Figure 3.2.

3.3 Perception Associative Memory

The Perceptual Associative Memory (PAM) is presented with a duplicate of the stored sensory information. PAM is responsible for establishing a relationship between identification, interpretation, and classification-based objects. For further processing, memory is encoded through systemic and semantical analysis and co-operation between Short-Term Memory (STM) and Working Memory (WM). The percepts are then passed to the conscious memory system to interpret it further.

The perceptional memory is the recollection containing different types of objects that a perceptual system can perceive. It contains various objects, properties, and patterns that a perceptual system can categorize. Every perceptual memory instance is a depiction of a classification used during the perception.

Perceptual associative memory, which is, the capacity to perceive incoming stimuli by identifying instances, categorizing them, and marking the associations between any of these entities and categories, is pervasive among animal species, and so is the awareness of such capabilities [55]. Perceptual learning happens quickly and easily, but declines according to an inverse sigmoid function; fresh, immature memories degrade exceptionally fast, while mature perceptual memories can last for several decades Preconscious observation is the first step in a constantly cascading sequence of cognitive processes in which each sense and acts on its environment [19].Perceptual associative memory working with respect to NiNA is shown in figure 3.3.

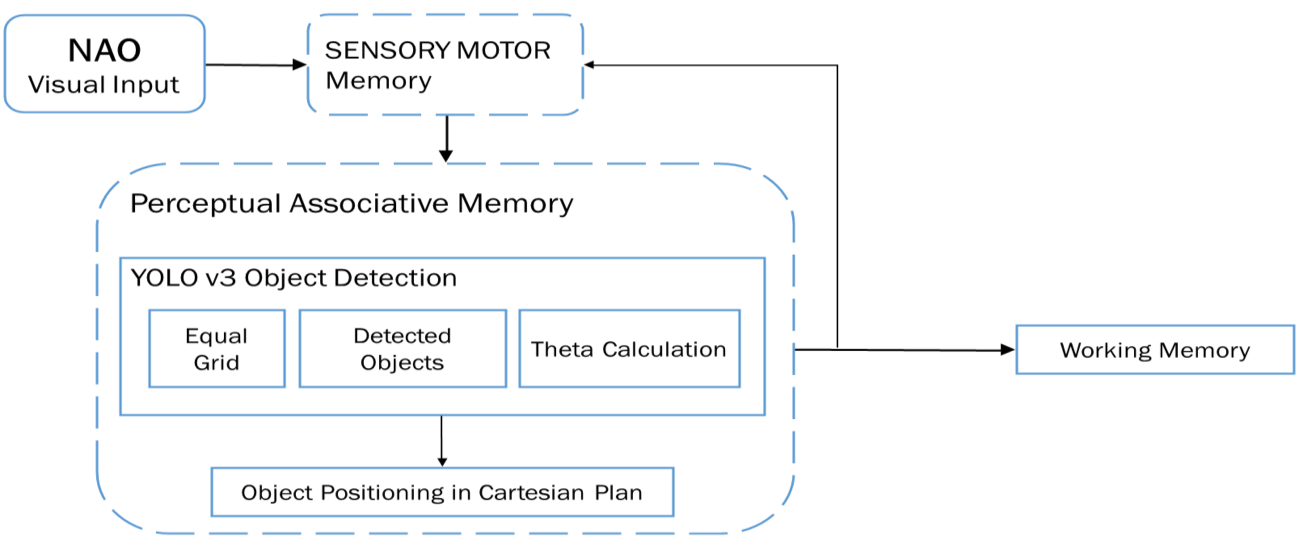


Figure 3-0-3: Perceptual Associative Memory

It receives signals from Sensory-motor memory and implements **Yolo v3 object detection algorithm** with following modification. First, convert the image frame into an **equal grid** and place the center at **(0, 0)** then get the **detected objects** list and their location in the image frame as shown in figure 3.4.

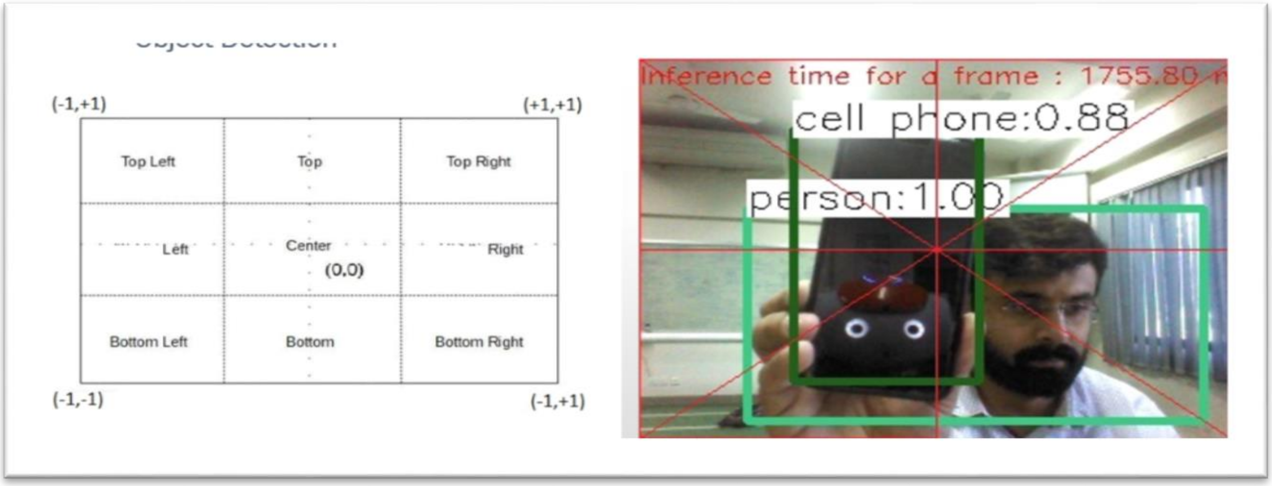


Figure 3-0-4: PAM Object Detection

Using an equal grid system **calculates theta** and creates a **Cartesian-Coordinate System**. Reconfiguring each object's position based on their thetas in the Cartesian-Coordinate System as shown in Figure 3.5. This will generate a spatial relationship of detected objects with NAO's field of vision. Now the robot will be perceiving each object as placed at its left or right. After that, encode information in the form of percepts and forward it Working memory and set PAM free flag true.

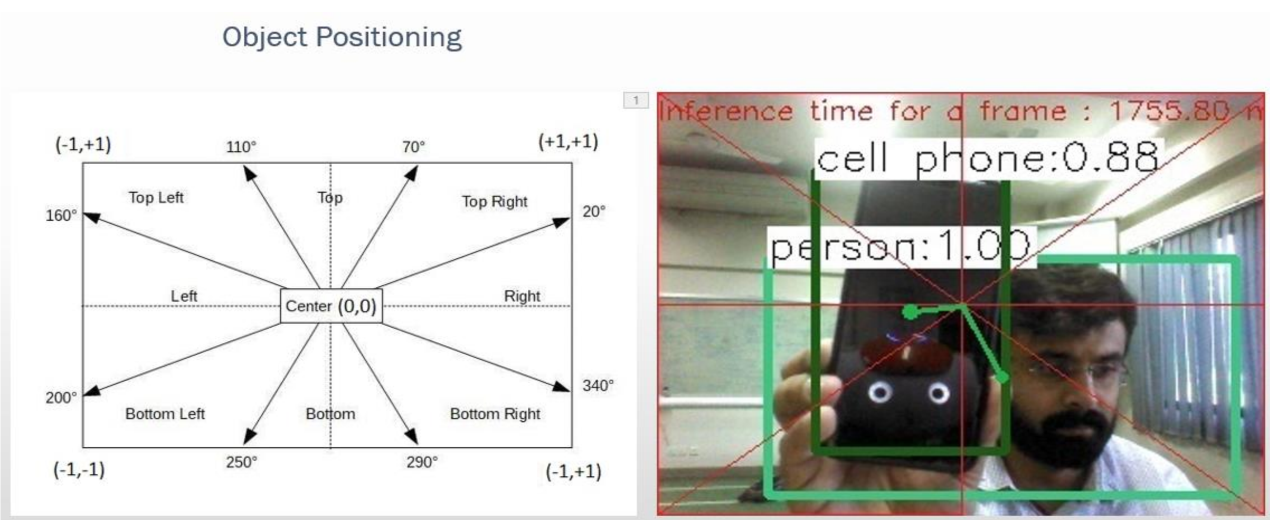


Figure 3-0-5: PMA Cartesian-Coordinate System

3.4 Working Memory

Throughout the learning cycle, Working Memory (MM) is responsible for interpreting and refining information. In NiNA WM acts as a central executive assisted by visual memory and spatial memory (with spatial maps and knowledge about the output size and location) as Shown in Figure 3.6.

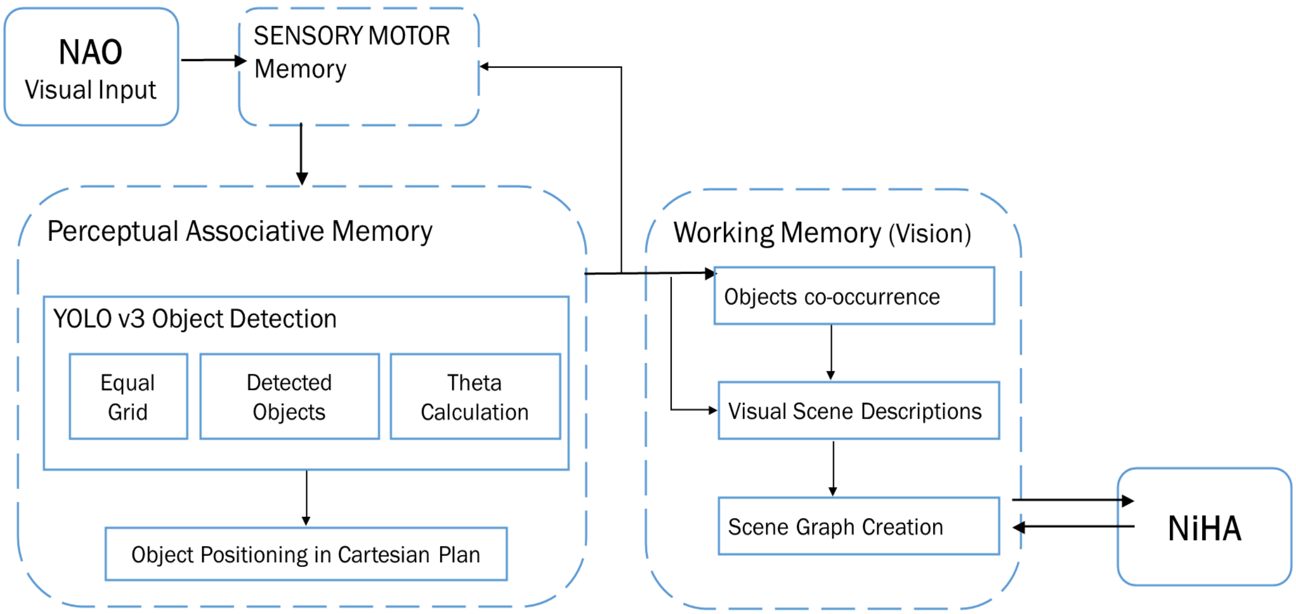


Figure 3-0-6: Working Memory

On receiving percepts from PAM **working memory** analysis spatial relation of each object with other objects and identifying **objects co-occurrence** as shown in Figure 3.7. These analyses are sent to generate the **visual scene description** with the help of some pre-defined phrases as shown in Figure 3.8.

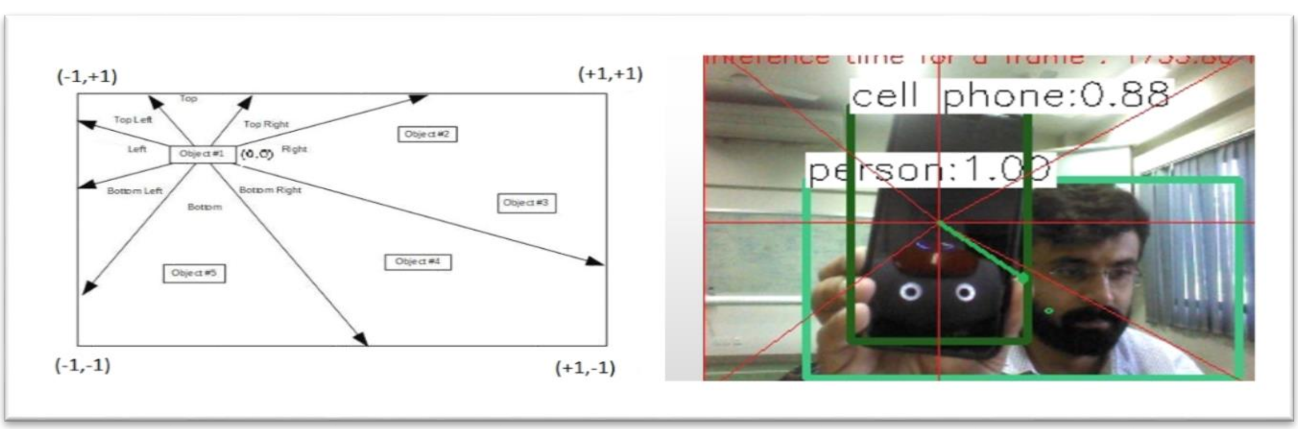


Figure 3-0-7: WM Object Co-occurrence

Figure 3-0-8: WM Visual scene description

These phrases then pass to the graph generation module to convert the visual scene description into the **scene graph** for the inner workings of NiHA's memory system as shown in Figure 3.9. Further, these pieces of information (codelets) use for selecting and generating action in procedural memory.

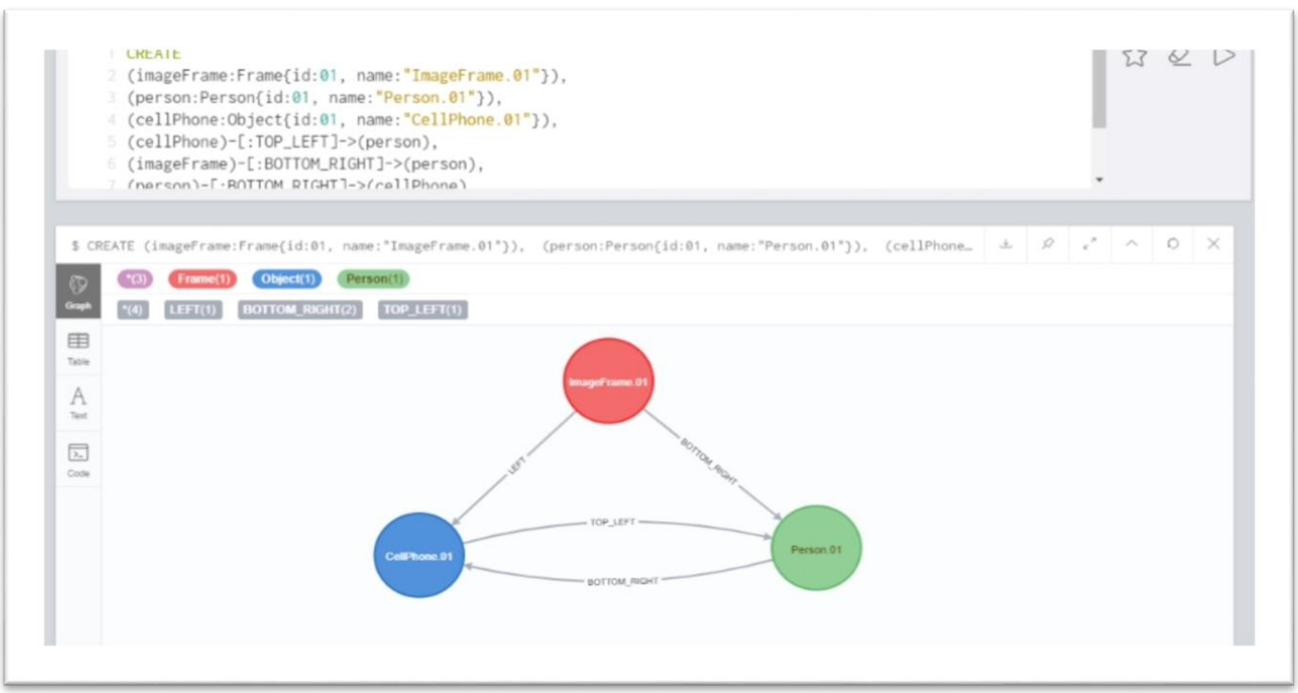


Figure 3-0-9: WM Scene Graph

3.5 Procedural Memory

The procedural memory is a recollection of an agents' behaviors and actions. This is a non-declaratory memory that refers to a kind of "how-to" knowledge that is typically a record of potential engine and actions. The functional principles are the standard representations of memory artifacts in procedure memory. [26]

It helps for selecting actions like move RightAram to left and point (open hand) then generating appropriate joint angle, stiffness, and speed of motion according to the selected action. These are sent to NAO to complete these actions using “ALMotionProxy” and speak through “ALTexttoSpeechProxy”.

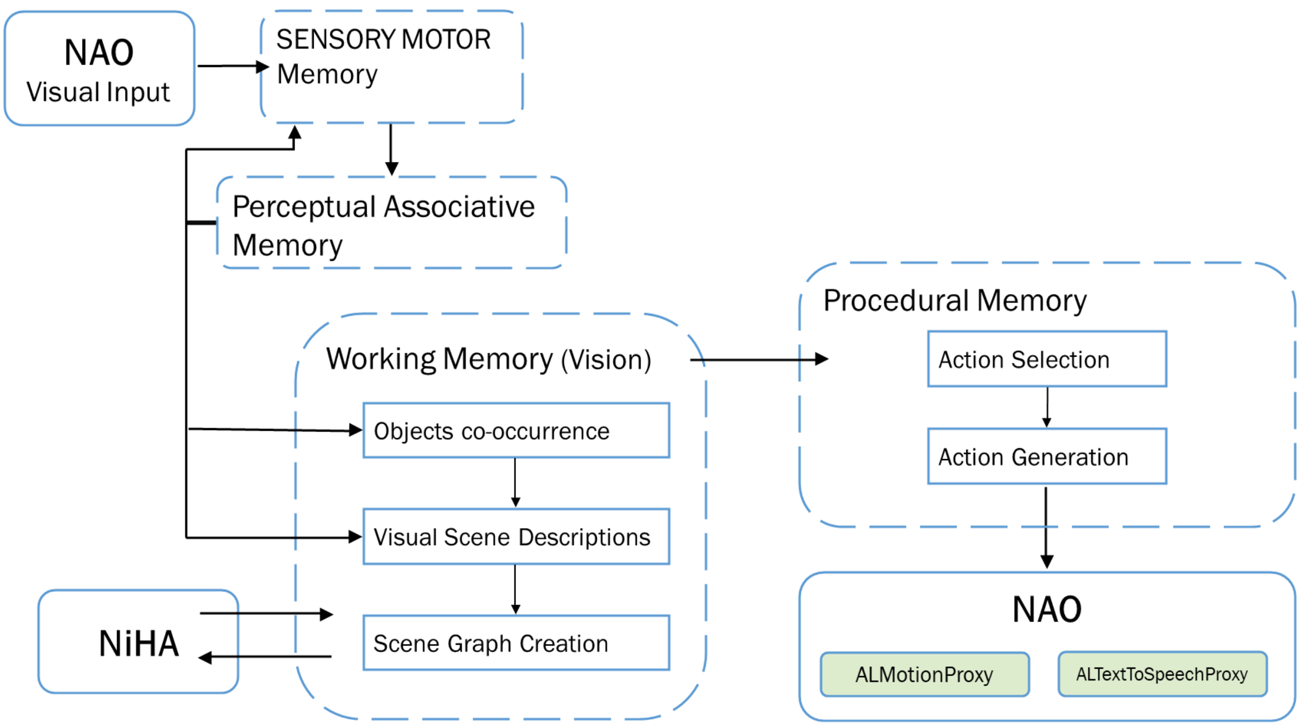


Figure 3-0-10: Procedural Memory

3.6 Implementation

For the implementation of these, I have used NAOqi 2.4.3 C++ SDK as it provides build on remote computers as well as on NAO too. Although for initial testing implemented C#, java, and ROS SDKs as Shown in Table 5.

NAOqi 2.4.3				
Programing Languages	Bindings running on		Choreograph support	
	Computer	Robot	Build App	Edit Code
C#	✓	×	×	×
Java	✓	×	×	×
ROS	✓	×	×	×
C++	✓	✓	×	×

Table 6: Naoqi Programing Languages

First Feature extraction & Object detection was implemented through **NAO object detection** module but left because you have to manually extract features and mark using NAOqi then name each object still different orientation of object was not detected.

Secondly, **point cloud library's** range function use for 3D mapping of objects although, it gives a depth map of the environment for object detection we require other algorithms. Next, we use **OpenCV stereo vision** which was not compatible with the NAO vision sensor. Then we use 2D Object detection of YOLO v3 on the COCO dataset as shown in Image 3.11.

Figure 3-0-11: YOLO v3 Object Detection

3.6.1 COCO Dataset

There are various ways to obtain a perception of the environment, including visual perception, and NAO robot also needs visual information through its camera. Visual Perception requires to process visual stimuli, obtain objects, and to get categories. For that purpose we have compared many common datasets; analyzed the properties of the MS COCO dataset collection with ImageNet; CIFAR10, and PASCAL VOC 2012. Each of these datasets differs considerably in size, list of classified groupings, and image categories. **Coco Dataset** is being used here because it has 80 common use types of the class although Image net has more of a class when it comes to detected objects COCO dataset had a clear lead as shown in Figure 3.12. And better ration of class vs detected object.

- **ImageNet** was intended to collect and capture a vast number of object classes, many of them are fine-grained.
- **PASCAL VOC's** core purpose is the identification of artifacts in natural images.
- **Cifar 10** This dataset has been created by Krizhevsky, Nair, and Hinton
- **MS COCO** is intended to identify and segment objects that occur in their natural context.

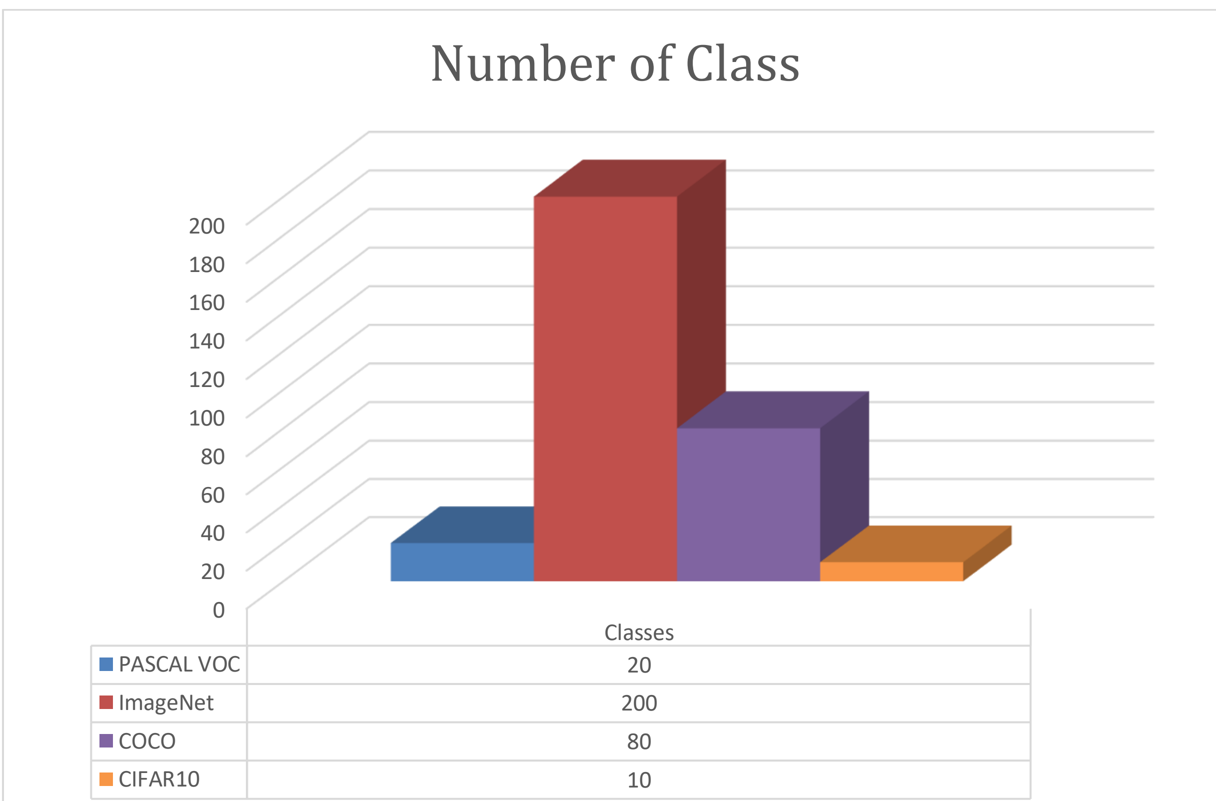


Figure 3-0-12: COCO Data Set Comparison (Number of Class)

Furthermore, each dataset has a different number of classes, Training and validation Images, Testing Images, and Detected objects. Such that SUN has the most number of classes 397, then Image Net (200), COCO (80), and least PASCAL VOC(20). But for training the Machine Learning model and testing Image Net, a huge Dataset is provided containing 516,840 total number of Images, then comes COCO dataset with 328,124, and lastly PASCAL VOC with 22,531. As compared to a huge dataset and a large number of classes, MS COCO has a great number of Detected objects (886,284) on its dataset, excluding test data which greatly affects its authenticity as compared to Image net with 534309; PASCAL VOC with 27,450. That analysis greatly helps us to decide that the COCO dataset will be more suited for our visual perception module because it provides objects identification, including segmentation, the right kind of information that vastly helps us to generate association for visual perceptual memory.

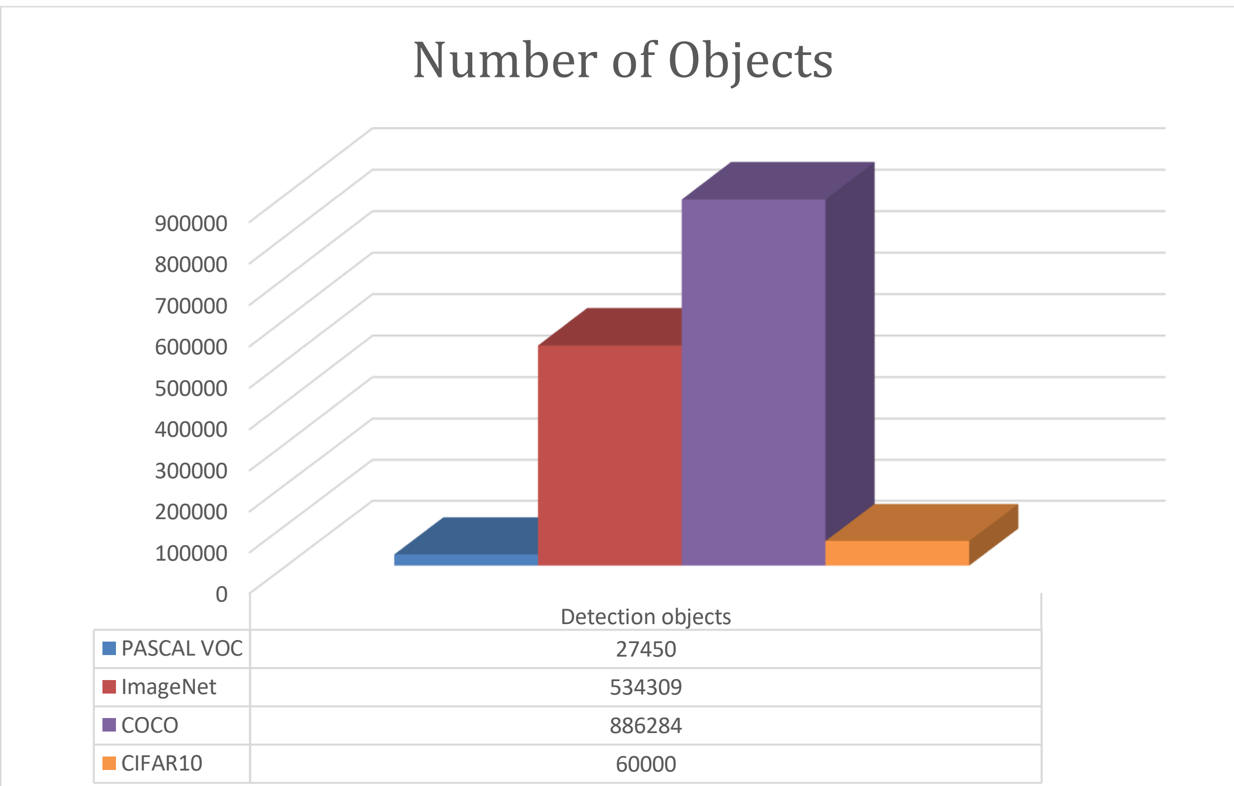


Figure 3-0-13: COCO Data Set Comparison

CHAPTER 4: RESULTS & LIMITATIONS

4.1 Object Detection

YOLO v3 pre-trained model is used for Object detection and it gives us an accuracy of **72 %** in our RISE Lab environment. And achieving average precision of **91%** as compare to **YOLO v3 having 57%**, **FPN FRCN 59%**, and **RetinaNet 61%** it is due to our control environment and YOLO best performance on Medium size object.

4.1.1 Accuracy = 72%

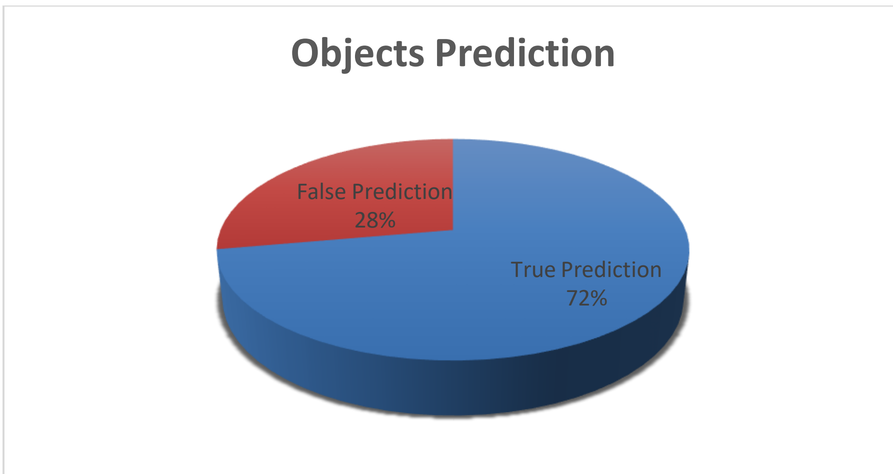


Figure 4-0-1: NiNA Object Detection Accuracy

4.1.2 Average Precision = 0.91

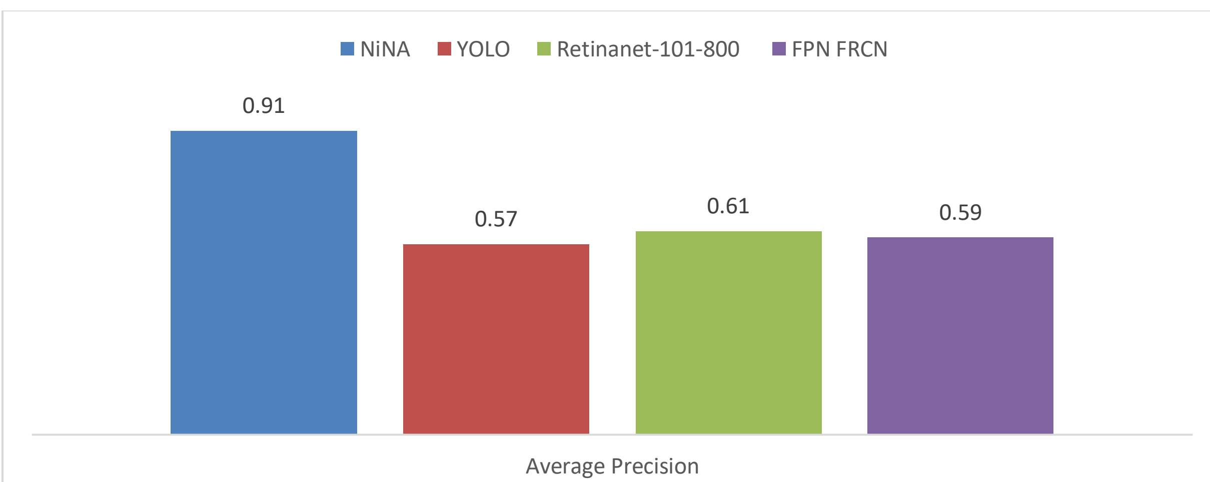


Figure 4-0-2: NiNA Average Precision

4.1.3 F1 Score = 0.82

Selected some of the most commonly found objects in our RISE Lab and perform our experiments on them. And calculated **F1 score** because Classification Accuracy alone cannot be trusted to select a well-performing model when classes are imbalance.

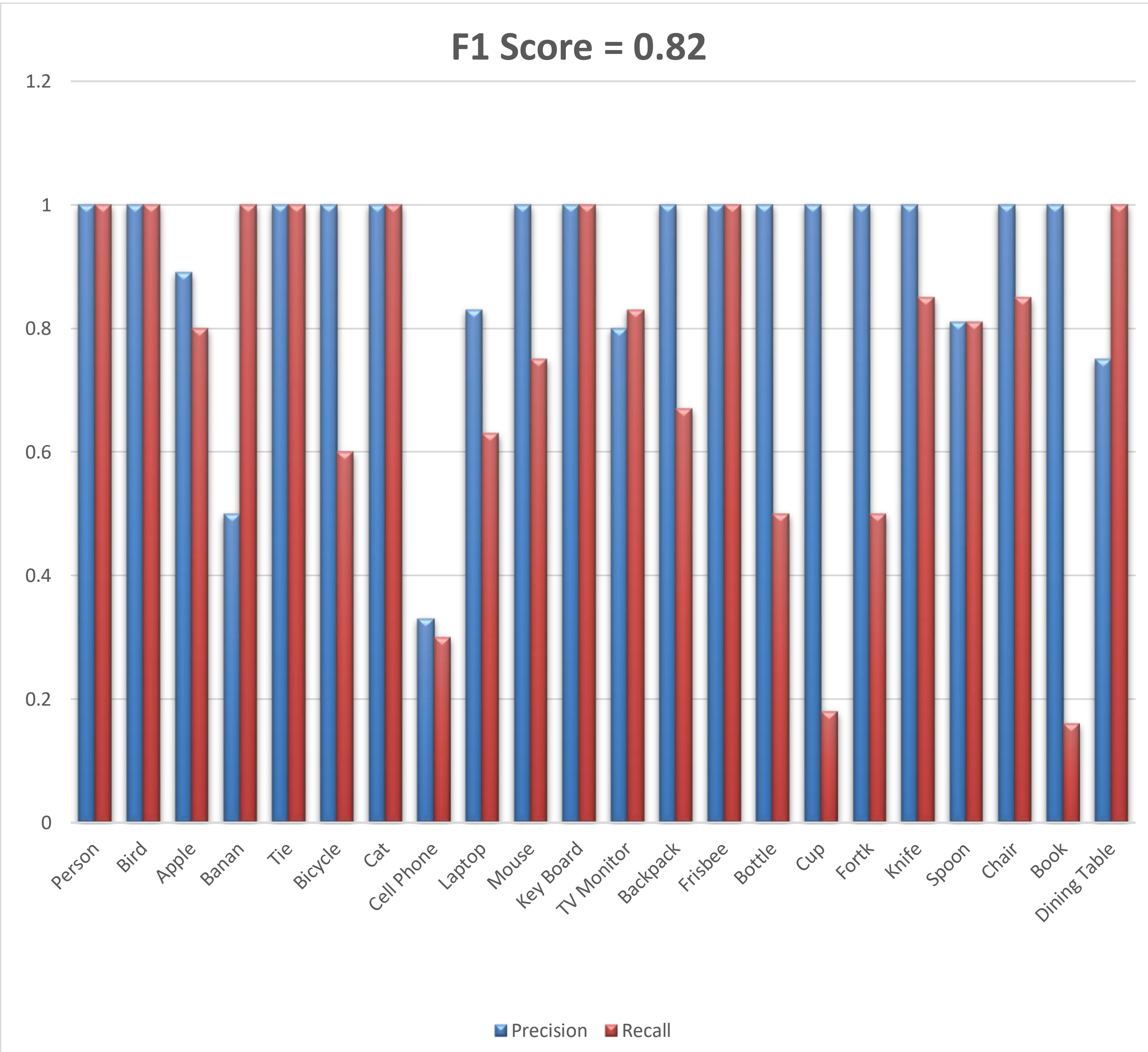


Figure 4-0-3: NiNA F1 Score

4.2 NAO Pointing Objects

4.2.1 Accuracy = 78%

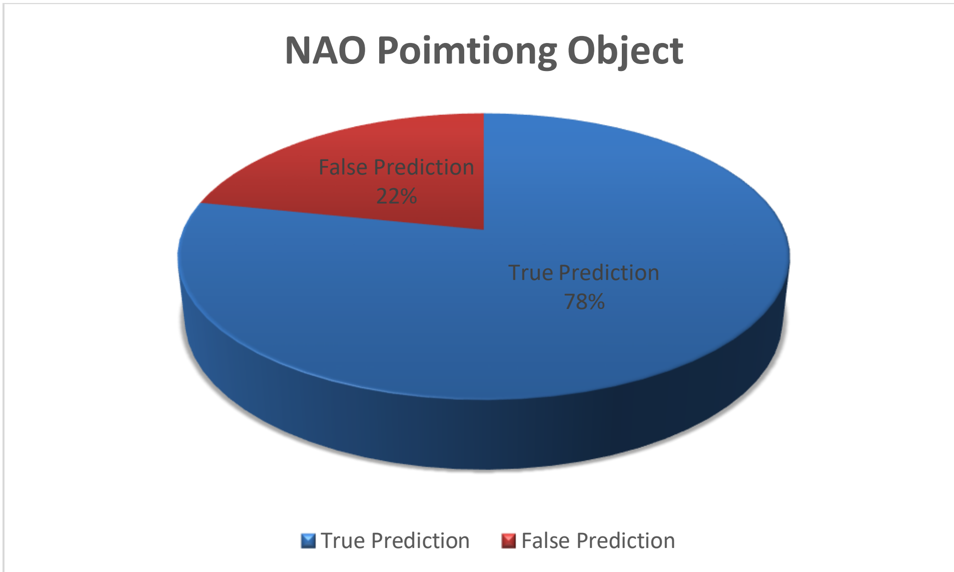


Figure 4-0-4: NiNA Pointing Objects Accuracy

4.2.2 *F1 Score = 0.78*

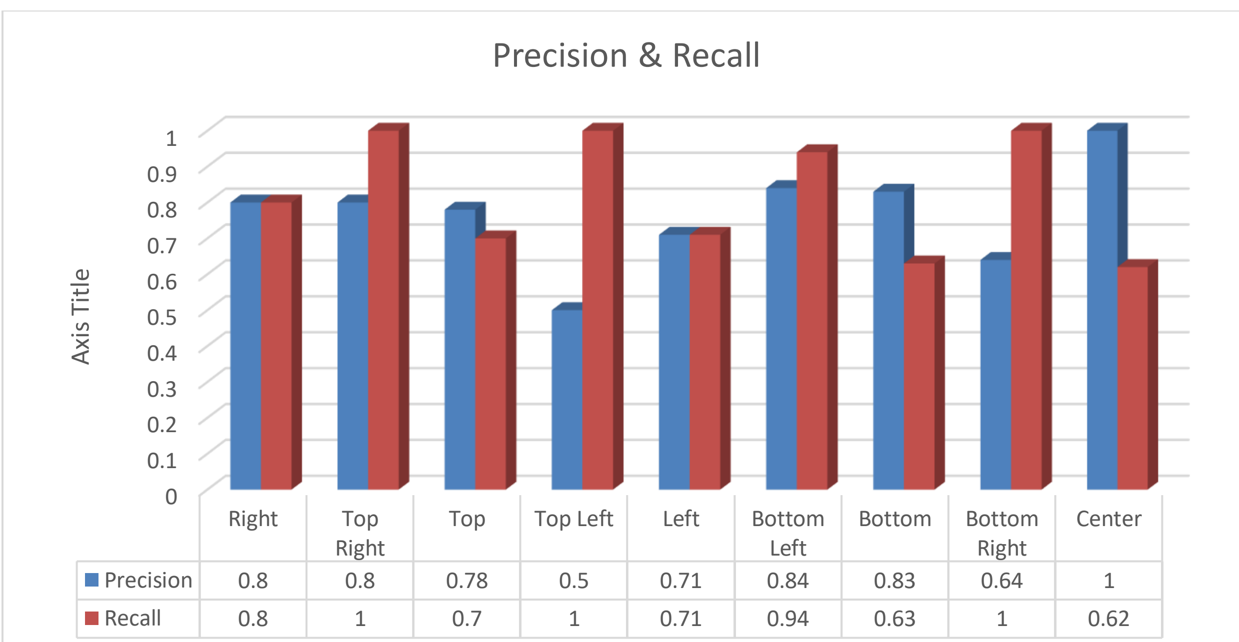


Figure 4-0-5: NiNA Pointing Objects F1 Score

4.3 Object Co-occurrence

For spatial relationships of objects co-occurrence, we got Accuracy of 85% which is higher than the test accuracy of research **"Identifying Spatial Relations in Images using Convolutional Neural Networks** [56]**" = 68.98%** and "**Visual Relationship Detection with Language Priors** [57]" = **70%.**

4.3.1 Accuracy = 85%

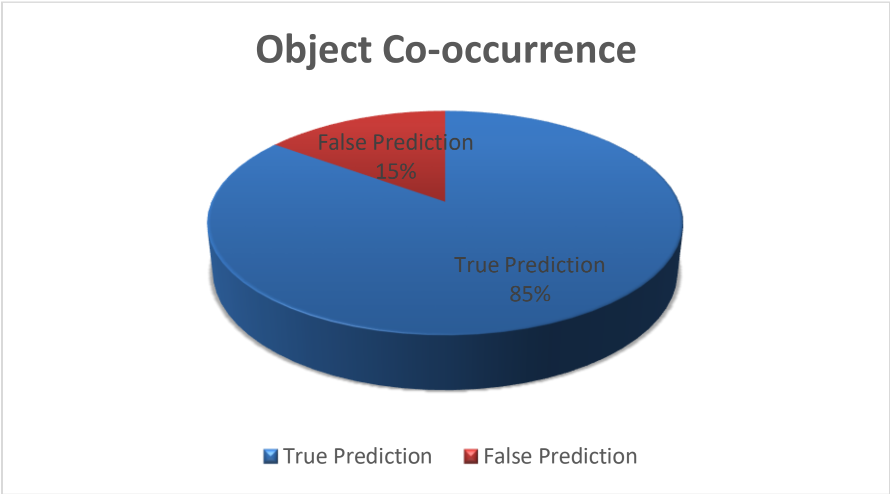


Figure 4-0-6: NiNA Object Co-occurrence Accuracy

4.3.2 **Accuracy of Identifying Spatial Relations in Images using Convolutional Neural Networks [56]**

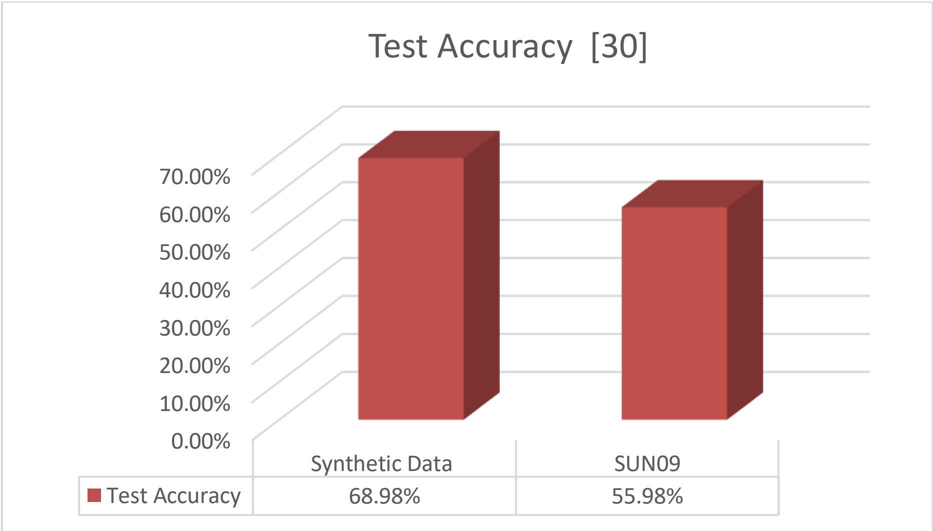
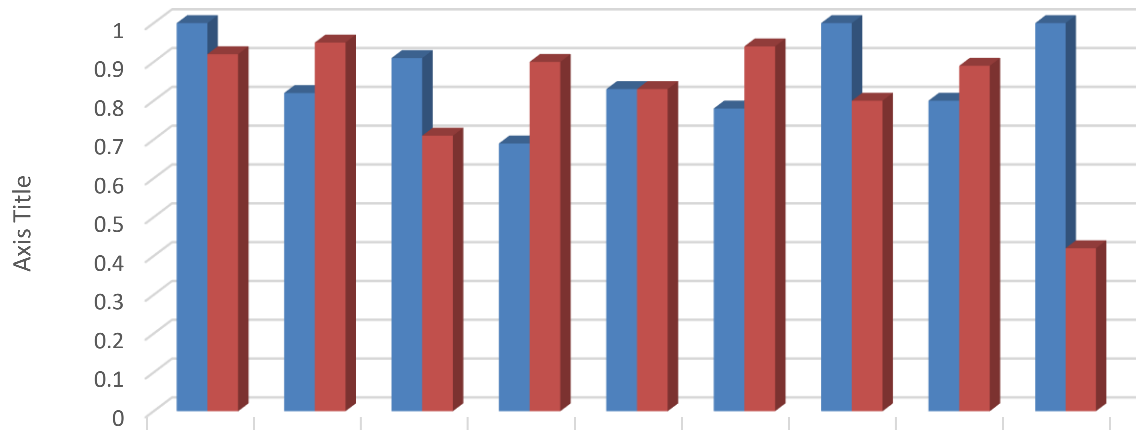


Figure 4-0-7: Object Co-occurrence Accuracy Comparison

4.3.3 F1 Score = 83%



	Right	Top Right	Top	Top Left	Left	Bottom Left	Bottom	Bottom Right	Center
Precision	1	0.82	0.91	0.69	0.83	0.78	1	0.8	1
Recall	0.92	0.95	0.71	0.9	0.83	0.94	0.8	0.89	0.42

Figure 4-8: NiNA Object Co-occurrence F1 Score

4.4 Limitations

- The Co-occurrence of Large objects with the tilted position in co-relation of small objects.
- When an object is onto another object.
- The object behind another Object

CHAPTER 5: CONCLUSION & FUTURE WORK

5.1 CONCLUSION

In this research, a novel approach for visual-spatial relations has been presented for object-object relation in a visual feed to identify the positioning of the different objects based on the placement of each object with respect to another object. Perceptual Associative Memory (PAM) has been employed to get more details from the visual feed by address, the correlation between each object in an environment to improve the visual co-occurrence of objects in a perceptual visual feed. Although the visual perceptual associative memory process; object detection and feature extraction; along with the perceived positions of objects in an environment needs to be elaborated in the context of other objects presented in the same environment by defining their relative direction and distance with other objects. Keeping the accuracy of object detection high, dividing the scene into 8 parts (quarters) and based on the positioning of objects in x,y plane developing correlation between each object.

It's relatively a basic technique. However, the importance of this unfolds with the combination of semantics from the audio feed that manifests the visual perception of an environment in a cognitive model. Through its working memory and with the help of other memory units present in NiHA Cognitive architecture agent which is essential for environment understanding. The extended desire of these efforts is to achieve higher-level cognitive competencies for social human-robot interaction. And through controlled environment observation and experimentation, this has been proven to be a valid possibility.

5.2 Future Work

- New Gesture Creation
- Stereo Vision for 3 Dimensional Scene Understanding
- Complete NiHA Framework Implementation

References

1] S. T. S. B. J. A. W. A. A. Wajahat Mahmood Qazi, "NiHA: A Conscious Agent," in *The Tenth International Conference on Advanced Cognitive Technologies and Applications, COGNITIVE 2018*, 2018.

2] Aldebaran, "NAO Documentation," Aldebaran Robotics, [Online]. Available: http://doc.aldebaran.com/2-1/home_nao.html.

3] I. Asimov, "Three laws that governed the behaviour of robots," in *Runaround*, 1942.

4] D. Coldewey, "World Economic Forum," TechCrunch, 26 June 2015. [Online]. Available: https://www.weforum.org/agenda/2015/06/what-is-collaborative-robotics/.

5] H. S. a. Y. E. Amit Gil, "A Cognitive Robot Collaborative Reinforcement Learning Algorithm," in *World Academy of Science, Engineering and Technology*, 2009.

6] J. L. L. J. J. B. Bidan Huang, "Humanoid Robots and Cognitive Systems Research:An Epistemological Case Study Based on the iCub".

7] A. V. Samsonovich, "On a roadmap for the BICA Challenge," *Biologically Inspired Cognitive Architectures,* vol. 1, pp. 100-107, 29 May 2012.

8] N. Mavridis, "A review of verbal and non-verbal human–robot interactive communication," *Robotics and Autonomous Systems, Elsevier,* no. 63, pp. 22-35, 2015.

9] G. Buttazzo, "Artificial Consciousness: Hazardous Questions(and Answers)," *Elsevier,* vol. 44, pp. 139-146, 2008.

10] I. Aleksander and B. Dunmall, "Axioms and Tests for the Presence of Minimal Consciousness in Agents," *Journal of Consciousness Studies,* vol. 10, 2003.

11] D. Gamez, "The Development and Analysis of Conscious Machines," *Department of Computing and Electronic Systems, University of Essex,* 2008b.

12] S. F. a. M. Ferkin, "An Ontology for Comparative Cognition: A Functional Approach," *Comparative cognition & Behavior Reviews,* vol. 1, pp. 36-52, 2006.

R. M. R. S. T. A. Catherina Burghart∗†, "A Cognitive Architecture for a Humanoid

13] Robot:," *IEEE-RAS International Conference on Humanoid Robots,* vol. 5, pp. 357-382, 2005.

14] I. a. T. J. K. Kotseruba, "40 years of cognitive architectures: core cognitive abilities and practical applications," *Artificial Intelligence Review, Springer*, pp. 1-78, 2018.

15] W. Qazi, "Modeling cognitive cybernetics from unified theory of mind using quantum neuro-computing for machine consciousness," *National College of Business Administration and Economics, School of Computer Science, Lahore, Pakistan,* 2011.

16] B. Baars, A Cognitive Theory of Consciousness., Cambridge University Press, 1998.

17] B. J. Baars and S. Franklin, "An Archiectural Model of Concious and Unconcious Brain Functions: Global Workspce Theory and IDA," *Neural Networks,* vol. 20, pp. 955-961, 2007.

18] C. L. A. S. ALEXEI V. SAMSONOVICH, "THE B-I-C-A OF BIOLOGICALLY INSPIRED COGNITIVE ARCHITECTURES," *International Journal of Machine Consciousness,* vol. 2, no. 2, pp. 171-192, 2010.

19] S. Franklin, "A Conscious Artifact?," *Journal of Consciousness Studies,* vol. 10, no. 20, pp. 47-66, 2003.

20] R. Sun, "The CLARION Cognitive Architecture: Toward a Comprehensive Theory of the Mind," in *The Oxford Handbook of Cognitive Science*, London;, Oxford University Press, 2017, pp. 1-29.

21] J. Anderson, "Human symbol manipulation within an integrated cognitive architecture," *Cognitive Science,* vol. 29, no. 3, 2005.

22] J. B. G. Anderson, "Human associative memory," in *Winston and Sons*, Washington, DC, 1973.

23] D. Vernon, G. Metta and G. Sandini, "The iCub cognitive architecture: Interactive development in a humanoid robot," in *IEEE*, London, UK, 2007.

24] A. V. Samsonovich, "Modeling Human Emotional Intelligence in Virtual Agents," in *Association for the Advancement of Artificial Intelligence,* 2013.

A. K. W. M. Q. Syed Tanweer Shah Bukhari, "Machine Imagination: A Step

25] Toward the Construction of Artistic World Through Storytelling," in *Recent Trends and Advances in Wireless and IoT-enabled Networks*, Springer International Publishing, 2019, pp. 197-205.

26] A. L. Paraense, K. Raizer, S. M. d. Paula, E. Rohmer and R. R. Gudwin, "The Cognitive Systems Toolkit and the CST Reference Cognitive Architecture," *Elsevier*, vol. 17, pp. 32-48, 2016.

27] S. Franklin, "A Foundational Architecture for Artificial General Intelligence,," *Adv. Artif. Gen. Intell. Concepts, Archit. Algorithms,* pp. 36-54, 2007.

28] J. S. a. S. F. R. McCall, "Sensory and Perceptual Scene Representation," in *J. Cogn. Syst*, 2010.

29] E. N. a. N. A. Taatgen, "Pre-attentive and attentive vision module," *Cogn. Syst*, pp. 211-216, 2013.

30] R. Sun, "Memory systems within a cognitive architecture,," *New Ideas Psychol,* vol. 30, no. 2, pp. 227-240, 2012.

31] D. Kieras, "Why EPIC was Wrong about Motor Feature Programming," in *Proceedings of the international conference on cognitive modeling*, 2009.

32] L. M. R. a. C. L. J. R. Anderson, "Working memory: activation limitations on retrieval," *Cogn. Psychol,* vol. 30, pp. 221-256, 1996.

33] B. Rohrer, "An implemented architecture for feature creation and general reinforcement learning,," in *Fourth International Conference on Artificial General Intelligence*, 2011.

34] D. Kieras, "Modeling Visual Search of Displays of Many Objects: The Role of Differential Acuity and Fixation Memory," in *Proc. 10th Int. Conf. Cogn. Model*, 2010.

35] A. V. S. a. K. A. D. Jong, "Designing a self-aware neuromorphic hybrid," in *Workshop on Modular Construction of Human-Like Intelligence*, 2005.

36] M. L. A. B. J. H. J. S.-V. a. R. P. J. Ruesch, "Multimodal saliency-based bottom-up attention a framework for the humanoid robot iCub," in *Proceedings of the IEEE International Conference on Robotics and Automation*, 2008.

J. M.-G. I. G.-V. L. R.-R. P. B. L. C. L. J. M. A. S. A. B. a. J. P. B. P. Bustos, 37] "Multimodal Interaction with Loki," in *Workshop of Physical Agents*, 2013.

38] S. B. a. C. L. R. Thomson, "Extending the Influence of Contextual Information in ACT-R using Buffer Decay," in *Annual Meeting of the Cognitive Science Society*, 2014.

39] R. Sun, "Autonomous generation of symbolic representations through subsymbolic activities," *Philos. Psychol,* vol. 26, no. 6, pp. 888-912, 2013.

40] D. E. M. S. M. a. T. S. D. E. Kieras, "Insights into working memory from the perspective of the EPIC architecture for modeling skilled perceptual-motor and cognitive human performance,," in *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control, A. Miyake and P. Shah, Eds. Cambridge University Press*, 1998.

41] L. J. M. J. P. B. A. R.-G. A. B. P. B. L. V. C. J. C. G. A. G.-O. R. F. a. F. F. R. Marfil, "Percepts symbols or Action symbols? Generalizing how all modules interact within a software architecture for cognitive robotics," in *Proceedings of the WAF2016*, 2016.

42] E. M. a. T. P. R. Sun, "A bottom-up model of skill learning," in *Proc. 20th Cogn. Sci. Soc. Conf*, 1998.

43] D. Vernon, "RobotCub: Development of a Cognitive Humanoid Robot," 2009.

44] D. P. a. C. Jones, "A model of object location memory," in *the 36th Annual Conference of the Cognitive Science Society*, 2014.

45] R. S. a. T. Peterson, "Autonomous learning of sequential tasks: Experiments and analyzes,," *IEEE Trans. Neural Networks,* vol. 9, no. 6, pp. 1217-1234, 1998.

46] S. F. K. C. D. M. a. R. T. T. Madl, "Towards real-world capable spatial memory in the LIDA cognitive architecture," *Biol. Inspired Cogn. Archit., ,* vol. 16, 2015.

47] J. G. T. a. A. M. Harrison, "Embodied Spatial Cognition," *Top. Cogn. Sci,* vol. 3, pp. 686-706, 2011.

48] M. B. I. P. K. I. a. K. T. H. Kato, "Virtual object manipulation on a table-top AR environment," in *Proceedings of the IEEE and ACM International Symposium on Augmented Reality*, 2000.

49] M. B. D. J. M. F. R. a. P. X. B. Rohrer, "Model-free learning and control in a mobile robot," in *Proceedings of the 5th International Conference on Natural Computation, ICNC*

2009, 2009.

J. P. B. P. B. L. V. C. F. F. R. F. F. J. G. A. I. J. L. R. M. C. P. C. R. A. R.-G. a. C.
50] S. A. Bandera, "CLARC: a Robotic Architecture for Comprehensive Geriatric Assessment," in *WAF2016,*, 2016.

C. v. H. a. L. F. D. Vernon, "The iCub Cognitive Architecture," *A Roadmap for*
51] *Cognitive Development in Humanoid Robots,* pp. 121-153, 2010.

R. S. a. P. Fleischer, "A Cognitive Social Simulation of Tribal Survival Strategies,"
52] *The Importance of Cognitive and Motivational Factors,* vol. 12, 2012.

a. S. W. J. E. L. S. D. Lathrop, "Exploring the Functional Advantages of Spatial and
53] Visual Cognition From an Architectural Perspective," *Cognitive Science Society,* vol. 3, pp. 796-818, 2011.

A. Baddeley, Human Memory - Theory and Practice (revised ed.), East Sussex, BN3
54] 2FA: Psychology Press, Taylor and Francis Group, 1997.

S. Franklin, "COGNITIVE ROBOTS: PERCEPTUAL ASSOCIATIVE MEMORY
55] AND LEARNING," in *IEEE International Workshop on Robots and Human Interactive Communication*, 2005.

A. G. O. Mandar Haldekar, "Identifying Spatial Relations in Images using
56] Convolutional Neural Networks,," in *arXiv*, 2017.

R. K. M. B. L. F.-F. C. Lu, "Visual Relationship Detection with Language Priors," in
57] *Stanford University*, 2015.