ADAPTIVE VISION ATTENTIVE ROBOT EYE FOR SERVICE ROBOTS IN DOMESTIC ENVIRONMENTS

Master of Science Dissertation

B.G.D. ACHINTHA MADHUSANKA

119107P

Department of Electrical Engineering

University of Moratuwa

Sri Lanka

May 2017

ADAPTIVE VISION ATTENTIVE ROBOT EYE FOR SERVICE ROBOTS IN DOMESTIC ENVIRONMENTS

A dissertation submitted to the
Department of Electrical Engineering, University of Moratuwa
in partial fulfilment of the requirements for the
degree of Master of Science

by

BATHTHALAMULLA GAMAGE DON ACHINTHA MADHUSANKA

Supervised by: Dr. Buddhika Jayasekara

Department of Electrical Engineering
University of Moratuwa
Sri Lanka

May 2017

DECLARATION

I declare that this is my own work and this dissertation does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant to University of Moratuwa the non-exclusive right to reproduce and distribute my dissertation, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature: Date:

The above candidate has carried out research for the Master's Dissertation under my supervision.

Signature of the supervisor: Date:

CONTENTS

ABSTRACT

Using eyes as an input modality for different control environments is a great area of interest for enhancing the bandwidth of human machine interaction and providing interaction functions when the use of hands is not possible. Interface design requirements in such implementations are quite different from conventional application areas. Both command-execution and feedback observation tasks may be performed by human eyes simultaneously. In order to control the motion of a mobile robot by operator gaze interaction, gaze contingent regions in the operator interface are used to execute robot movement commands, with different screen areas controlling specific directions. I have developed of adaptive vision attentive robot eye for a service robot.

In this project a methodical approach has been followed to the design and develop an interactive robotic eye for adapting robot attention to user command request about the distance of an object based on the visual attention of the robot. In a human robot interaction, the humans may use command request, which focus or search object a feedback whether the movement is “near”, “middle”, and “far”. The actual quantitative meaning of those terms depends on spatial arrangement of the domestic environment where the attention is focused on. Therefore, spatial information of the environment is analyzed to adapt robot’s perception about the distance of an object, which is in its vision field. The process includes the mechanical and electrical designs, in the design process close attention has been paid to the human bio-mechanics to realize a design that reaches anthropomorphism to a closer degree.

The proposed method is capable to mimics key visual functions of the human brain promises to robot eye maneuver quickly and safely through adaptive vision field through the domestic environments. The motion of changing adaptive vision field which used a focus or search objects in domestic environment more human-like manner using depth map analysis. Also the proposed robotic eye is designed in such a way that it can be used as a platform for facilitating further developments in integrating more interactive features to robotic eye.

ACKNOWLEDGEMENT

Though the work in this dissertation is my own, many people have been involved in its creation both directly and indirectly, and these are the people I would like to express my thanks to.

First of all, I want to express my gratitude to my supervisor Dr. Buddhika Jayasekara for believing in this project as much as I have, and for supporting the process with guidance and insights I could not have been without.

I would like to extend our thanks to the Industrial Automation program coordinator Dr. D.P Chandima. My deep sense of appreciation goes to the academic staff of the department of Electrical Engineering, University of Moratuwa, for the invaluable advices and guidance.

Lastly we thank all the individuals who assisted us in making this endeavor a success.

LIST OF FIGURES

LIST OF TABLES

ABBREVATIONS

ASIMO	Advanced Step in Innovative Mobility
GPS	Global Positioning System
ARM	Adaptive Resolution Method
DOF	Degree of Freedom
EOM	Extraocular Muscles
CAD	Computer Aided Design
CMOS	Complementary Metal Oxide Semiconductor
DC	Direct Current
PC	Personal Computer
IC	Integrated Circuit
PWM	Pulse Width Modulation
USB	Universal Serial Bus
OpenCV	Open source Computer Vision
CAMShift	Continuously Adaptive Mean Shift
HSV	Hue Saturation Value
ROI	Region of Interest

CHAPTER 1: INTRODUCTION

1.1. Overviews of robots

Robotics is the branch of mechanical engineering, electrical engineering and computer science that deals with the design, construction, operation, and application of robots, as well as computer systems for their control, sensory feedback, and information processing.

These technologies deal with automated machines that can take the place of humans in dangerous environments or manufacturing processes, or resemble humans in appearance, behavior, and or cognition. Many of today's robots are inspired by nature contributing to the field of bio-inspired robotics.

The concept of creating machines that can operate autonomously dates back to classical times, but research into the functionality and potential uses of robots did not grow substantially until the 20th century. Throughout history, it has been frequently assumed that robots will one day be able to mimic human behavior and manage tasks in a human-like fashion. Today, robotics is a rapidly growing field, as technological advances continue; researching, designing, and building new robots serve various practical purposes, whether domestically, commercially, or militarily [1]. Many robots are built to do jobs that are hazardous to people such as defusing bombs, finding survivors in unstable ruins, and exploring mines and shipwrecks. As more and more robots are designed for specific tasks this method of classification becomes more relevant. For example, many robots are designed for assembly work, which may not be readily adaptable for other applications. They are termed as assembly robots.

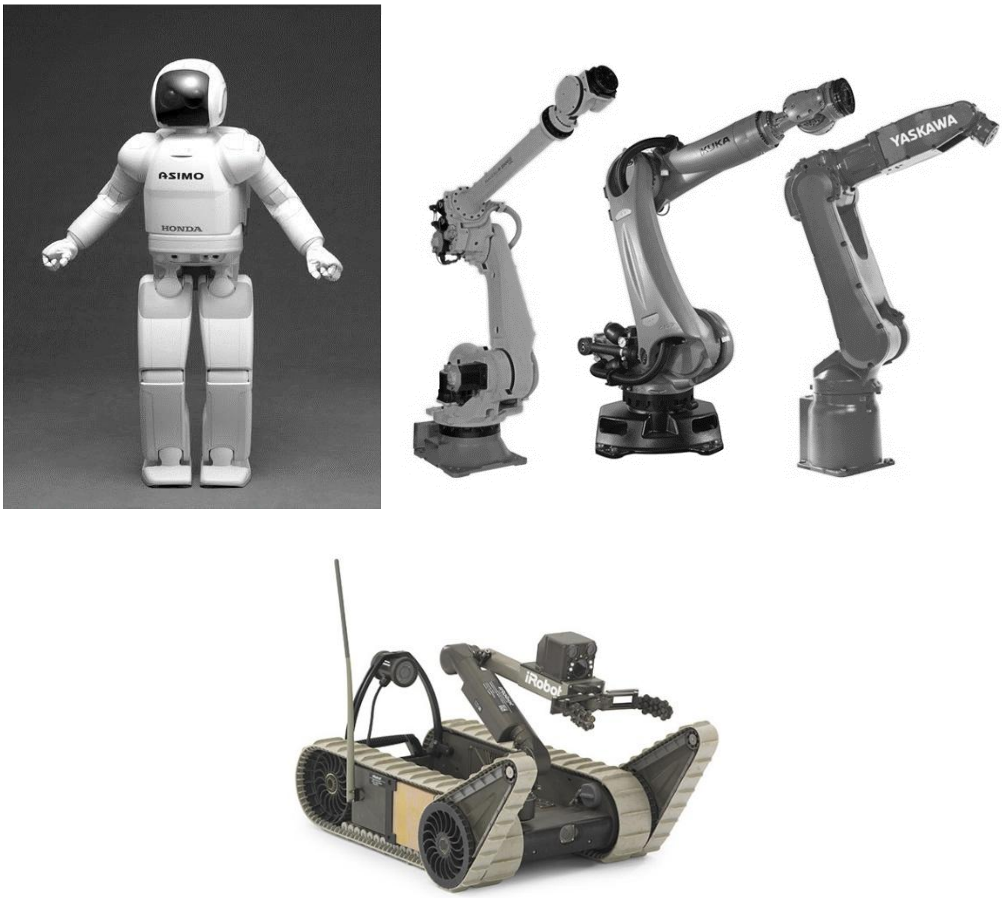


Figure 1.1: Robots currently developed in the world

Though a significant percentage of robots in commission today are either human controlled, or operate in a static environment, there is an increasing interest in robots that can operate autonomously in a dynamic environment [2]. These robots require some combination of navigation hardware and software in order to traverse their environment. In particular unforeseen events (e.g. people and other obstacles that are not stationary) can cause problems or collisions. Some highly advanced robots such as ASIMO has particularly good robot navigation hardware and software. Also, self-controlled cars are capable of sensing the environment well and subsequently making navigational decisions based on this information [3]. Most of these robots employ a GPS navigation device with waypoints, along with radar, sometimes combined with other sensory data such as lidar, video cameras, and inertial guidance systems for better navigation between waypoints.

1.2. Robotic applications in service robots

Service Robotics is the branch of robotics aiming at the development of robots able to assist humans in their environment. Service robots are going to be the major application field for technologies developed by the autonomous robot research community. Service robots must be designed to interact with people, so human-robot interaction technologies are crucial for service robotics. An important ability they will have is the capability to entertain and play with humans. In a word they can be "companions" to humans with specialized abilities. Possible applications for service robots will be to assist elderly people or to allow parents to stay close to their children through telepresence, or they can act as physical avatars for virtual meetings [4]. Moreover, they may be used for surveillance and security within houses or company buildings.

The purpose of this service robots are to facilitate the integration of different research activities in order to short-term the development of commercial service robots. More functions such as autonomous navigation, object and people recognition, human-robot interaction, object manipulation, ethical issues, and many others, currently treated as separate independent problems, have to be combined in order to produce effective solutions to achieve effective service robots.

Industrial and service robots differ significantly in terms of specifications. This can be seen when the required positioning accuracy is considered, or how the robots are integrated into the overall system. While traditional industrial robots perform their tasks in clearly structured environments with external safeguards, service robots usually work in unstructured environments and collaborate directly with humans [5].

While industrial robots are made safe by being deactivated when somebody comes close, service robots have to interact with people. As a result, they require more complex safety concepts in order to ensure safe operation, perhaps even going as far as proximity sensors and tactile skin. Industrial safety standards can be applied to service robotics wherever it makes sense to do so. However, at the same time they must not be overdone and end up running up exorbitant costs. Some smaller aspects, such as gripping technology and kinematics, can be applied to service robotics applications relatively easily. The manufacturer needs to consider the far more varied requirements of service robotics.

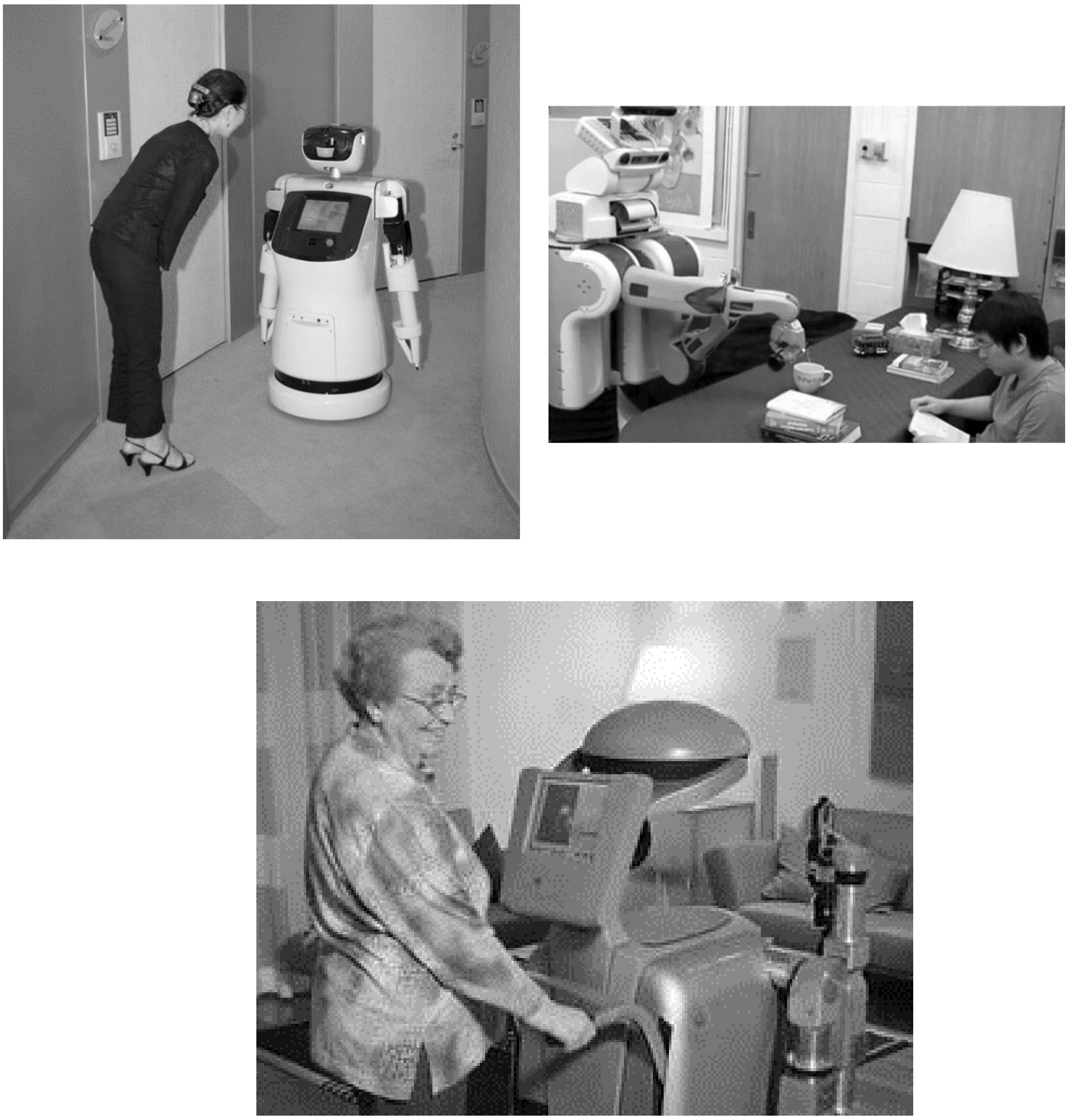


Figure 1.2: Service robots currently developed in the world

1.3. Robotic eyes

In many experiments in psychology, human-computer interaction, and other fields, researchers want to monitor precisely what subjects are looking at. Gaze can reveal not only what people are focusing their attention on but it also provides clues about their state of mind and intentions [6].

Mobile systems to monitor gaze include eye-tracking software and head-mounted cameras. But they're not perfect; sometimes they just can't follow a person's fast eye movements, and sometimes they provide ambiguous gaze information.

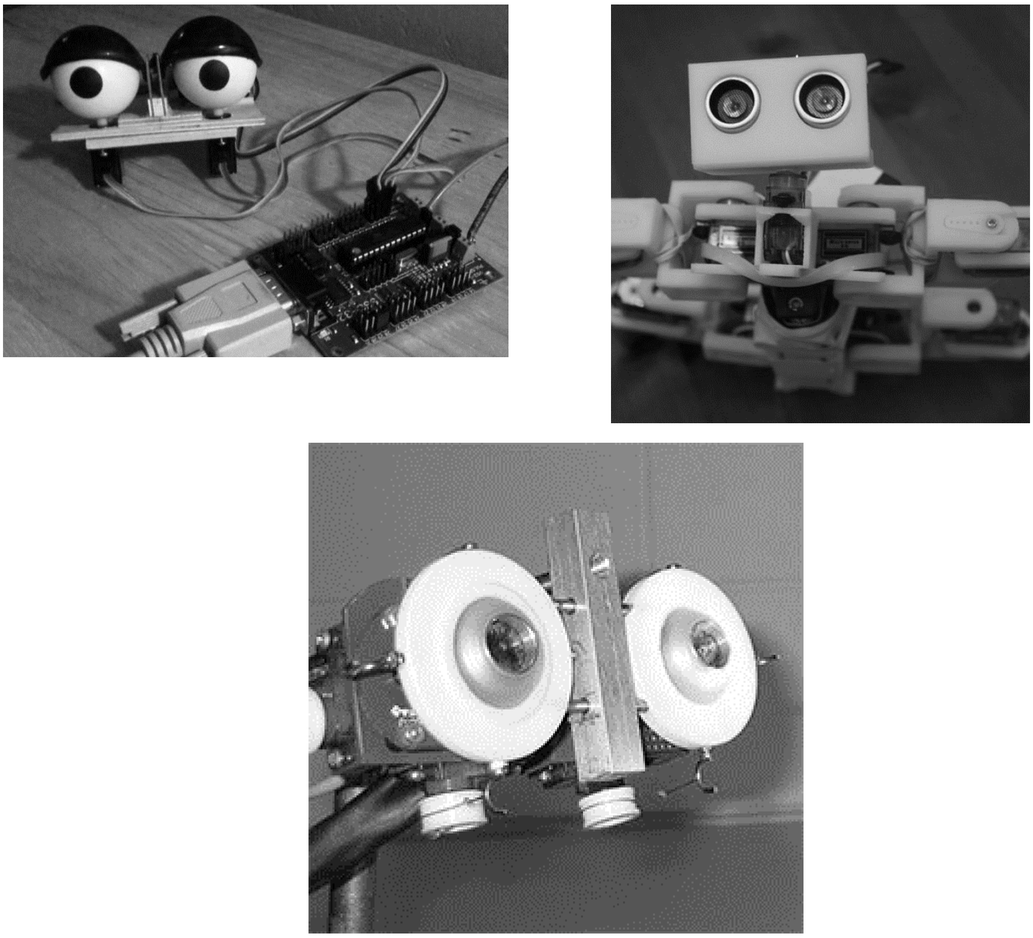


Figure 1.3: Applications of robot eye

1.4. Applications of robotic eyes

In biological vision systems, the attention mechanism is the responsible of selecting the relevant information from the sensed field of view so that the complete scene can be analyzed using a sequence of rapid eye saccades. This attention behavior has been imitated by artificial vision systems in order to optimize computational resources. Probably one of the most influential theoretical models of visual attention is the spotlight metaphor that has inspired many concrete computational models [7]. These approaches are related with the feature integration theory, a biologically plausible theory proposed to explain human visual search strategies. According to this model, these attention mechanisms are organized into two main stages.

First, in a pre-attentive task-independent stage, a number of parallel channels compute image features. The extracted features are integrated into a single saliency map which codes the saliency of each image region. The most salient regions are selected from this map. Second, in an attentive task dependent stage, the spotlight is moved to each salient region to analyze it in a sequential process. Analyzed regions are included in an inhibition map to avoid the spotlight moving to an already visited region. Thus, while the second stage must be redefined for different systems, the pre-attentive stage is general for any application [8]. Although these models have good performance in static environments, they cannot in principle handle dynamic environments due to their impossibility to take into account the motion and the occlusions of the objects in the scene. In order to solve this problem, propose an attention mechanism which incorporates depth and motion as features for the computation of saliency.

The previously described methods deploy attention at the level of space locations (space-based models of visual attention). The models of space-based attention scan the scene by shifting attention from one location to the next to limit the processing to a variable size of space in the visual field. Therefore, they have some intrinsic disadvantages. In a normal scene, objects may overlap or share some common properties. Then, attention may need to work in several discontinuous spatial regions at the same time [9]. Only if different visual features, which constitute the same object, come from the same region of space, an attention shift will not be required. On the contrary, other approaches deploy attention at the level of objects instead to a generic region of space. Object-based models of visual attention provide a more efficient visual search than space-based attention. Besides, it is less likely to select an empty location. In the last few years, these models of visual attention have received an increasing interest in computational neuroscience and in computer vision [10]. These models reflect the fact that the perception abilities must be optimized to interact with objects and not just with disembodied spatial locations. Thus, visual systems that follow this approach will segment complex scenes into objects which can be subsequently used for recognition and action. However, recent physiological research shows that, in natural vision, the pre-attentive process divides a visual input into raw or primitive objects instead of well-defined objects. Some authors use the notion of proto-objects to refer to these primitive objects that are defined as units of visual information that can be bound into a coherent and stable object.

Nevertheless, space-based and object-based approaches are not mutually exclusive, and several researchers have proposed attentional models that integrate both approaches. Thus, combine object-based and feature-based theories in the model of visual attention. In its current form, this model is able to replicate human viewing behavior. However, it needs input images to be manually segmented. That is, it uses information that is not available in a pre-attentive stage, before objects are recognized [11]. Another approach following the space- and object-based integration is the one proposed by which employs a Bayesian model to describe the visual attention mechanism. Finally, some models build their saliency maps taking into account not only visual information but also using another senses.

1.5. Existing Robotic eyes and their features

A dynamic visual memory to store the information gathered from a moving camera on board a robot has been proposed in [12] implementing an attention system to choose where to look with this mobile camera, and a visual localization algorithm that incorporates this visual memory. The visual memory is a collection of relevant task-oriented objects and 3D segments, and its scope is wider than the current camera field of view. The attention module takes into account the need to re-observe objects in the visual memory and the need to explore new areas. In [13] has been described research and development of a visual servoing system for autonomous satellite capture using an on-board manipulator with binocular hand-eye cameras. Taking account of the special aspects of space environment such as lighting and available computing power, to realize a safe and reliable real-time visual servoing operation, cooperative visual marker search processing, time delay processing, and capture strategy have been carefully designed. Novel adaptive and individualized robot-mediated technology for children with autism spectrum disorder has been proposed in [14] to seek bridge this gap by developing composed of a humanoid robot with its vision being augmented by several wall-mounted cameras for real-time head tracking using a distributed architecture. Based on the cues from the child's head movement, the robot intelligently adapts itself in an individualized manner to promote joint attention.

The development of a reliable vision-based eye-gaze detection system for an analysis of multiple users' intention and communication has been proposed in [15]. Headgear design for the gaze detection, a recording system, an off-line computation, an algorithm of the gaze extraction, and its calibration method, are mentioned. Applying the proposed system to a test case, sufficient precision of the gaze detection was confirmed.

In [16] has been presented a novel adaptive controller for image-based visual servoing of robots with an eye-in-hand camera using angle and distance features. The key idea lies in the development of the depth-independent interaction matrix and the proposal of an adaptive algorithm for estimating the unknown geometric parameters of the features in the 3-D space. The Lyapunov theory is used to prove the asymptotic convergence of the image error to zero based on the nonlinear robot dynamics.

An eye-in-hand device for hand gesture-based human-robot interaction has been proposed in [17]. As an eye-in-hand configured vision system, the device is simply worn on human wrist and provides hand images in close proximity due to physical constraints. Taking advantage of an eye-in-hand configuration, the proposed device guarantees that a target hand is projected always onto images captured from a camera. On the basis of the device, an adaptive histogram-based hand region segmentation algorithm is also proposed especially to enhance the performance against background and illumination changes.

Based on a study of the engagement process between humans, [18] has developed and implemented an initial computational model for recognizing engagement between a human and a humanoid robot. That model contains recognizers for four types of connection events involving gesture and speech: directed gaze, mutual facial gaze, conversational adjacency pairs and backchannels. To facilitate integrating and experimenting with our model in a broad range of robot architectures, they have packaged it as a node in the open source robot operating system framework. They have conducted a preliminary validation of computational model and implementation in a simple human-robot pointing game.

In the humanoid soccer league competition, the vision system is used to collect various environment information as the terminal data to finish the functions of object recognition, coordinate establishment, robot localization, robot tactic, barrier avoiding, has been proposed in [19]. Therefore, a real-time object recognition and high accurate self-localization system of the soccer robot becomes the key technology to improve the performance. In this work proposed an efficient object recognition and self-localization system for the humanoid soccer league rules of competition. This research proposed two methods, firstly, the object recognition part; the real-time vision based method is based on the Adaptive Resolution Method (ARM). It can select the most proper resolution for different situations in the competition. ARM can reduce the noises interference and make the object recognition system more robust as well. Secondly, the self-localization part, they proposed a new approach, adaptive vision-based self-localization system, which uses the trigonometric function to find the coarse location of the robot and further adopts the measuring artificial neural network technique to adjust the humanoid robot position adaptively.

An autonomous grasping system using visual servoing and mobile robots has been proposed in [20]. While this kind of system has many potential significant applications, there have been several key challenges, for example, localization accuracy, visibility and velocity constraints, obstacle avoidance, and so on, to prevent the implementation of such a system. The main contribution of this research is to develop an adaptive nonlinear model predictive controller to meet all these challenges in one single controller. In particular, the model of the vision-based mobile grasping system is first derived. Then, based on the model, a nonlinear predictive control strategy with vision feedback is proposed to deal with the issues of optimal control and constraints simultaneously. Different from other work in this field, in order to improve the performance, an adaptive mechanism is proposed in the paper to update the model online so that it can track the nonlinear time-varying plant in a real time manner.

The design of the social robot Probo has been proposed in [21]. The robot will be used in hospitals, as a tele-interface for entertainment, communication and medical assistance. Therefore, it requires the ability to express emotions. In order to do so, an emotional interface is developed to fully configure the display of emotions. These emotions represented as a vector in an emotion space are mapped to the degrees of freedom used in the robot. Besides emotions, the interface includes a control for the point of attention and a module to create and store animations. A 3D virtual model is created, acting as a virtual replica of the robot, providing realistic visual feedback to evaluate the design choices for the facial expressions.

Moving target tracking and measurement with a binocular vision system has been used in [22] to achieve the visual servoing, mimicking the movements of human eyes. First, a robust scheme, which combines adaptive background subtraction and Camshift algorithms, is proposed to detect and track a moving object. Then, it discusses the problems of camera parameter calibration and position measurement of moving objects. A new approach is introduced to realize the extrinsic parameter calibration of a pan-tilt camera, providing successful tracking and accurate measurement of a moving target, even as it is outside the view of camera.

1.5.1. Sensors and actuators

In general robot eyes use vision sensors to receive feedbacks from the external environment. Also use motors to find external feedbacks from the environment.

1.5.2. Interactive features of robotic eyes

There is an increasing interest in advanced human-robot interfaces due to a growing need for service robots, designed to perform a variety of tasks in human inhabited environments. Robot eye movements are particularly important for human-humanoid interaction, because they constitute a highly attended and communicative part of the human body. A key issue for the acceptance of such systems by generic users lies on the ability to generate human-like motions and postures.

Due to different bio-mechanical dynamic characteristics, the eyes are quicker in achieving the goals set by the neurological commands. In the control of saccades, eyes rapidly achieve the target point. Then, the head moves slower toward the target, while the eyes counter rotate to compensate head motion.

In humans several stimuli and physiological subsystems contribute to coordinate eyes and neck motions: visual information, acoustic cues, afferent and efferent copy signals, the vestibular system, motion prediction, etc. The overall system is of great complexity and to provide a neuro-biological compatible and complete model. More interactive features intend to model the behavior of biological gaze control systems as observed by an external agent. For the design of human-humanoid interfaces, it is the external robot behavior that influences human perception, rather than the driving internal models.

Developing computational perception systems that emulate this ability becomes a critical step in designing robots that are able to cooperate with people as capable partners, that are able to learn from natural human instruction, and that are intuitive and engaging for people to interact with, but that are also able to navigate in initially unknown environments or to grasp an object. In order to accomplish these tasks it is typically assumed that the perception system of a robot should imitate the ability of natural vision systems to select the most salient information from the broad visual input.

Specifically, the perception system used to extract, from the visual input data, the information that the robot will need to accomplish both navigation and human-robot interaction behaviors. Besides it is interesting, to achieve an intuitive interaction with people that the robot is able to perceive the real world in a similar way that people do.

1.5.3. Object tracking

Object tracking is considered as important subject within the area of computer vision. Availability of high definition videos, fast processing computers and exponentially increasing demand for highly reliable automated video analysis has created a new and great deal for modifying object tracking algorithms. Some of the important applications of object tracking are:

- **Automated video surveillance:** In order to monitor the happenings in a particular area, to detect and recognize moving objects and to detect unlikely events and to report suspicious, criminal activities applications are developed in computer vision system.

- **Robot vision:** for navigation of robot, different obstacles in the path are identified by the steering system in order to avoid collision. In case the obstacles are itself in motion then we need a real time object tracking system.

- **Monitoring of traffic:** Highway traffic can be continuously monitored using moving cameras. The vehicle breaking any type of law or involved in any illegal activities can be detected and tracked using an object tracking system.

- **Animation purpose:** Animation can be supported by using object tracking Algorithm.

- **Human computer interaction:** it can be used for automatic attendance system in many areas and to record the in and out time of the object.

Object tracking can be performed stage wise stage and the stages are:

i. **Object detection**

The computer technology associated with computer vision and image processing which deals with the task of detecting and identifying instances of semantic objects belonging to a certain class in video and images. Face detection, pedestrian, detection, vehicle tracking etc. are the well-researched domains of object detection. Object detection is widely used in computer vision including areas like image retrieval and surveillance of video. In a video surveillance system moving detection algorithms can be broadly classified into two categories depending on their relative movement between scene of surveillance and camera.

If the camera is ideal and fixed then the surveillance scene will remain unchanged leading to method called static background detection method. If camera is not static then camera and scene of surveillance will have some relative movement and the method is called as moving background detection method. In static background detection method position and size of pixels in the background will remain same in different frames of an image sequence .Thus it can use method of method of difference of the pixels in term of intensity or color value to determine moving region and extract moving object at the same previous position in the different frames [26]. This method of extraction is called as inter-frame difference method.

ii. Object tracking

Tracking is the problem of estimating the trajectory of an object as it moves around a scene. Ultimate aim of object tracking is to associate objects targeted in upcoming video frames. This can be very difficult in case relative motion between moving objects and frame rate is high. Change of orientation of tracked object with passing of time can increase the complexity of the process. Thus to deal with the above problem we are employ a model for motion of the object to show how the image of the target might get change for every possible motion of the object. Basically tracker assigns consistent and unique labels to the objects tracked in different frames of a video. Additionally, depending on the domain of tracking, a tracker can also provide centroid information of the object, orientation, shape, or are of an object [27].

Tracking objects can become complex due to following reasons:

- Projection of the 3D world on a 2D image will cause loss of information.
- Presence of noises in the image.
- Irregular object motion.
- Non rigid nature of objects.
- Occurrence of occlusion (partial or full).
- Objects shape is complex.
- Change in the illumination or intensity of the scene.

iii. Dynamic template matching

The generated templates from detection module are passed on to the tracking module, which initiate tracking the moving object with a given input reference template. The tracking module makes use of template-matching to search for the input template in the future frames grabbed by the video. A new template is generated in case the object is lost during tracking due to change in its appearance and used further. Generations of such templates are dynamic which helps to track the object in robust manner. The theory behind template matching is given below [27].

- The matching process moves the template image to all possible positions in a larger source image and computes a numerical index that indicates how well the template matches the image in that position.
- Matching is done on a pixel-by-pixel basis of both input template and masked template.

In this implementation two web cameras have been used as a robot eye to grab the video frames from the ideal environment. Once the object has been detected it is tracked by employing an efficient template matching algorithm. The templates used for the matching purposes are generated already and saved before tracking. This ensures that any change in the pose of the object does not hinder the tracking procedure. To automate the tracking process the robot eye coupled with DC motors, which is synchronized with a tracking algorithm.

1.6. Novel trends in Robotic eye development

Adaptive vision attentive robot eye motivates the novel trends in robotic eye development research, which is to develop robot eye with pan-tilt movement that can be used more human-like interaction. Those trends are crucial for service robots in domestic environment to robustly and reliably interact with the real world and development of robot eye with pan tilt movements. However, vision attentive for human-like interaction was not completely resolve as so far. In order further research progress, need to be implemented vision attentive system with human-like interaction more tracking performances.

1.7. Development of an adaptive vision attentive robotic eye for service robot

Design and development of the mechanical and electrical structure and developing interactive features in the robot eye are the major steps of the project. Design phase of the project is comprised of designing both hardware and electrical systems thereafter software implementation and adding interactive features are completed.

1.7.1. Objectives

Main goal with this project is to design and development of adaptive vision attentive robot eye for service robots in domestic environments. This project will perform several objectives as followed.

- Design and development of robot eye with pan tilt movements for a service robot.
- Development of adaptive vision attentive features for keeping attention in human-robot interactions.
  - Development of vision attentive system focusing to the given location in the environment.
  - Development of vision attentive field according to the human-like interaction.

    - Development more human-like pan-tilt movement for given vision attentive scenario.
- Implementation of adaptive vision attentive robot eye in different domestic scenarios.

1.7.2. Motivation

The design and construction of human-like robot eye that can perceive and interact with the environment depends significantly on their perception capabilities. In this dissertation I present the human-like robot eye, which has been designed to be used both as part of my robot eye and as a stand-alone robot eye for studying various visual perception tasks in the context of object tracking, object searching and human-like robot eye interaction. The eyes have a common tilt and can pan independently. Each eye is equipped with two digital color cameras, with peripheral vision for fovea vision to allow simple vision-motor behaviors. Among these are tracking and searching motions towards relevant regions, as well as more complex visual tasks such as hand-eye coordination. We present the mechatronic design concept, the motor control system, the sensor system and the computational system. Considering the information acquired I was motivated to develop a robotic eye that would help in promoting human friendly interactive features in service robots.

1.7.3. Salient features

Designed robotic eye is capable of performing basic eye movements and its motors are well positioned to get these movements with close to a human-like. Two cameras are used for achieving vision and main processing tasks are carried out using a PC. Motor controlling is done through a microcontroller.

- 3DOF eye movements (common tilt and separate pan movements)
- Binocular vision
- Interactive eye movements more human-like

Developed interactive features are object tracking, object searching, tracking pre-defined objects and identifying the depth. And also it has been developed to give responses with the coordination of eye movements. Vision feedback and key board inputs are taken as inputs to the system.

1.8. Overview of the thesis

This thesis contains six chapters which describes the process followed in design and development of an adaptive vision attentive robot eye. First chapter offers brief overview about the robots and provides more detailed description about the existing robotic eyes and the current trends of robotic eye development together with a brief introduction of the proposed features of the robotic eye that was proposed.

The second chapter provides details about the design of the mechanical design of robot eye which includes a description about the biomechanics of the human eye system. This chapter also offers information about the hardware selection for the mechanical system. Design details about the electrical and control systems of the robotic eye are given under chapter three. The forth chapter of the report has been dedicated for presenting the interactive robot eye which including adaptive vision attentive features carried out with regard to the robotic eye.

The fifth chapter provides information about the results in adaptive vision features of object tracking and object recognition. Chapter six is dedicated problems encounter and further developments to be suggested.

CHAPTER 2: MECHANICAL DESIGN OF ROBOT EYE

2.1. Bio-Mechanics of human eyes

Vision is one of our most important senses. Rapid and accurate eye movements are crucial for coordinated direction of gaze. Many kinds of disorders that impair human vision are associated with pathological eye movements. Understanding how the eye moves is driven by the demand for better treatment for these diseases. For instance, strabismus, or binocular misalignment, is a common visual defect in which two eyes do not look at the same point at the same time when focusing.

Each eye has six muscles, called the Extraocular Muscles (EOMs). Strabismus is typically caused by uncoordinated extraocular muscles. Strabismus can be treated surgically; this involves manipulating the unbalanced EOMs. Because of the difficulty of predicting the mechanical effects of operations, current diagnosis and treatment are mostly based on simple heuristics, intuition, and experience. As a result, the success rates of strabismus surgeries are not satisfying. Many operational treatments are found ineffective in improving vision, and re-operations are generally required. To advance the knowledge of binocular misalignment and human eye movement in general, much effort has been devoted to reexamine the anatomy and biomechanics of the orbital plant.

Computational modeling provides a powerful tool to analyze biological systems, complementary to experimental studies. It aims to build realistic models for individual components of the system to describe the underlying mechanism. Such a model normally employs physical principles and computational methods. Validation is performed through simulation and comparison to the empirical findings. Computational models are not isolated from experimental observations. Incorporating anatomical and physiological properties is critical for their realism and accuracy, and for their utility in scientific studies and clinical applications. In turn, modeling efforts and predictions can provide insights into the structures and aid in the design of better experiments.

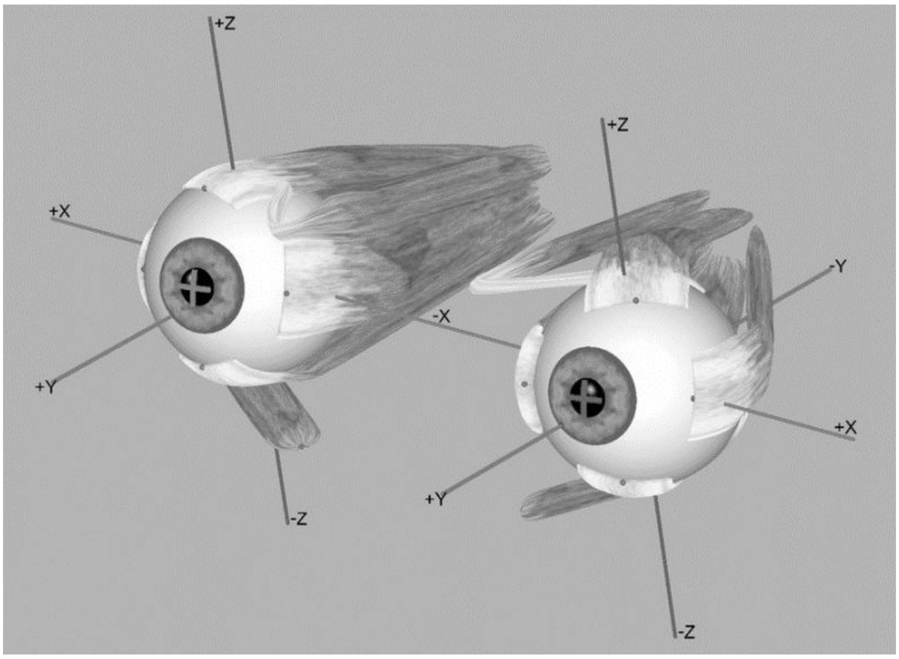


Figure 2.1: Bio-mechanical models of the eye muscles

Studying human eye movement has other significant implications. The oculomotor plant is an unique mechanical system. It is organized in a seemingly simple way – three pairs of extraocular muscles work as three agonist and-antagonist pairs and control 3D rotation with three degrees of freedom.

2.2. Construction of robot eye mechanism

Each human eye has six muscles. As it is a globe inside a socket three motions can be considered, abduction/adduction, elevation/depression and rotation. The muscles have combined actions to achieve these motions, as described in Table 2.1. Each eye is completely independent of the other.

Table 2.1: Eye muscles action

Muscle	Action
Superior rectus	Elevates, adducts, and rotates eyeball medially
Inferior rectus	Depresses, adducts, and rotates eyeball laterally
Lateral rectus	Abducts eyeball
Medial rectus	Adducts eyeball
Superior oblique	Abducts, depresses, and medially rotates eyeball
Inferior oblique	Abducts, elevates, and laterally rotates eyeball

The human oculomotor system combines several basic movements: saccades, smooth pursuit, vergence, vestibuloocular reflex, optokinetic reflex, microsaccades and accommodation. Saccadic and smooth pursuit eye movements occur when the eyes pursue an object. During smooth detection, the eye tries to match the (angular) speed of the tracked target, usually at relatively low speeds. Saccadic eye movements are high speed jumping movements, in the range of a few hundreds degrees per second. The saccadic eye movement occurs when the eye ball movement is not able to track an object or when the human searches outside of the view.

The range of the motion could not be found in research, so it had to be obtained through measurements conducted by the researcher. Figure 2.2 below depicts the range of eye positions corresponds to the mapping of these moments onto the eyeball model to achieve measurements of the angular positions. Also downward angle was easily achieved due to the tendency for the eyelids to close over the eye when looking down.

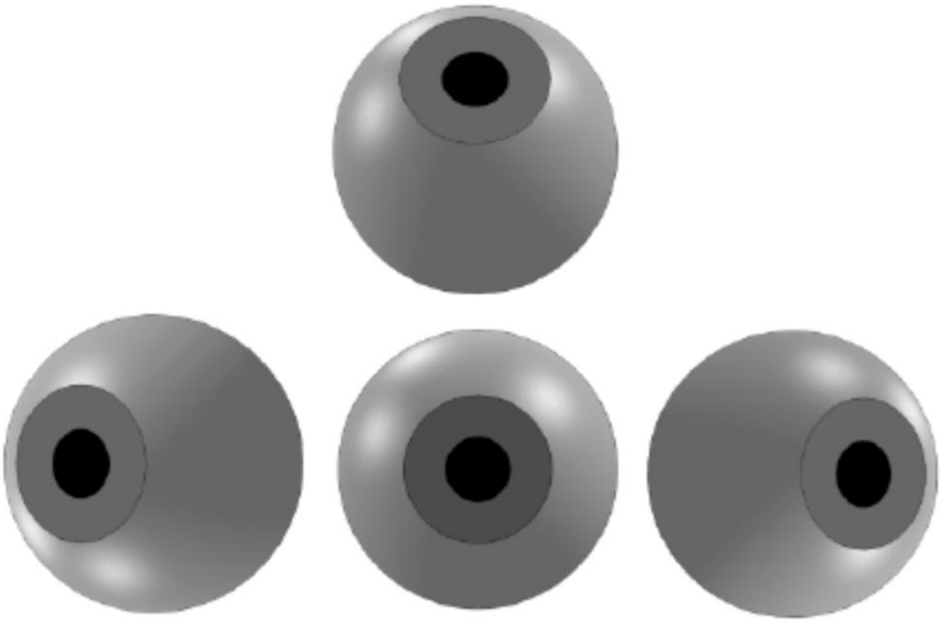


Figure 2.2: Corresponding eye orientations in CAD

Table 2.2 presents the range of movements for each joint, and the eye movement for human adults during saccadic movements. The saccade speed increases with the motion amplitude. Hence, the speed during small amplitude saccades resemble those of smooth detection. These data also show how the effort is divided amongst eye degrees of freedom when some redundancy exists (e.g. eye pan movements). This information has been used for the design of the robot eye specifications.

Table 2.2: Selected specifications for robot eye movement

<table>
<tr><th></th><th>Movement</th><th colspan="2">Constraining angle</th></tr>
<tr><td rowspan="4">Eye</td><td rowspan="2">Pan</td><td>Left</td><td>45°</td></tr>
<tr><td>Right</td><td>45°</td></tr>
<tr><td rowspan="2">Tilt</td><td>Up</td><td>40°</td></tr>
<tr><td>Down</td><td>40°</td></tr>
</table>

The eyes mechanism has three degrees of freedom. Both eyes have individual pan movements and a common tilting movement. Movements are achieved using three servo drives. The two cameras which are used as the eyes of the robotic, have been mounted on top of two servo motors (rotating motors) in such a way that their common tilting axis goes through the center of each camera lens. Panning axis of the cameras coincide with the rotational axis of the respective servo drives.

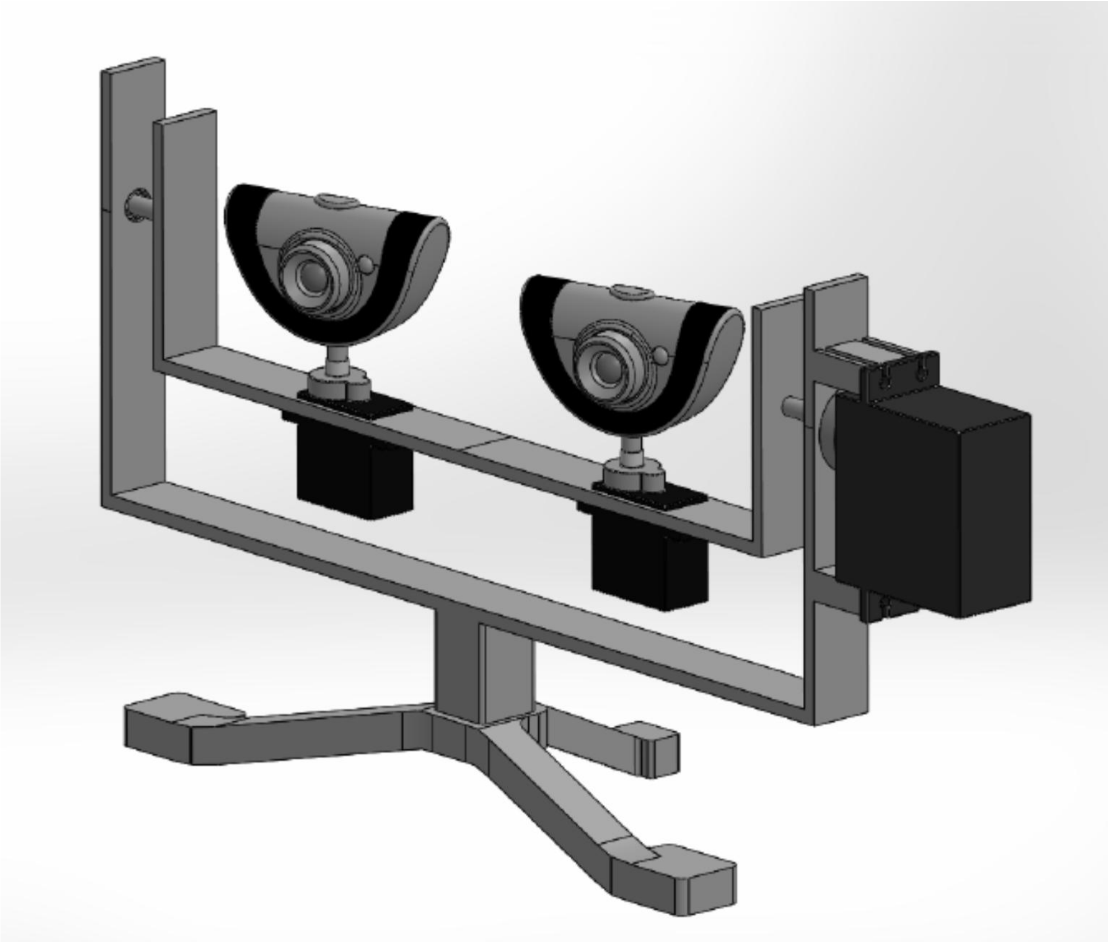


Figure 2.3: Configuration of eye mechanism

The common tilting movement has been achieved using another servo motor (tilting motor). Actuator lengths have been selected to fix as above mentioned considering the size of cameras, servo motors and all the other parts.

2.3. Hardware selection for mechanical system

During the design process, the preliminary design for the robot eye based upon both for the average human range of motion for each degree of freedom. Additionally, servo motors that were equal to or larger than the size of the motors that were to be designed in order to insure that there was space available within the outer shell of the robot eye such that tilt motor could fit inside. The range of speeds achieved in each degree of freedom was also used according to specification the servo motors. Additionally, the design was fine tuned to fit the pan-tilt motors that were purchased and to eliminate collisions of components while in motion.

2.3.1. Motor selection

The measured speeds specified in section 2.2 was used to determine the pan-tilt servo motors used for each degree of freedom. Using the kinematic relationship of the input link speeds to the output link speeds in each linkage, the necessary speeds of each motor can be obtained. Table 2.3 presents the output link speeds provided with the motor speeds needed to create those output link speeds, as well as the motor speed specification (s/60°) needed to pick each servo motor.

Since the pan-tilt motors are directly attached to the outputs, the angular speeds are the slight different. The torques required for each motor are calculated based on the design with each non-servo motor component. This is the material that was decided on for the final assembly since it is light, quick to manufacture, and relatively inexpensive. Table 2.3 below depicts both pan-tilt motor specification. It is important to note that the torques have been sufficient to account for added torques that will be caused by friction.

Table 2.3: Pan-tilt motor specification

<table>
<tr><th>Motor type</th><th>Figure</th><th>Specification</th></tr>
<tr><td rowspan="6">Pan motor (SG90 9g Micro Servo)</td><td rowspan="6">
</td><td>Weight: $9g$</td></tr>
<tr><td>Dimension:
$22.2\ x\ 11.8\ x\ 31\ mm$</td></tr>
<tr><td>Stall torque: $1.8\ kgf \cdot cm$</td></tr>
<tr><td>Operating speed: $0.1\ s/ 60\ degree$</td></tr>
<tr><td>Operating voltage: $4.8V - 5V$</td></tr>
<tr><td>Dead band width: $10\ \mu s$</td></tr>
<tr><td rowspan="7">Tilt motor (MG995 Metal Gear Servo)</td><td rowspan="7">
</td><td>Weight: $55g$</td></tr>
<tr><td>Dimension:
$40.7\ x\ 19.7\ x\ 42.9\ mm$</td></tr>
<tr><td>Stall torque: $8.5\ kgf \cdot cm\ (4.8\ V\ ), 10\ kgf \cdot cm\ (6\ V)$</td></tr>
<tr><td>Operating speed: $0.2\ s/ 60^{\circ}\ (4.8\ V), 0.16\ s/ 60^{\circ}\ (6\ V)$</td></tr>
<tr><td>Operating voltage:
$4.8\ V\ and\ 7.2\ V$</td></tr>
<tr><td>Dead band width: $5\mu s$</td></tr>
<tr><td>Stable and shock proof double ball bearing design</td></tr>
</table>

All specifications in the torque and speed tables are specified for servo motor operations at 5 volts. With this information in mind and based on the torques and speeds required for each degree of freedom, the above servo motors have been decided on. The pan motor was decided to be controlled with SG90 9g Micro Servos, since they have a maximum torque of 9 g-cm and a maximum speed of 0.10 s/60°. This motor is sufficient for camera movement and is desirable because it has a low profile and both servos could fit within the pan mechanism without any collisions with the links or the tilt of the aluminum profile.

The MG995 Metal Gear Servo was decided on for the tilt motor since it has a maximum torque of 10 kg-cm and a maximum speed of 0.16 s/60°. This motor has sufficient torque to rotate the entire robot eye assembly. Both of these values met the requirements for each of the three degrees of freedom, plus there were few other servo motor options that supplied less torque while still meeting the requirements.

2.3.2. Camera selection

As an open physical platform for embodied research, that can be used by the research community from different types of science fields, the robot eye system cannot be very complex. So, in order to easy assembly and maintenance procedures, the mechanical system architecture is also completely modular, in such a way that easy to remove and replace a certain module, without having to disassemble the entire structure.

Figure 2.4: Vision camera

To allow the robot to interact with more human-like and to have all desired behavior to web cameras were applied. These cameras are very easy to integrate because the vision sensor is mounted as robot eyes, connected to the electronics with a flexible cable. In this way, while the electronics are fixed to a non-moving part of the eye-system. Below depicts vision camera specification.

Image sensor:	CMOS
High definition:	16M Pixel
Dynamic resolution:	640 × 480
Frame rate:	30fps
Focus:	8mm - infinity
Weight:	99g

2.3.3. Selection of material for structural links

3mm thickness extruded aluminum is selected due to high strength to weight ratio. Ability to bear the required level of stresses has been verified using the SolidWorks model. Another major reason for selecting aluminum for structural links is the availability and the reasonable cost.

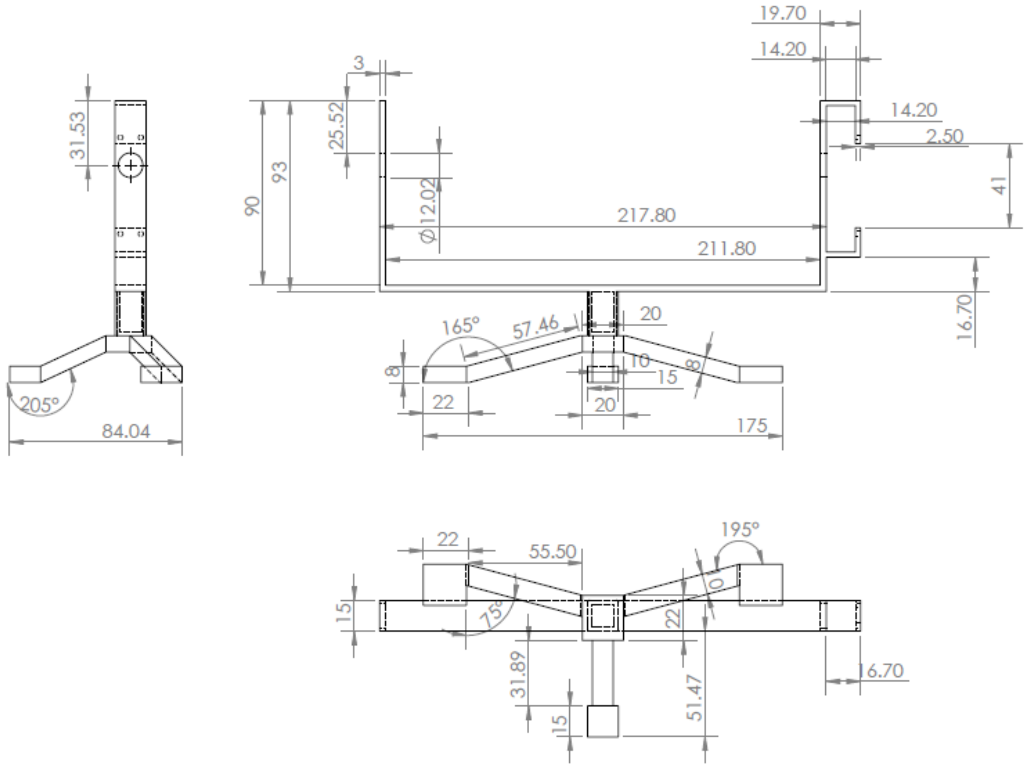


Figure 2.5: Dimension for structural links of robot eye mechanism

CHAPTER 3: DESIGN OF ELECTRICAL AND CONTROL SYSTEMS

3.1. System overview

The electrical system should be capable of supplying power to all the actuators of the robotic eye while facilitating flexible controlling of those actuators. As the selected actuators have different voltage and current ratings, a system had to be implemented to supply the required voltage and current for each actuator while ensuring that safety is provided to all the components. Effects on the controlling hardware due to the electromagnetic effects of the DC motors too had to be reduced to ensure the safety of the costly control equipment. The total electrical and control system layout design is shown in the figure 3.1 below.

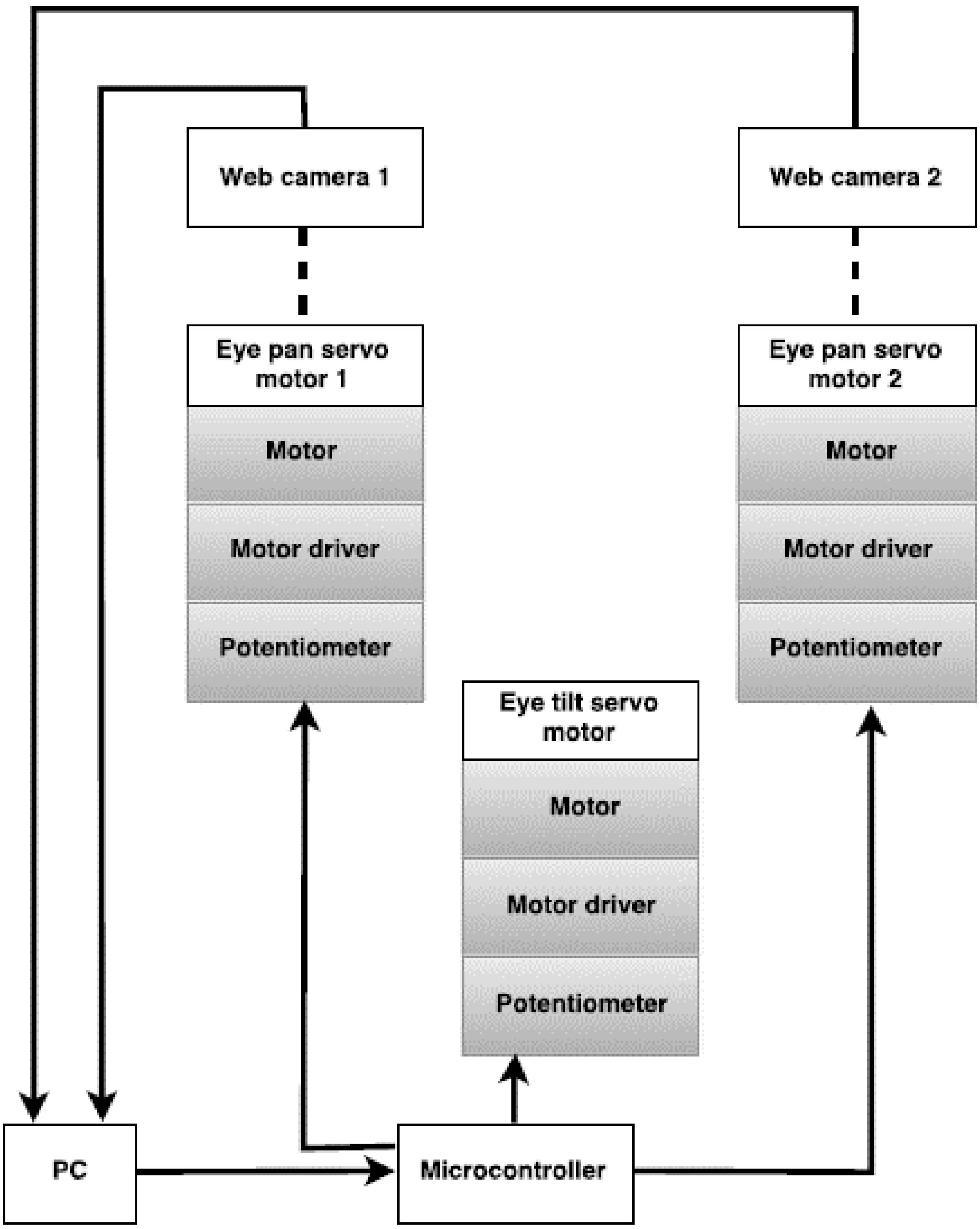


Figure 3.1: Electrical and control system overview

As shown in the figure the three DC motors that are used to implement the three degrees of freedom in the robot eyes are controlled via separate motor drives by the micro controller. The servo motors that are used to implement the three degrees of freedom in the eye mechanism are controlled directly by the microcontroller.

The absolute encoders which are of potentiometer type are also connected to the microcontroller to supply position feedback of the three DC motors for controlling purposes. The power supply for the potentiometers are also supplied by the microcontroller which assures that the supplied voltage is accurately at 5V. This has a good impact on the accuracy of the readings given by the encoders.

The two cameras which act as the vision feedback sensors are connected directly to a station PC which carries out the complex calculations involved with image processing and all the other high level algorithms associated with interactive features that have been implemented on the robotic eye platform once it was constructed.

The Station PC and the microcontroller are connected with each other through a serial data link (RS232). This data link is a bi-directional link which allows passing of data from PC to the microcontroller to provide commands and position information calculated by image processing functions in some operational modes and for monitoring of different variables controlled via the microcontroller using the PC interface.

3.2. Microcontroller selection

The robotic eye has 3 degree of freedoms and therefore, the controller should be able to handle all those three motors. In the design there are 3 servo motors which have shaft coupled encoders. For controlling of the three servo motors there is a requirement of generating 3 PWM signals and for the DC motors there is a requirement of another 3 PWM signals plus 3 digital outputs. Therefore the microcontroller should have a minimum of 3 PWM output channels. Since DC servo motors having potentiometer type encoders the microcontroller need to have minimum of 3 analog to digital conversion channels. USB communication also is preferable feature for ease of establishing serial communication with the PC [24]. The microcontroller should have high processing power also. Considering all the factors we selected the Arduino Uno as the microcontroller.

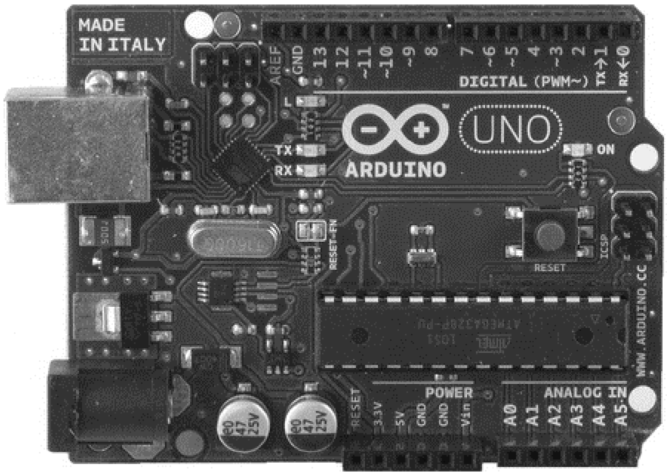


Figure 3.2: Arduino UNO microcontroller board

Key features:

- Microcontroller-ATmega328P
- Analog Input Pins 6
- Operating Voltage 5V
- Clock Speed 16 MHz
- Digital I/O Pins 14 (of which 6 provide PWM output)
- Flash Memory 32 KB of which 0.5 KB used by boot loader
- SRAM 2 KB
- EEPROM 1 KB
- USB communication facility
- Open source
- Libraries are freely available
- User friendly

3.3. Selection of softwares

The software was implemented during the entire of the project. First, object tracking purpose was used to obtain data from the objects in the domestic environment as sequence of movements in front of a robot eye. Second, searching-object purpose was used to obtain data from the objects in the domestic environment as sequence of movements in front of a robot eye.

3.3.1. Selection of software for image processing

Interactive featured implemented in the robotic eye are based on the vision feedback. Therefore it is required to perform image processing for the implementation of vision based interactive features. There are number of software and libraries available for development of computer vision applications.

3.3.1.1.Comparison of available options

In computer vision, the most of researches are used OpenCV using C++, OpenCV using Python, or MATLAB. Until recently computer vision was a research area in its infancy. Most of the computer vision tool of choice was MATLAB, and for the longest time OpenCV reduced in comparison to what MATLAB and its community had to offer. In order to deciding the applicable software package is used a comparison of the available packages. OpenCV and MATLAB software were considered for the most suitable comparison due to image processing [23].

Table 3.1: Comparison between OpenCV and MATLAB

OpenCV	MATLAB
Need to worry about low level programming issues	Easy to use
Free software	Built in memory management
Super-fast compared to the MATLAB	Has a good help section
Short development path from prototype code to embedding code	Very slow to execute code
Inexpensive	Expensive
Specially designed for computer vision	Generic programming language use for many applications

OpenCV is an open source computer vision and machine learning software library. OpenCV was built to provide a common infrastructure for computer vision applications and to accelerate the use of machine perception in the commercial products. Although MATLAB offers an efficient high level platform for prototyping and testing algorithms, its performance doesn’t compete with a well-designed and optimized C/C++

implementation of OpenCV. Therefore we decided to use OpenCV for image processing.

3.3.1.2.OpenCV library

OpenCV is an open source computer vision and machine learning software library. OpenCV was built to provide a common infrastructure for computer vision applications and to accelerate the use of machine perception in the commercial products. Being a licensed product, OpenCV makes it easy for businesses to utilize and modify the code.

The library has more than 2500 optimized algorithms, which includes a comprehensive set of both classic and state-of-the-art computer vision and machine learning algorithms. These algorithms can be used to detect and recognize faces, identify objects, classify human actions in videos, track camera movements, track moving objects, extract 3D models of objects, produce 3D point clouds from stereo cameras, stitch images together to produce a high resolution image of an entire scene, find similar images from an image database, remove red eyes from images taken using flash, follow eye movements, recognize scenery and establish markers to overlay it with augmented reality, etc. The library is used extensively in companies, research groups and by governmental bodies.

It has C++, C, Python, Java and MATLAB interfaces and supports Windows, Linux, Android and Mac OS. OpenCV leans mostly towards real-time vision applications and takes advantage of instructions when available. There are over 500 algorithms and about 10 times as many functions that compose or support those algorithms. OpenCV is only a library for computer vision, so it is required to use an interfacing programming language [23]. OpenCV is written in C++ and its primary interface is in C++, but it still retains a less comprehensive (though extensive) older C interface. Apart from that full interfaces are available in Python and Java also. A brief comparison of features of these interfaces is shown below.

Table 3.2: Comparison features of programing languages

C++	Python	Java
Need to have good programming skills	Have automated memory management	Recently implemented interface
Have official tutorials	Easy syntax	Less resources available
Portability on embedded system	Portability on embedded system	Portability on embedded system
Fast execution	Slow execution	Slow execution

3.3.2. Serial communication

For establishing serial communication between the pc and the Arduino, tserial library by Thierry Schneider was used.

3.3.3. Software used for programming the controller

Arduino is an open-source prototyping platform based on easy-to-use hardware and software. Arduino boards are able to read inputs. To do so it’s need use the Arduino programming language (based on wiring), and the Arduino Software (IDE), based on processing. Arduino boards are completely open-source, empowering users to build them independently and eventually adapt them to their particular needs. The software, too, is open-source, and it is growing through the contributions of users worldwide.

There are many other microcontrollers and microcontroller platforms available for physical computing. Parallax Basic Stamp, Netmedia's BX-24, Phidgets, MIT's Handyboard, and many others offer similar functionality. All of these tools take the messy details of microcontroller programming and wrap it up in an easy-to-use package. Arduino also simplifies the process of working with microcontrollers but it offers some advantages for users.

- Inexpensive:

  Arduino boards are relatively inexpensive compared to other microcontroller platforms. The least expensive version of the Arduino module can be assembled by hand, and even the pre-assembled Arduino modules cost less than $50.

- Cross-platform:

  The Arduino Software (IDE) runs on Windows, Macintosh OSX, and Linux operating systems. Most microcontroller systems are limited to Windows.

- Simple, clear programming environment:

  The Arduino Software (IDE) is easy-to-use for beginners, yet flexible enough for advanced users to take advantage of as well. For teachers, it's conveniently based on the Processing programming environment, so students learning to program in that environment will be familiar with how the Arduino IDE works.

- Open source and extensible software:

  The Arduino software is published as open source tools, available for extension by experienced programmers. The language can be expanded through C++ libraries, and people wanting to understand the technical details can make the leap from Arduino to the AVR C programming language on which it's based. Similarly, you can add AVR-C code directly into your Arduino programs if you want to.

- Open source and extensible hardware:

  The plans of the Arduino boards are published under a Creative Commons license, so experienced circuit designers can make their own version of the module, extending it and improving it. Even relatively inexperienced users can build the breadboard version of the module in order to understand how it works and save money.

CHAPTER 4: ADAPTIVE VISION ATTENTIVE ROBOT EYE

4.1. Overview of adaptive vision attention

In humans, the mechanism that determines which part of the sensory input is currently most promising is called selective attention. Accordingly, its call robots that attend to the most promising part of their sensor data “Attentive Robots”. The term “attention” is used in many contexts and many definitions exist. It is a term of common language, it is an active research area in psychophysics since many decades, and it is frequently used in machine vision and robotics to refer to mechanisms that focus further processing on regions of interest. The latter perspective of attention is very broad, in principle, any pre-processing method of sensor data could be called attentional since it focuses further processing on parts of the data. It believes that closely mimicking the human system has the advantage that it results in human-like behavior, which is beneficial for systems that should interact with humans in a natural and natural manner. Therefore, in this article, I focus on methods that are based on concepts of human perception. Following this direction, one of the best fitting definitions of attention comes from attention is the without cognitive process of selectively concentrating on one aspect of the environment while ignoring other things.

While the concept of attention exists for all senses, most research focuses on the visual part of attention. This is true both for human visual attention, due to the fact that vision is the most important sense in humans, and for computational attention system. Thus, with a few exceptions the approaches mentioned in this article focus on analyzing visual data. In this research, I will give an overview of the current state of the art in computational adaptive vision attention systems for autonomous robots. While being far from a thorough overview, I aim to give the reader an impression of what adaptive vision attention systems can do for object tracking and object searching robot eye.

Robot eye module generates the object tracker trajectory of an object over time by locating its position in every frame of the video. In order to implementation of the adaptive vision field based on human robot activities, the tasks of detecting the objects in domestic environment and establishing correspondence between the object instances across frames are performed simultaneously. The possible object region in every frame is obtained by means of adaptive robot attention control module as shown in figure 4.1, and then it's tracking the object across frames.

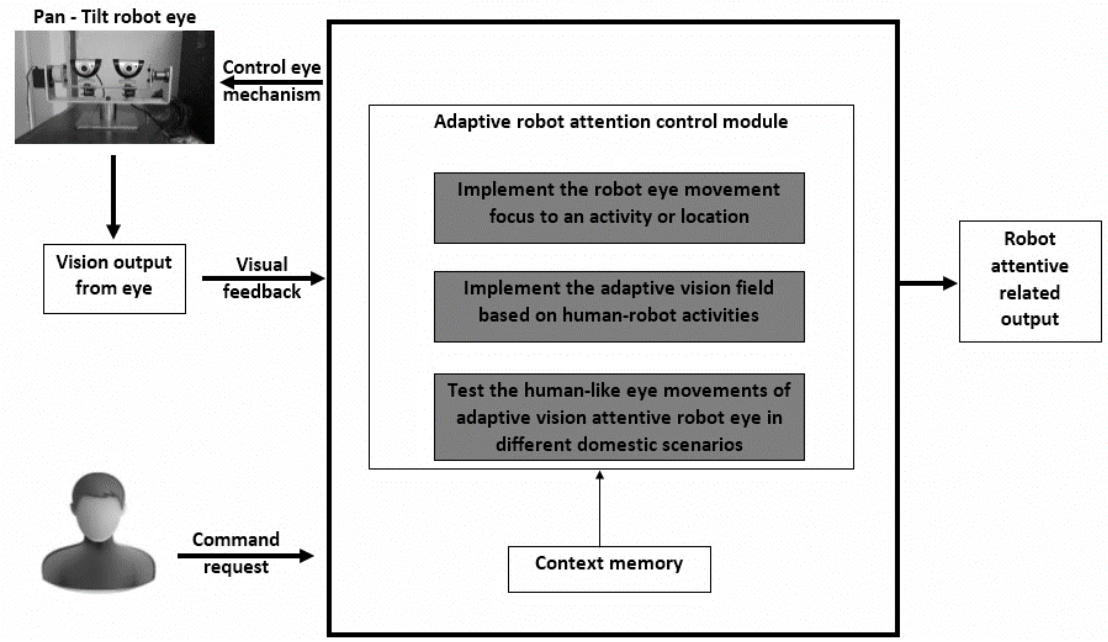


Figure 4.1: Adaptive robot attention control module

4.2. Implement the robot eye movement focus to an activity or location

To focus to the activities or location of the robot eye, developed a 2-DOF of a camera orientation system. The use of these systems to imitate the human eye movements. Both versions operate in the same way, three DC motors, respectively, to simulate the robot eye movements focus to an activity or location and orient the robotic eye in manner similar to the human eye. In this systems, supporting the robotic eye and allowing it to rotate around the axes requires mounting the robotic eye in front of a rotary joints accordingly.

4.3. Implement the adaptive vision field based on human-robot activities

The human-robot vision system that mimics key visual functions of the human brain promises to robots maneuver quickly and safely through adaptive vision field through the domestic environments. The motion of changing adaptive vision field which used a search objects in domestic environment. Then, the distance to measure from focused object to other objects and according to that the robot eye search more human-like manner each object by using depth map analysis . The user was asked to search object a feedback whether the movement is **"near"**, **"middle"**, and **"far"**. The adaptive field of view of the robot eye in a domestic scenario is shown in the Figure 4.2.

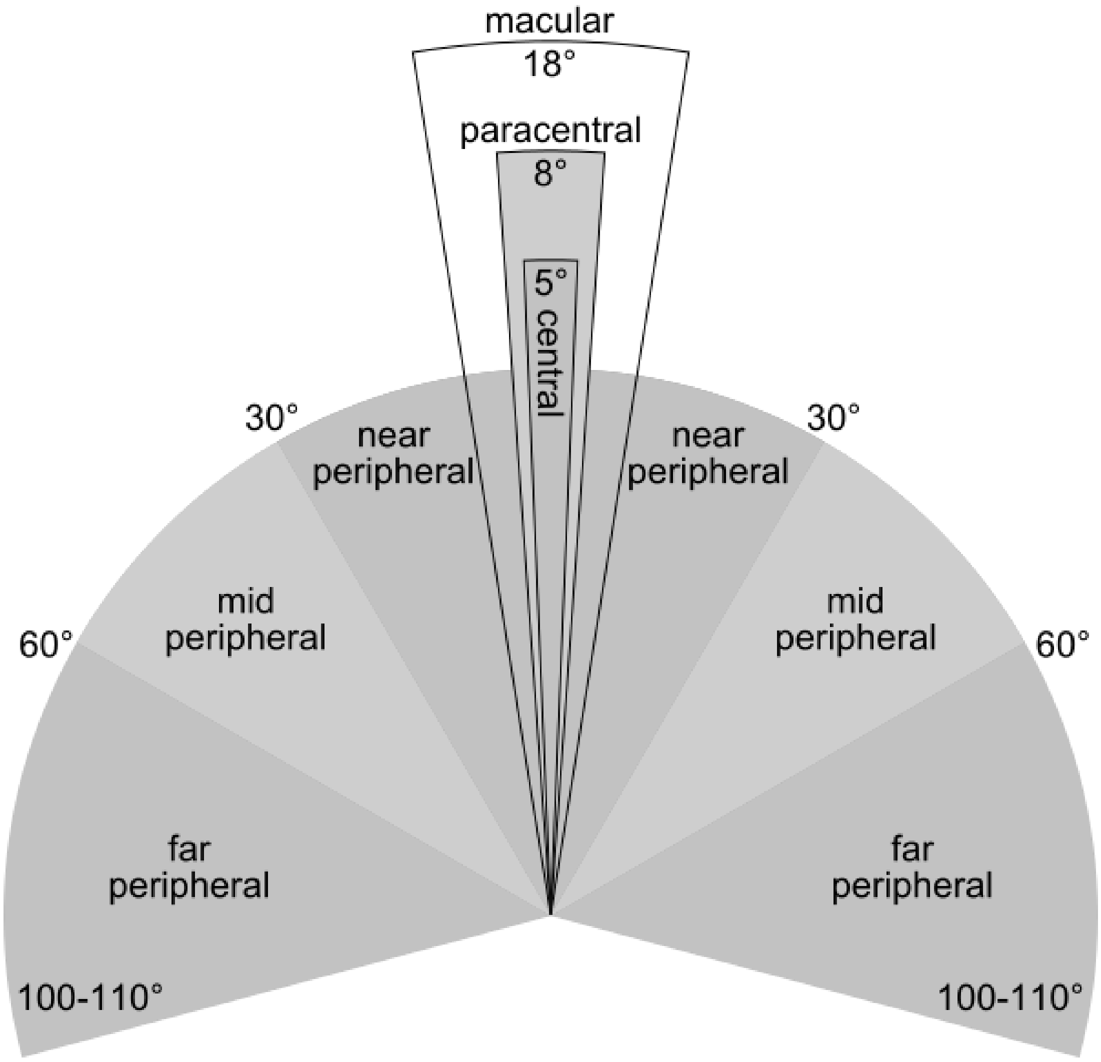


Figure 4.2: The adaptive field of view of the robot eye in a domestic scenario

4.4. Test the human-like eye movement of adaptive vision attentive robot eye in different domestic scenarios

To realize human-like eye movement characteristics in a robotic eye camera system, an effective mechanical structure and actuators are necessary. The mechanism should be designed with a minimum moment of inertia in its moving parts in order to maintain a compact volume. A compact design for the mechanism is crucial, as two of these eye-camera robotic models are to be put side by side together with other complex structures and mechanisms within robot eye module.

Robot attention system identifies potential objects in domestic environment using the vision system, and focuses on these objects to collect detailed images using the vision system, so that these images can be further processed for object focusing. Identifying potential objects correctly due to the presence of confusing backgrounds and the vast appearance and size variations amongst the items that refer to as an objects.

Rotation of the robot eye is an involuntary reaction of the human vision system that is not yet fully understood. The capacity of a human observer to detect movements or the significance that it might or might not have as a source of social information is varying. Thus, robot eye limit the design of the system to orient the robot eye around the two most important axes of the human eye pan and tilt, as equipping it with a rotation response as shown in Figure 4.3. Since, in normal conditions, humans orient their eyes only in a large range, but robot eye design to cover a workspace of ± 35 degrees for both orientations with respect to the environment and domestic scenarios. This design criterion is consistent with the design requirements of the prototypes presented in Chapter 2.

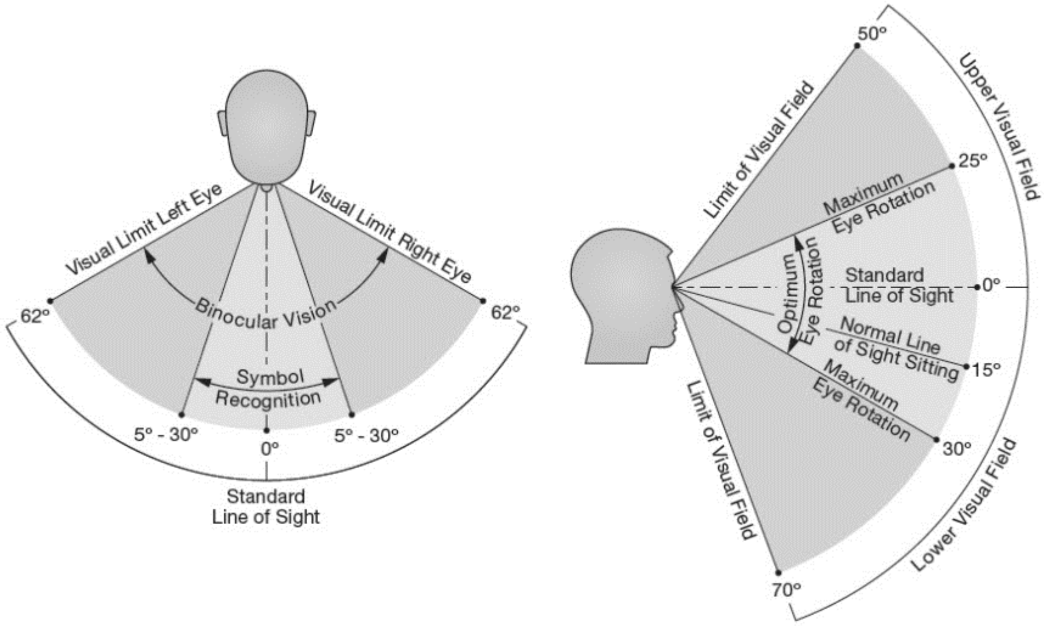


Figure 4.3: The adaptive field of view of the human eye in a domestic scenario

4.5. Object tracking

Object detection and tracking is considered as important subject within the area of computer vision. Availability of high definition videos, fast processing computers and exponentially increasing demand for highly reliable automated video analysis has created a new and great deal for modifying object tracking algorithms. CAMShift algorithm is based on object tracking in the domestic environment. Video analysis has three main steps mainly: interesting moving objects detection, the tracking the object detected from frame to frame and visualizing and analyzing to identify the behavior of the object in the entire video [25]. When the object being tracked moves out of the viewing range of the camera, the setup is automatically adjusted to move the camera so as to keep the object range given in table 2.2. The system is capable of handling entry and exit of an object. The performance of the proposed object tracking system has been demonstrated in real time in domestic environment by including a variety of disturbances and a means to detect a loss of track as shown in Figure 4.5. Also Figure 4.4 shows an object focusing/tracking algorithm.

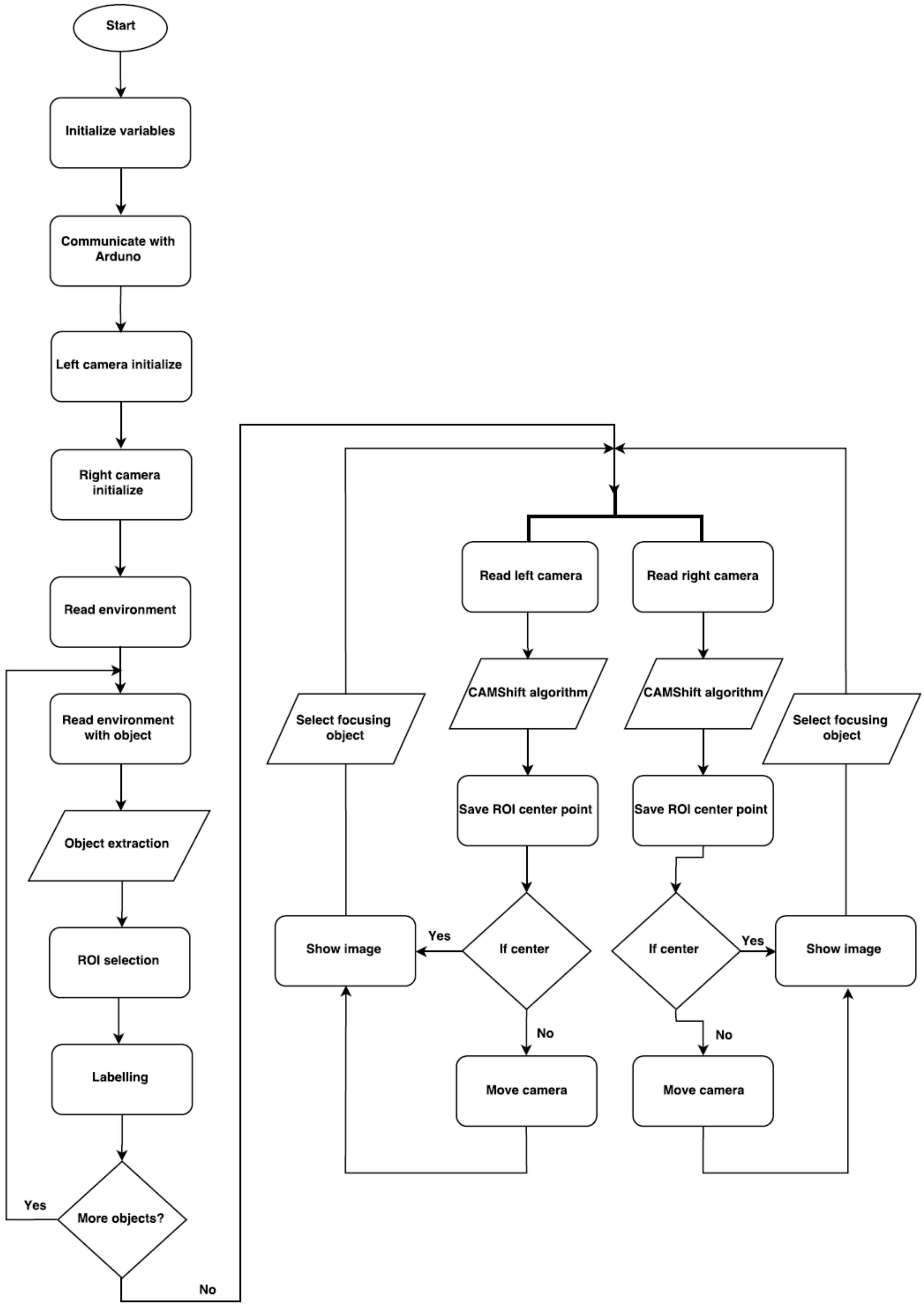


Figure 4.4: Object focusing/tracking algorithm

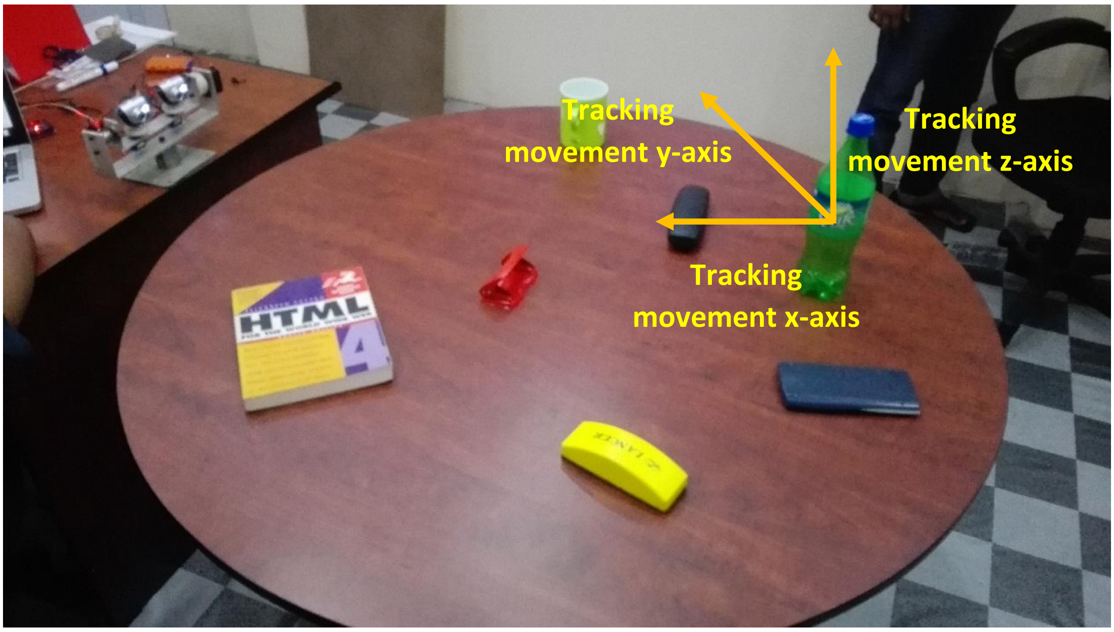


Figure 4.5: Tracking objects in domestic environment

4.6. Object searching

Object searching can be defined as the process of segmenting an object of interest from a video scene and keeping search of its position in order to extract useful information. Real-time object tracking and searching is a critical tasks in many computer vision applications such as surveillance, driver assistance, gesture recognition, service robots and man machine interface. Here, object searching in domestic environment is done using OpenCV software on circle platform and the implementation of the searching system. CAMShift algorithm is based on object searching in the domestic environment. In this development, algorithms have been developed and tested on laptop computers.

Due to its programming flexibility, easy availability and user friendly, OpenCV has been used to implement object searching scenario. The computational complexity of object searching algorithms is quite high. The major contribution of this research is implementation of object searching in domestic scenarios using OpenCV.

Searching is made simple by imposing constraints on the appearance of objects. In domestic scenario, I have tried to minimize the number of constraints on the distance between the each objects. Searching algorithm is capable of handling the each distance between an each objects and directly measured from the initial focused object (object#1) by the robot eye as shown in Figure 4.6.

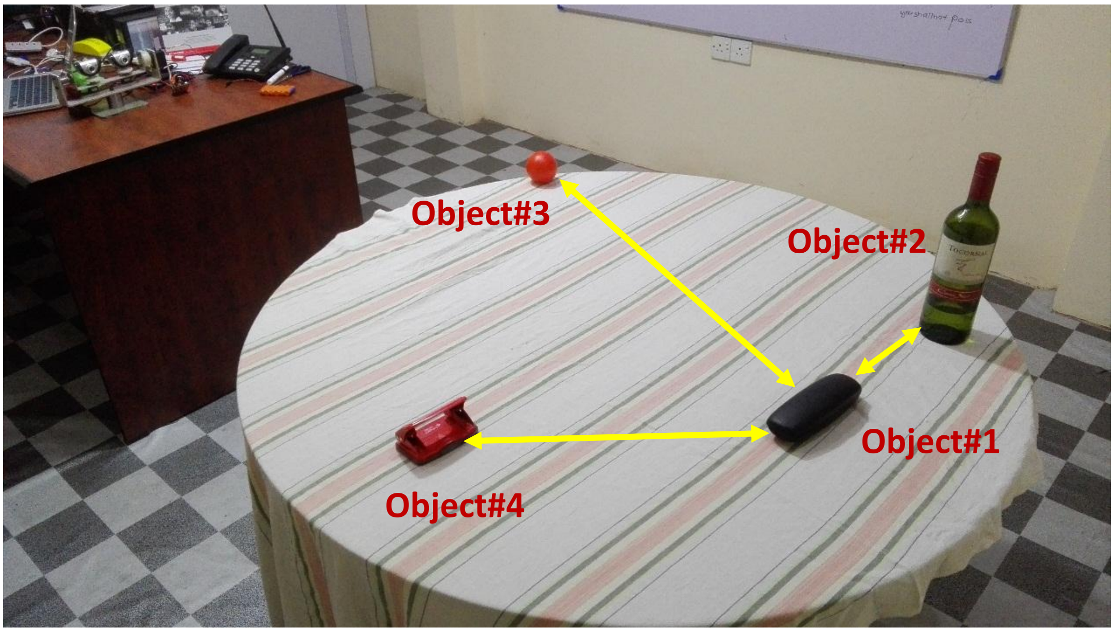


Figure 4.6: Searching objects in domestic environment

Template-based searching fits better into my scenarios. It uses a reference template of the object and searches it using image differences. These works nicely for well textured objects and small inter frame displacements. In order to improve the efficiency of the searching and to deal with more complex objects and camera motions, approaches were proposed.

4.7. Object tracking and searching algorithm using CAMShift algorithm

The CAMShift algorithm was derived from the earlier Mean Shift algorithm and is a simple, yet very effective, color-based tracking technique. It is applied as basic component for many advanced trackers and finds applications in fields such as perceptual user interfaces, face tracking, video surveillance, video editing, and computer vision based game interfaces, and others. CAMShift essentially climbs the gradient of a back projected probability distribution computed from re-scaled color histograms to find the nearest peak within an axis aligned search window. With this, the mean location of a target object is found by computing zero, first and second order image moments.

The position and dimensions of the search window are updated iteratively until convergence. One of the main drawbacks of standard CAMShift tracking is, that it is prone to tracking failures caused by objects with similar colors. The reason for this is that only the peak of the back-projected probability distribution is tracked without paying attention to color composition. For the same reason, objects with similar colors cannot be distinguished. Stable tracking despite appearance changes due to lighting or perspective and partial or full occlusion, and re-detection of lost objects are other problems of standard CAMShift.

Robot eye tracking is a process by which the point of gaze or the movement of the eyes is measured relative to the moving object in domestic environment. The video frames obtained from the web camera are subjected to pre-processing in order to help the searching and tracking algorithms work more efficiently and precisely. OpenCV software is chosen computer interface which is a library mainly meant for real time image processing that contains the functions used in my proposed algorithms. In this research work which propose to merge template matching algorithm to overcome these inaccuracies. Once detection is successful, template matching algorithm is used to track the objects in domestic environment after which the region of interest is targeted and searched for circular contours representing the dark, circular nature of the eye movement.

The traditional CAMShift algorithm uses color histogram as characterization of the target template for matching and only contains the frequency of a certain color value occurred in the image and loses the space location information of the pixel, when the background color is similar to the object and the emergence of occlusion and the background and occlusion will generate interference, resulting in the tracking method based on color histogram losing the target; meanwhile, the traditional color histogram is a global color histogram, computation is larger and is not conducive to real-time tracking. Therefore, the improved algorithm will continue the good characteristics of the color histogram and use a simplified main color histogram and an edge direction histogram which has robustness for the situation that the occlusion and back- ground object color are similar, together to describe the moving object, thus achieving the accurate and robust moving object tracking in complex environments. For the above deficiencies for CAMShift algorithm, this paper will improve it, and the specific improvement measures are as follows, firstly the quadratic difference method based on background motion estimation for object extraction is used, and the tracking process integrates the simplified main color histogram and the above concerned edge direction histogram as the target characteristics secondly, the obtained center is the initial search point, for similarity detection of the target tem plate and the candidate target, and we can search the optimal matching position and then to get the exact location of the target object in the current frame.

The specific implementation process of the improved CAMShift algorithm is shown in Figure 4.7. The improved algorithm introduces the object detection algorithm based on background motion parameter estimate, to achieve the effective tracking of the traditional CAMShift algorithm in the dynamic context, thus expanding its range of applications.

Meanwhile, in order to improve its tracking robustness in dynamic moving object tracking in video surveillance system context, the improved algorithm continues the good characteristics of the color histogram and uses a simplified main color histogram and an edge direction histogram which has robustness for the situation that the occlusion and background object color are similar, together to describe the moving object, which greatly improves the anti-jamming capability and robustness of tracking.

4.7.1. Improved CAMShift algorithm

To track targets, traditional CAMShift basically works as follows. First, target`s initial search window is selected and its color histogram is computed. Each frame of the sequence afterwards is converted to a probability distribution image relative to the target`s histogram. Then the new size and location of the target are computed via meanshift from this converted image, and are used as the initial size and location of the target for the next iterations of the algorithm.

In the next part explain the tracking approach and discuss how it overtakes the traditional CAMShift. For this research which use the tracking algorithm is primarily based on conventional CAMShift. The principles of CAMShift algorithm can be summarized in the following steps.

1. Search window initialization

In the majority of works proposed on CAMShift, the initial location of a target is selected manually by a user, and also use the same approach in algorithm. After manually locating the target by a surrounding rectangle called Region of Interest (ROI), its two dimensional color histogram is calculated for further processing in the next steps.

2. Color histogram generation

To obtain robust results, use Hue Saturation Value (HSV) color space in algorithm for the color histogram generation. HSV color space separates out color (H) from its saturation (S) and brightness (V) values, which would improve the tracking performance. For target representation, only choose hue (H) and saturation (S) color channels that represent the target`s main color features.

3. Motion segmentation

To do the motion segmentation in the algorithm which use the image differencing method which is one of the simplest and most used techniques to detect moving objects. The pixel by pixel intensity difference of the current frame and the reference background, a model is computed and the resulting image is threshold to segment the frame`s foreground from its background. The result is a coarse map of temporal changes, after some simple morphological operations on this image, the main moving parts are extracted to generate the image of foreground objects.

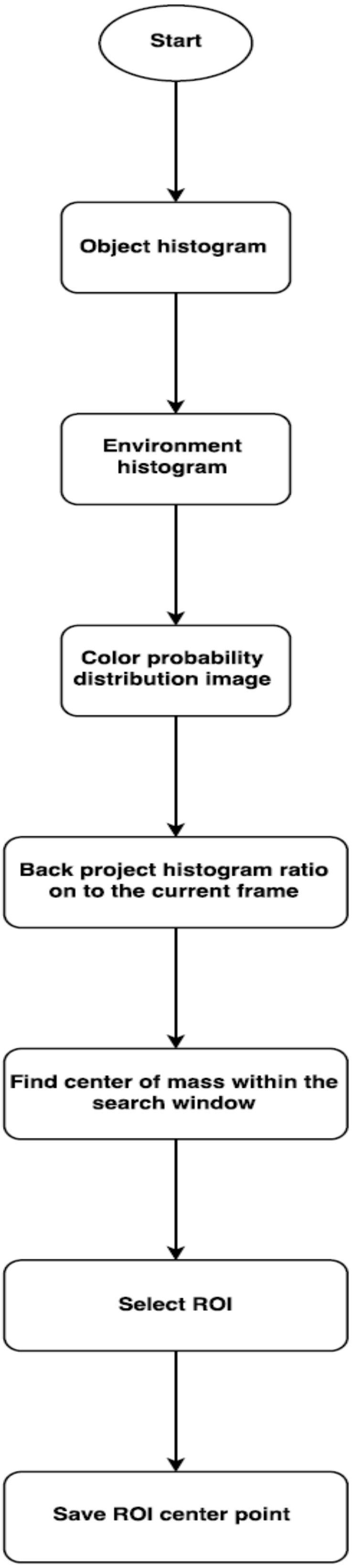
4. Probability distribution image generation

A common method to generate a probability distribution image is histogram back projection. Histogram back projection of the target histogram with a frame generates a probability distribution image in which each pixel`s value associates with the corresponding bin of the target histogram. In step 3 we calculate the back-projection of the target histogram with the resulting image of step 2. This will produce a probability distribution image which will be used by Meanshift to calculate the target`s new position.

5. Meanshift application

To calculate the new location of a target, the Meanshift algorithm is used. Meanshift takes a probability distribution image and an initial search window, computes the window`s center of mass, and then re-centers the window at the computed center of mass. This movement will change what is under the window, and so the re-centering process is repeated until the movement vector converges to zero. The last calculated center of mass will be the new location of the target.

Figure 4.7: Implementation process of improve CAMShift algorithm



4.7.2. Tracking objects with robot eye movements in domestic scenarios

According to tracking approach the object is represented as a rectangle shape. The robot eye takes the positional information of x and y co-ordinates of the system of the moving object as an input from the object template module as given command request by user input. This information is then used to extract a square image template from the last acquired frame. The module keeps on searching it in the frames captured from that point. Whenever found it displays a red overlaid rectangle over the detected object. If the template matching doesn't yield any result that signifying that the object has changed its appearance a new template is can be selected based on the user interest on the real time video sequence. Further on the Cartesian co-ordinates of the tracked object are calculated and passed on visual feedback from the pan-tilt robot eye to PC through serial communication for analysis.

An overview of the proposed system as a tracking objects with robot eye in domestic scenarios is given in Figure 4.8. The system starts with an initialization phase, and then uses the template based CAMShift tracking algorithm to track the object in domestic environment.

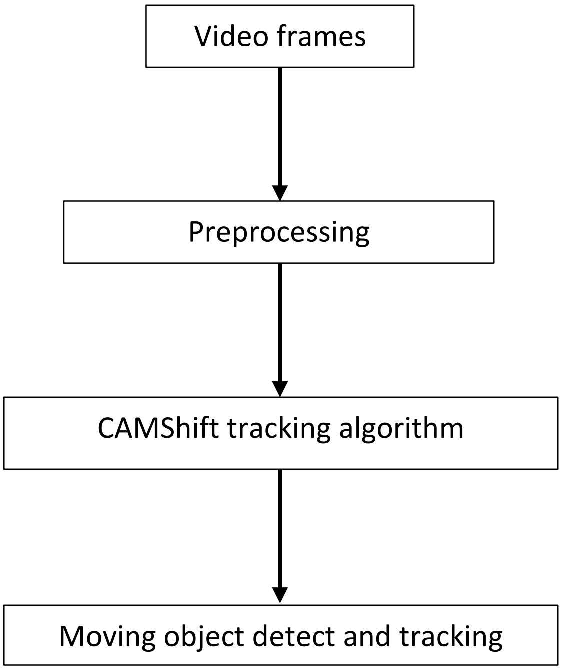


Figure 4.8: Object tracking system in video

Also the feature plays a very important role in the area of object tracking. Feature extraction techniques are helpful in various image processing applications object recognition. As features define the behavior of an image, they show its place in terms of storage taken, efficiency in classification and obviously in time consumption also. An overview of the proposed system as a feature extraction with robot eye in domestic scenarios is given in Figure 4.9.

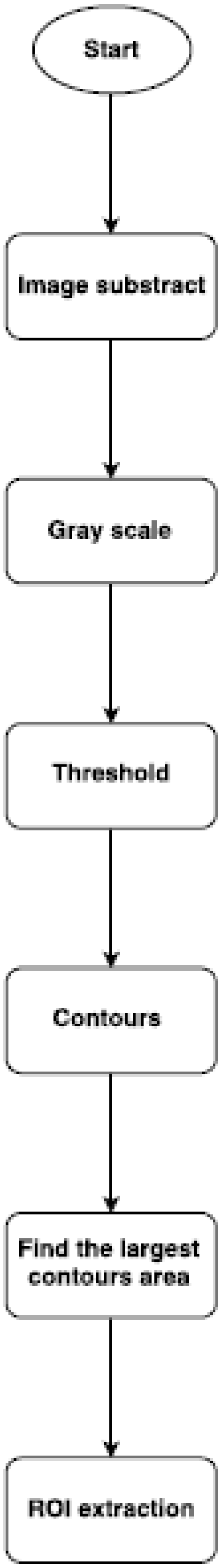


Figure 4.9: Feature extraction system in video

4.7.3. Searching objects with more human-like manner in domestic environment

The robot eye was placed in domestic environment and it was rotate by using commands of a human user. First, robot eye focused object by given user command. Then, the distance to measure from focused object to other objects and according to that the robot eye search more human-like manner each object by using depth map analysis . The user was asked to search object a feedback whether the movement is **"near"**, **"middle"**, and **"far"**. The robot eye in a test scenario is shown in the figure 4.10.

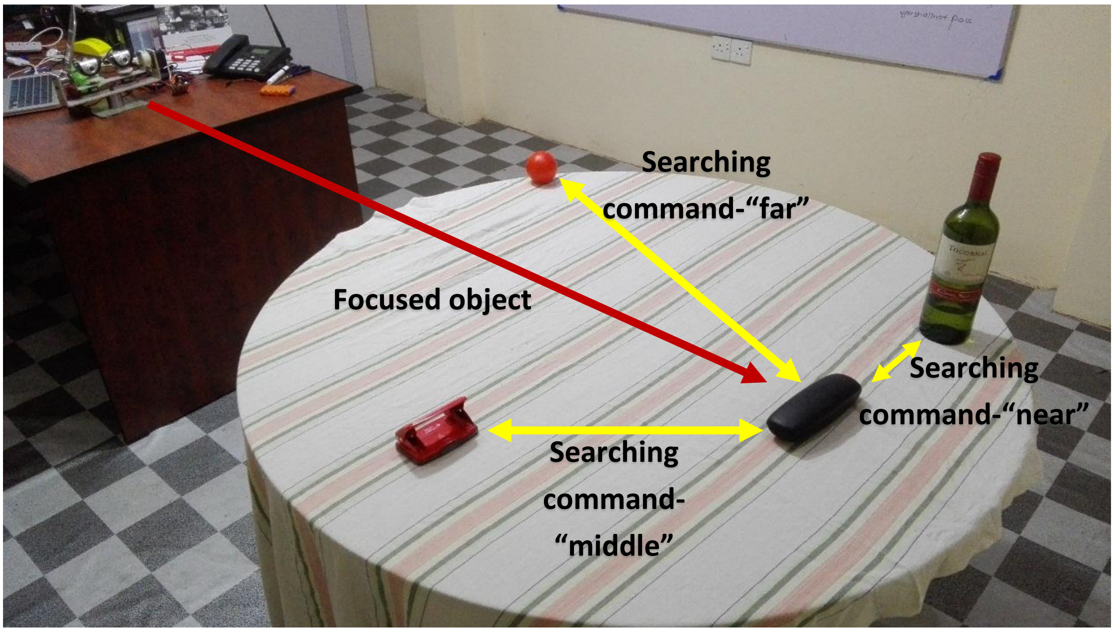


Figure 4.10: Searching objects with more human-like manner in domestic environment

Object searching is considered as important area of computer vision. Availability of high definition videos, fast processing computers and exponentially increasing demand for highly reliable automated video analysis has created a new and great deal for modifying object searching algorithms. CAMShift algorithm is based on object searching in the domestic environment. When the object being searched within the viewing range of the camera, the setup is automatically adjusted to move the camera so as to keep the object range given in table 2.2. An algorithm of the proposed system as a searching objects with robot eye in domestic scenarios is given in Figure 4.11. The system starts with an initialization phase, and then uses the template based CAMShift tracking algorithm to search the object in domestic environment.

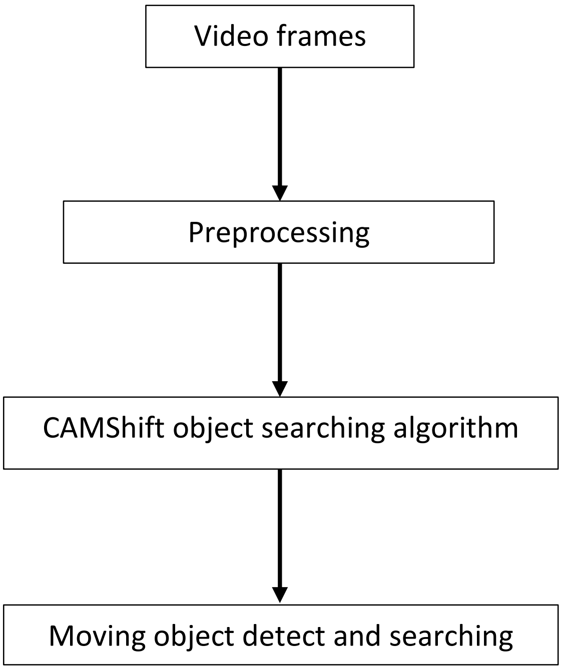


Figure 4.11: Object searching system in video

4.7.4. Searching objects in domestic scenarios with robot eye movements

CAMShift algorithm is used to search and find the object to domestic environment. This is different domestic scenario which use search object to environment. Also, for this process has used same object searching algorithm as given in Figure 4.11. The user was asked to search object a feedback whether the movement is **"near"**, **"middle"**, and **"far"**. The robot eye in a test scenario is shown in the Figure 4.12.

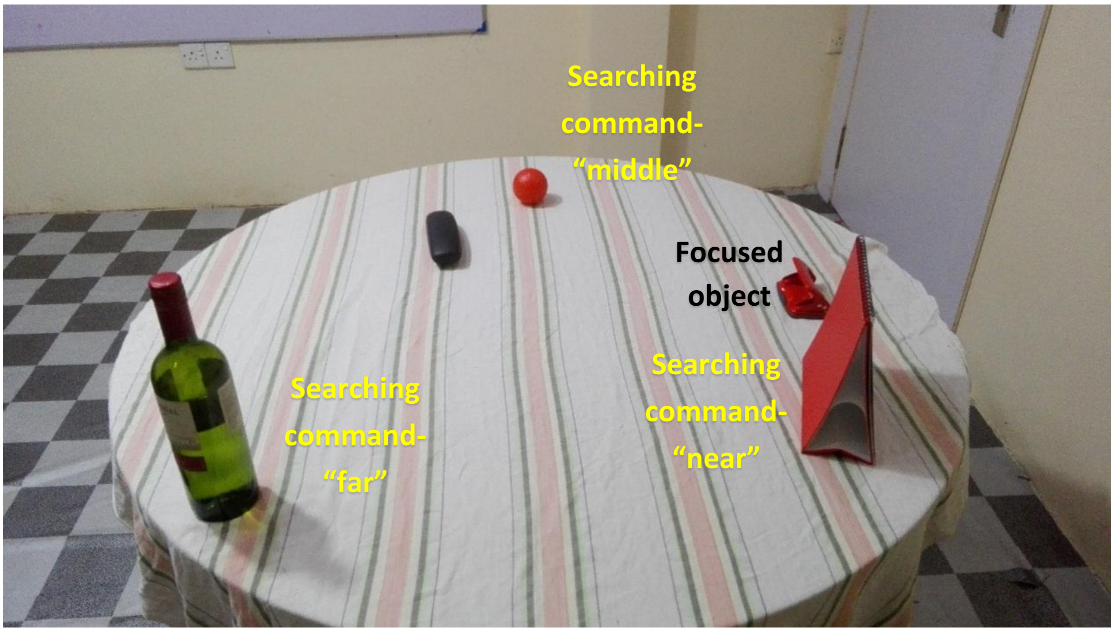
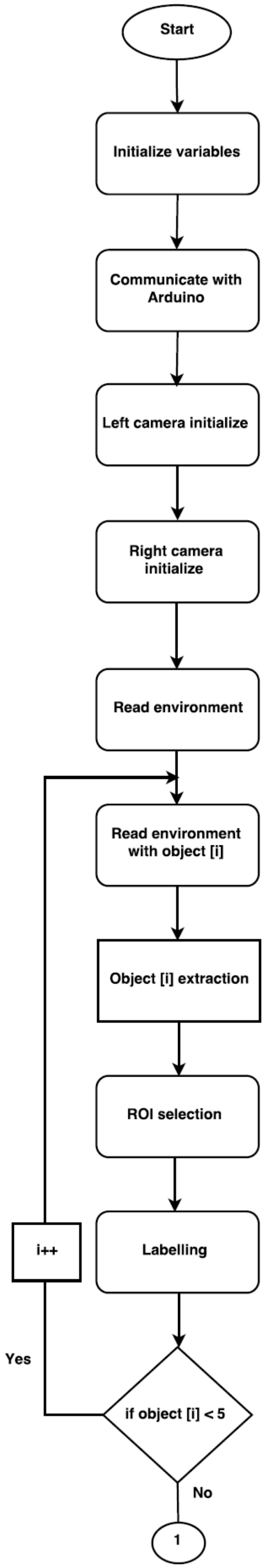


Figure 4.12: Searching objects in domestic scenarios with robot eye movements

Motivation of the objects searching was developing more human friendly interactive features on robot eye. Searching on objects which moves in front of the robot eye is a more sophisticated human-like interactive feature. This can also exhibit the sensors motors coordination to a greater extent.

Humans have movable eyes therefore when humans are gaze holding a moving object, the eye movements is occurred in inter related manner. Human can maintain the fixation of moving object with the coordinated eye movements. This involves keeping the visual axes of the eyes directed at the moving object. Following Figure 4.13 flow chart is shown the algorithm of the robot eye while searching to an objects.



Start
Initialize variables
Communicate with Arduino
Left camera initialize
Right camera initialize
Read environment
Read environment with object [i]
Object [i] extraction
ROI selection
Labelling
i++
Yes
if object [i] < 5
No
1

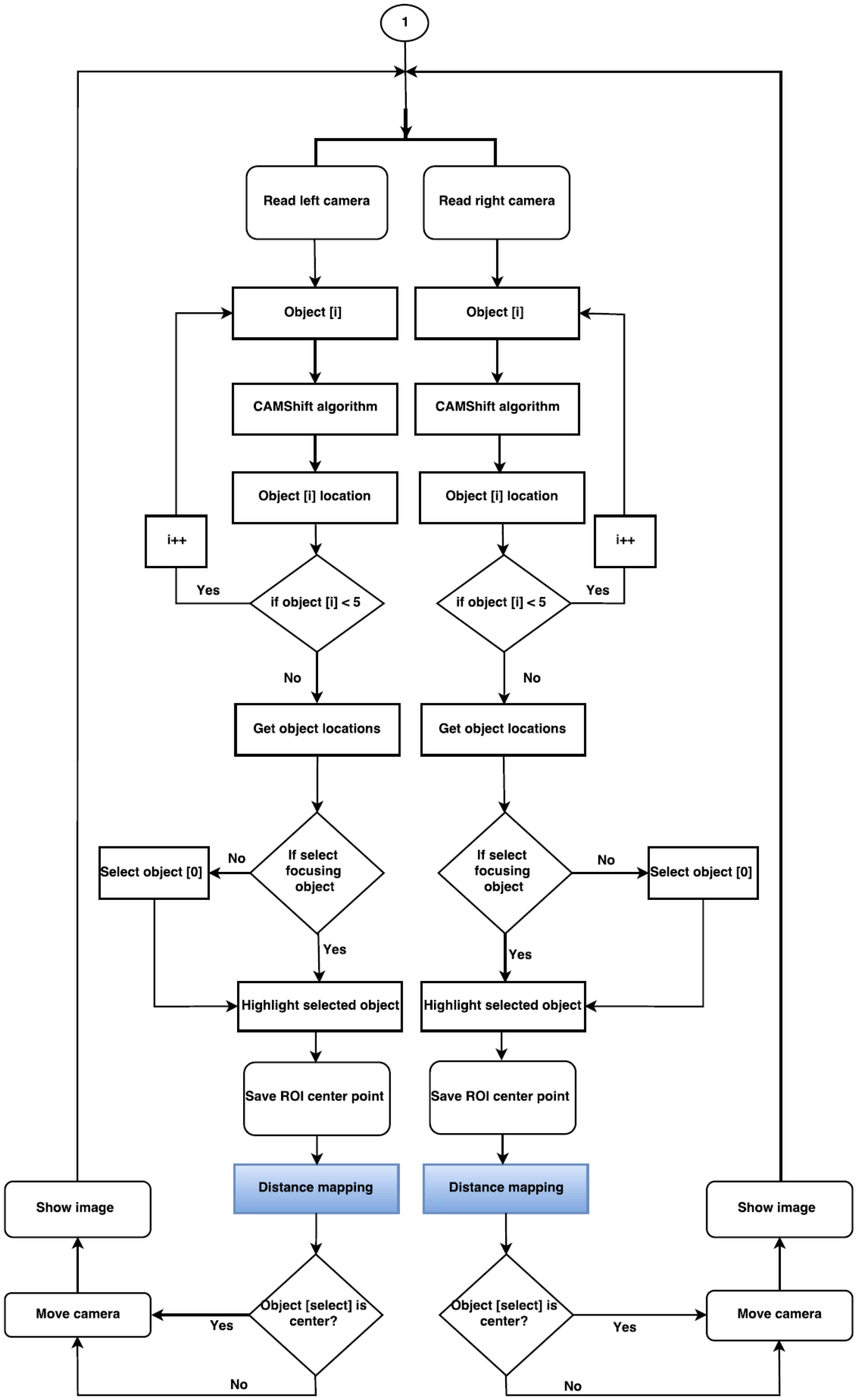


Figure 4.13: Object searching algorithm

The robotic eye has a binocular vision system. From the two cameras the position of the object relative to the frame of each camera can be obtained. Then the distance mapping used to identify the depth of image to utilize the searching movements such as “near”, “middle”, and “far” for maintaining the fixation as described below Figure 4.14.

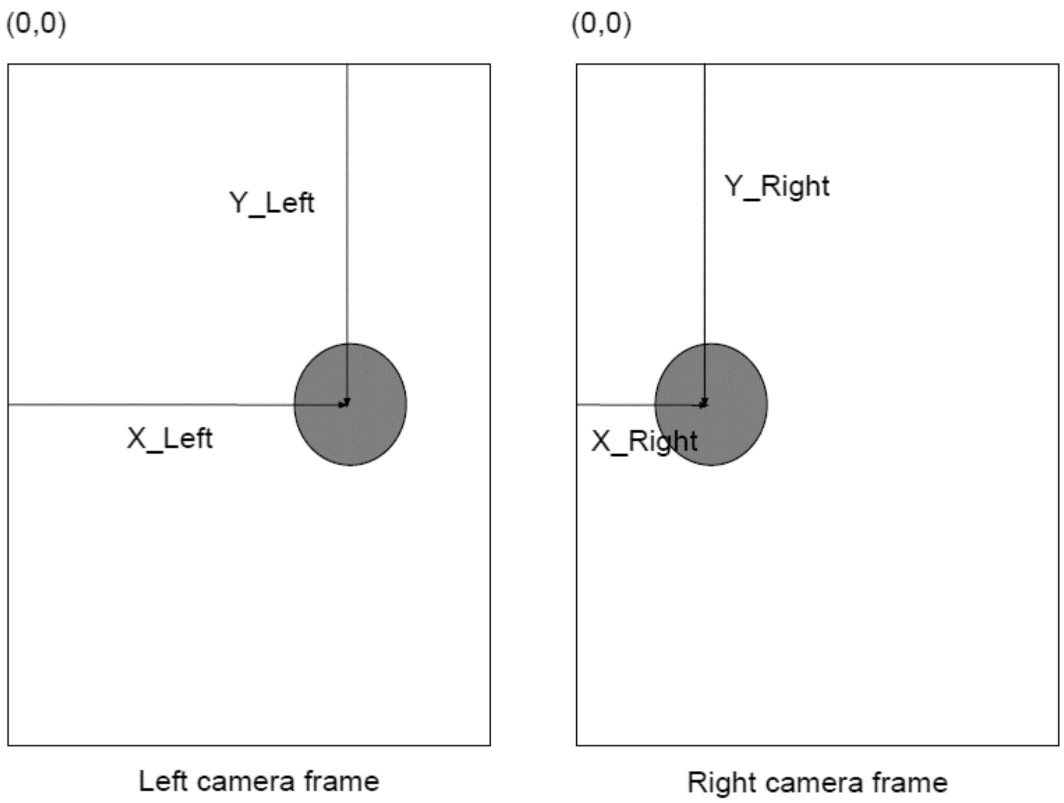


Figure 4.14: Coordinate frames relative to the distance mapping

The flow chart in the Figure 4.15 shows the distance mapping algorithm. The combined results along with a commanding signal are then sent to the microcontroller for actuator trajectory planning.

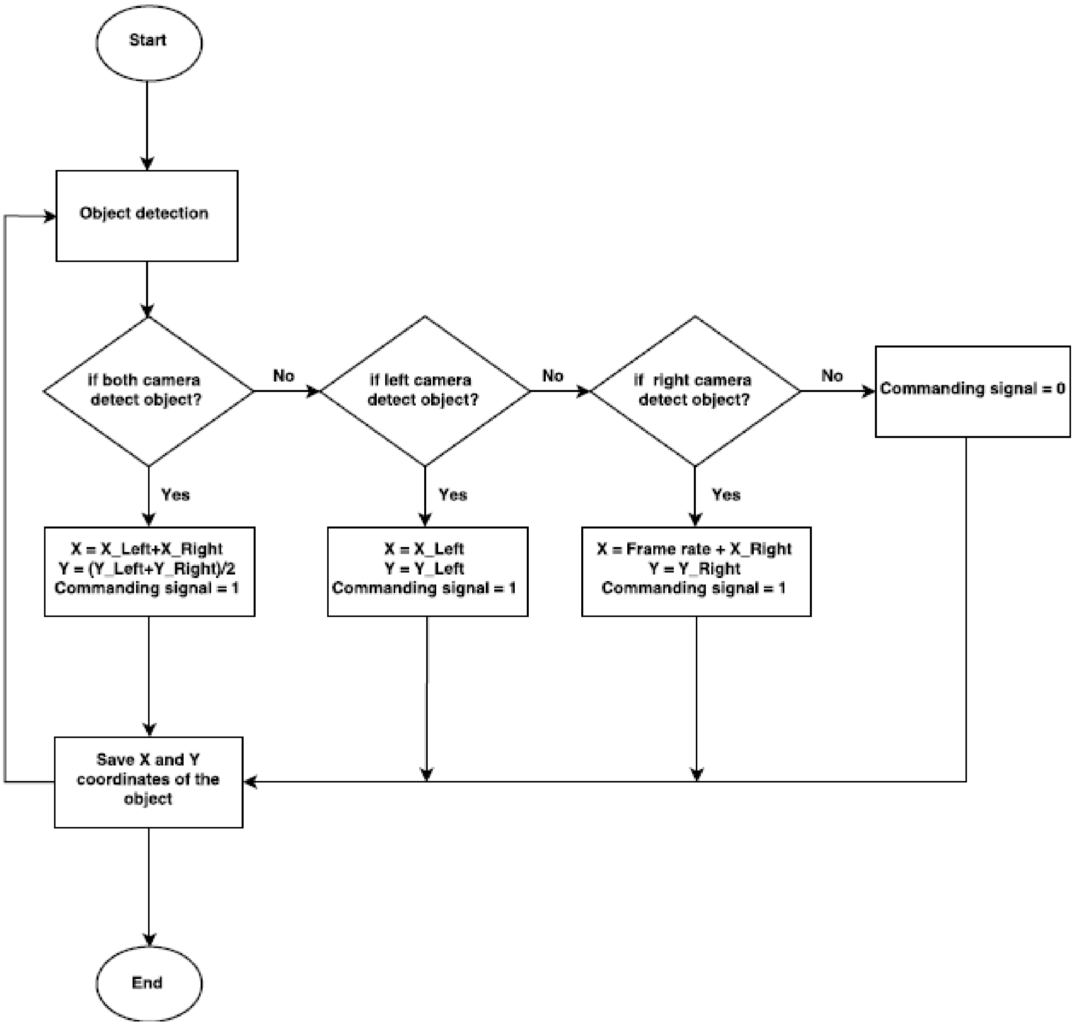


Figure 4.15: Distance mapping algorithm

The trajectory planning of the actuators and controlling of the actuators are done by the microcontroller. When the commanding signal = 1, movements of the actuators are calculated. According to those calculated angles commands are issued to drive them to required position.

The eye's 3 DOFs are directly controlled with a visual servoing mechanism, having the image positions of the object as feedback. The eye is then controlled in order to maintain the gaze of eyes as far as possible with their joint limits. This choice is motivated by biological behavior. The rationale behind this is the ability to track the object in any direction even if it moves very fast.

4.7.5. Perception of depth producing human-like reaction when searching objects in domestic scenarios with robot eye movements

When an object is searching, the person will react by closing his eyes through the environment for finding particular object. These are automatic reactions that happen very quickly without intention of the person. In this feature used to make an attempt to reproduce that natural human behavior in the robotic eye as close as possible. This interactive feature can be used for entertainment purposes and well as for human behavioral studies.

In the actual natural human reaction, eye is moved with the head in such scenario. But here only reproduce the eye movements. The eye movements are very fast and they vary from person to person. It was hard to find the proper data for such behavior modeling because of the unavailability of the literature.

The most important aspect of this interactive feature is the searching object. For searching object the robotic eye should have an idea about the distance between the object and the eye. Figure 4.16 shows the flow chart of the algorithm implemented for this interactive feature. To get a perception of the distance to the object, the disparity between $X - coordinates$ and $Y - coordinates$ of the center of detected object from left and right cameras are being analyzed.

$$\boldsymbol{Disparity\_X = X\_Left - X\_Right}$$
$$\boldsymbol{Disparity\_Y = Y\_Left - Y\_Right}$$

When the object is close to the eye, the disparity will increase and if the object is far away disparity tends to decrease. So, using two threshold values for disparities we divided the depth space in to three regions namely **"far"**, **"middle"** and **"near"**. The portion of the depth space where the object is in can be identified from that.

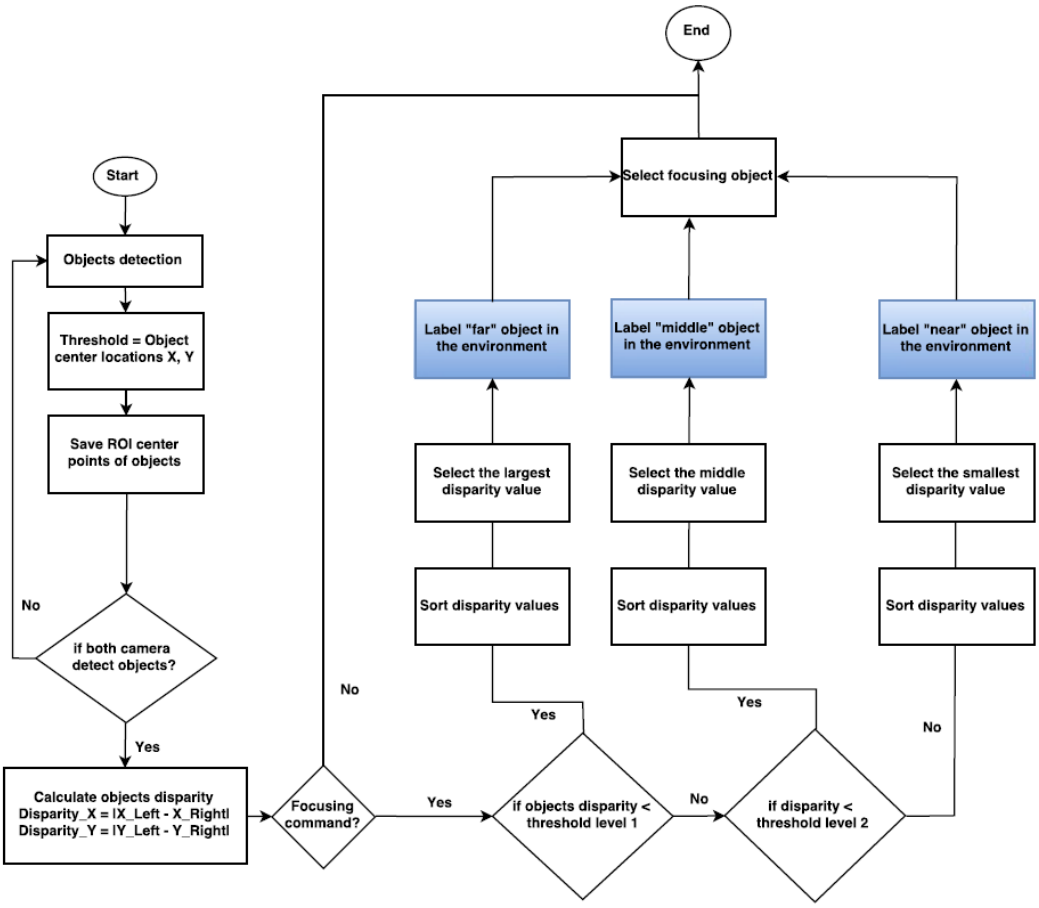


Figure 4.16: Human-like reaction when searching object

This threshold value can be used identify the center locations of objects in the domestic environment that should be there to produce responses. When this condition is satisfied the algorithm concludes that the objects are placed in the domestic environment and commanding signals are sent to the microcontroller to move robot eye to produce a reflexive reaction.

CHAPTER 5: RESULTS

The final robot eye design included two eyes that could pan individually and tilt together as shown in figure 5.1. The eyes would need to move at relatively similar speed and with similar range of motion as human eyes. Each eye contains a web camera to be used as part of an eye tracking and searching system or for general observation of what the robot is to be interacting with.

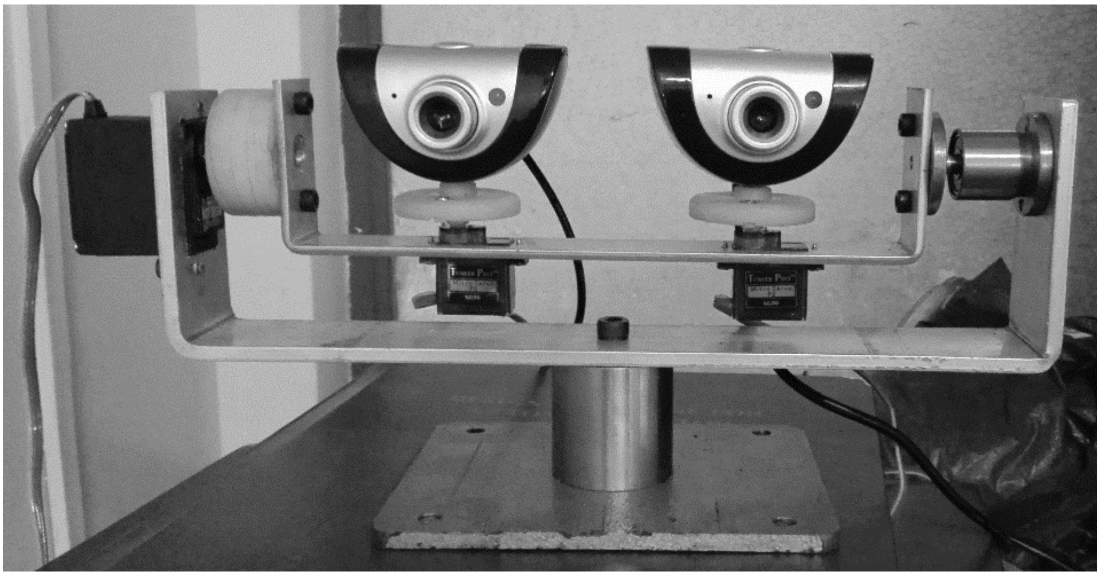


Figure 5.1: Final robot eye design

5.1. Results of the object focusing mode

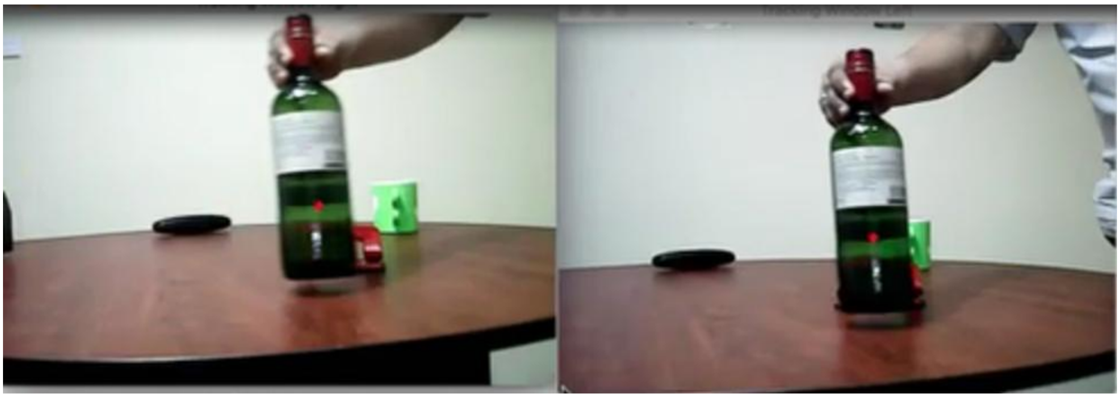
Results are presented of the experiment performed using the objects focusing in domestic environment as shown in below figures. The performance of the system was probabilistically studied by carrying out case studies. When the robotic eye is operated in a domestic environment the performance detection of background objects which is good exposure level of the camera. Moving object could be detected because of the ability of the robot eye to grab moving objects due to high frame rate of the cameras (30fps). Viewing angle of the cameras are low compared to the human eyes due to the domestic scenarios. Therefore the binocular vision is achievable to robot eye limited movement in given field which moved through the mid-point between the cameras.

5.1.1. Case I: Results of the object focusing mode (focusing to bottle)

Following results are presented the experiment performed of the robot eye using the web camera view to object focusing to the bottle in domestic environment as shown from Figure 5.2(a) to 5.2(f).

Left eye Right eye

Figure 5.2 (a): Results obtained for focusing mode to the bottle in domestic environment ($T = 0s$)



Left eye Right eye

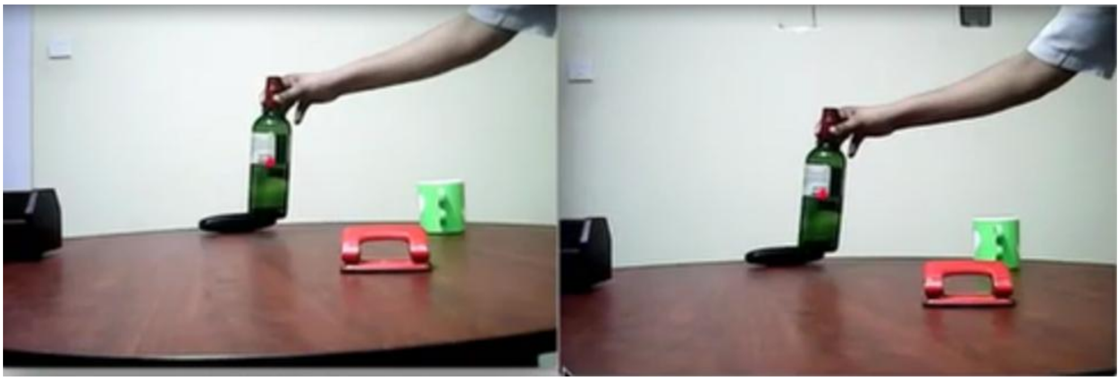
Figure 5.2 (b): Results obtained for focusing mode to the bottle in domestic environment ($T = 10s$)

Left eye Right eye

Figure 5.2 (c): Results obtained for focusing mode to the bottle in domestic environment ($T = 20s$)

Left eye Right eye

Figure 5.2 (d): Results obtained for focusing mode to the bottle in domestic environment ($T = 30s$)



Left eye Right eye

Figure 5.2 (e): Results obtained for focusing mode to the bottle in domestic environment ($T = 40s$)

Left eye Right eye

Figure 5.2 (f): Results obtained for focusing mode to the bottle in domestic environment ($T = 50s$)

Following results are presented the experiment performed of the robot eye using the environmental view to object focusing to the bottle in domestic environment as shown from Figure 5.3(a) to 5.3(f).

Figure 5.3 (a): Results obtained for focusing mode to the bottle in environmental view of domestic environment ($T = 0s$)

Figure 5.3 (b): Results obtained for focusing mode to the bottle in environmental view of domestic environment ($T = 10s$)

Figure 5.3 (c): Results obtained for focusing mode to the bottle in environmental view of domestic environment ($T = 20s$)

Figure 5.3 (d): Results obtained for focusing mode to the bottle in environmental view of domestic environment ($T = 30s$)

Figure 5.3 (e): Results obtained for focusing mode to the bottle in environmental view of domestic environment ($T = 40s$)

Figure 5.3 (f): Results obtained for focusing mode to the bottle in environmental view of domestic environment ($T = 50s$)

Figure 5.4 is presented the experiment performed of the robot eye joint angle movements for duration of $50s$.

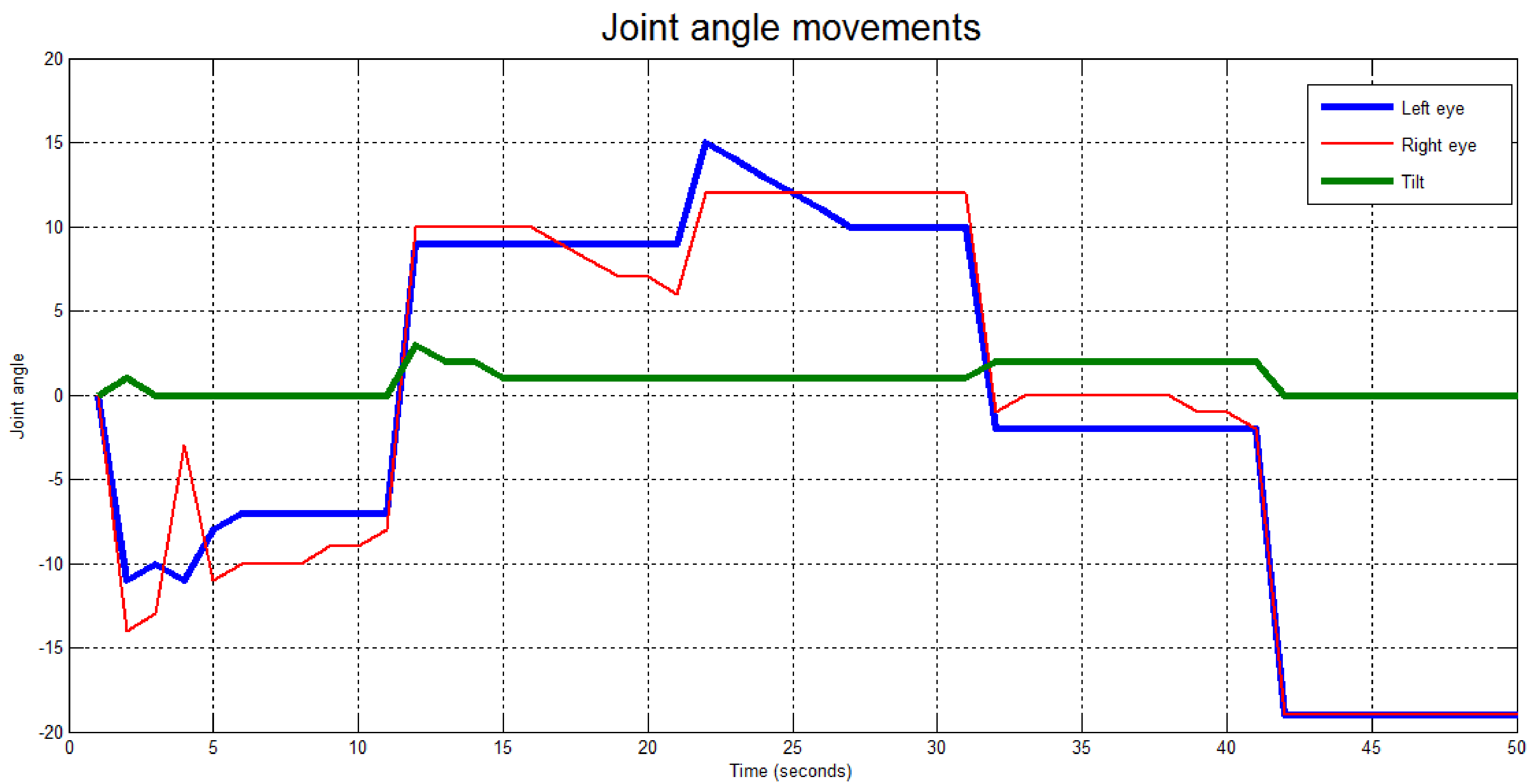
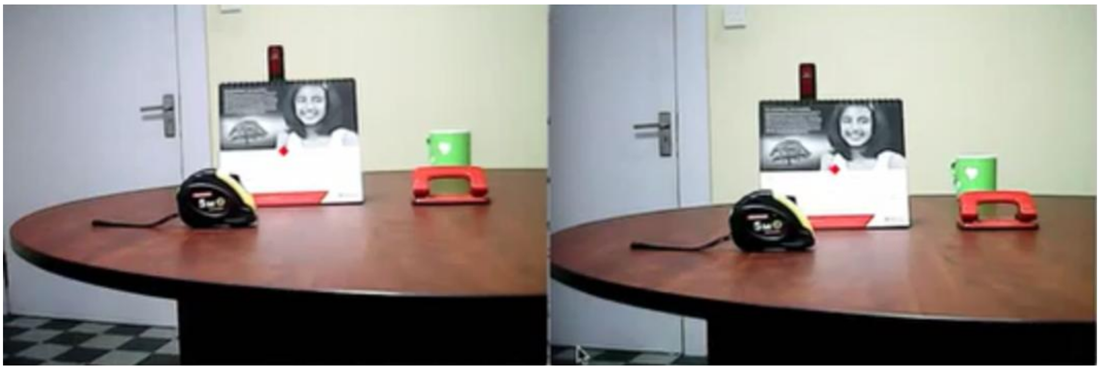


Figure 5.4: Joint angle movements of focusing to bottle

5.1.2. Case II: Results of the object focusing mode (focusing to calendar)

Following results are presented the experiment performed of the robot eye using the web camera view to object focusing to the calendar in domestic environment as shown from Figure 5.5(a) to 5.5(f).



Left eye Right eye

Figure 5.5 (a): Results obtained for focusing mode to the calendar in domestic environment ($T = 0s$)

Left eye Right eye

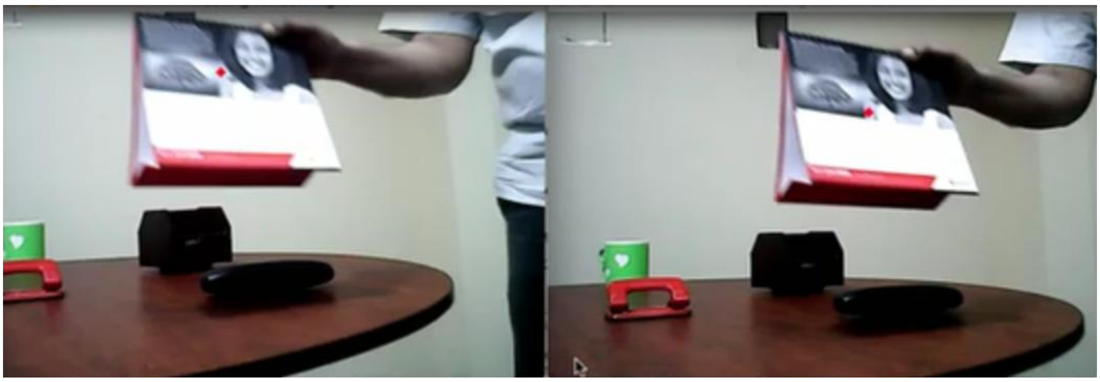
Figure 5.5 (b): Results obtained for focusing mode to the calendar in domestic environment ($T = 10s$)

Left eye Right eye

Figure 5.5 (c): Results obtained for focusing mode to the calendar in domestic environment ($T = 20s$)

Left eye Right eye

Figure 5.5 (d): Results obtained for focusing mode to the calendar in domestic environment ($T = 30s$)



Left eye Right eye

Figure 5.5 (e): Results obtained for focusing mode to the calendar in domestic environment ($T = 40s$)

Left eye Right eye

Figure 5.5 (f): Results obtained for focusing mode to the calendar in domestic environment ($T = 50s$)

Following results are presented the experiment performed of the robot eye using the environmental view to object focusing to the calendar in domestic environment as shown from Figure 5.6(a) to 5.6(f).

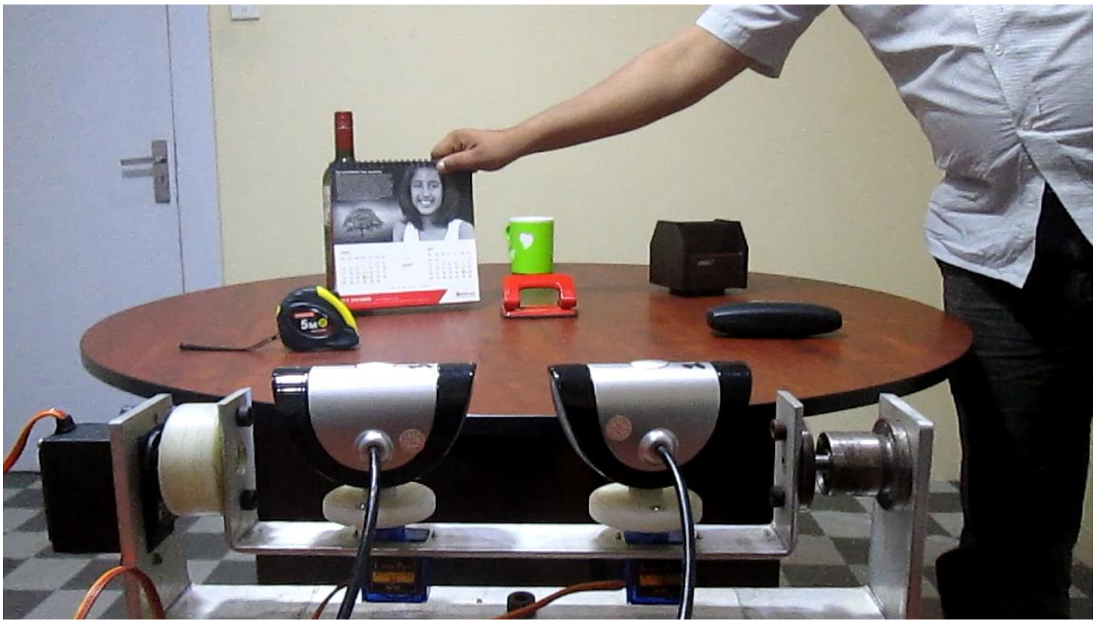


Figure 5.6 (a): Results obtained for focusing mode to the calendar in environmental view of domestic environment ($T = 0s$)

Figure 5.6 (b): Results obtained for focusing mode to the calendar in environmental view of domestic environment ($T = 10s$)

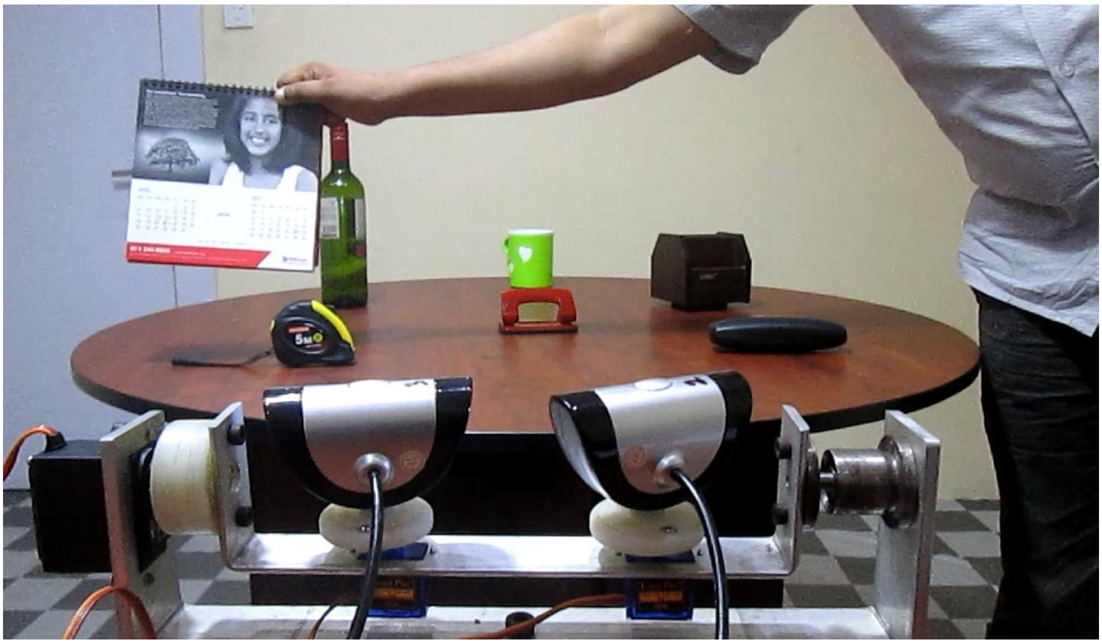


Figure 5.6 (c): Results obtained for focusing mode to the calendar in environmental view of domestic environment ($T = 20s$)

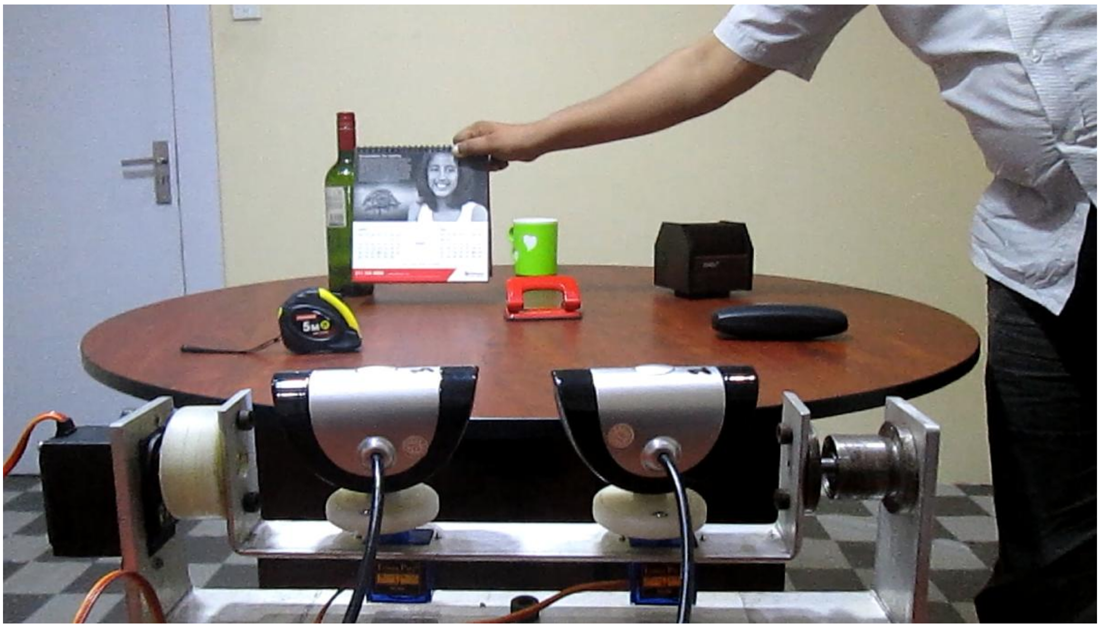


Figure 5.6 (d): Results obtained for focusing mode to the calendar in environmental view of domestic environment ($T = 30s$)

Figure 5.6 (e): Results obtained for focusing mode to the calendar in environmental view of domestic environment ($T = 40s$)

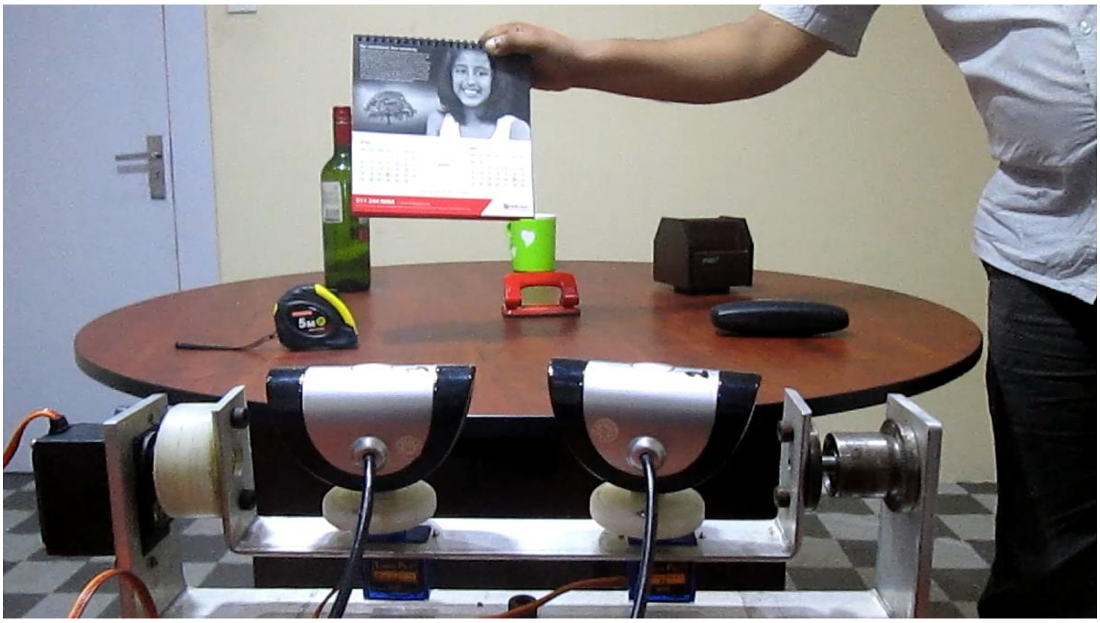


Figure 5.6 (f): Results obtained for focusing mode to the calendar in environmental view of domestic environment ($T = 50s$)

Figure 5.7 is presented the experiment performed of the robot eye joint angle movements for duration of $50s$.

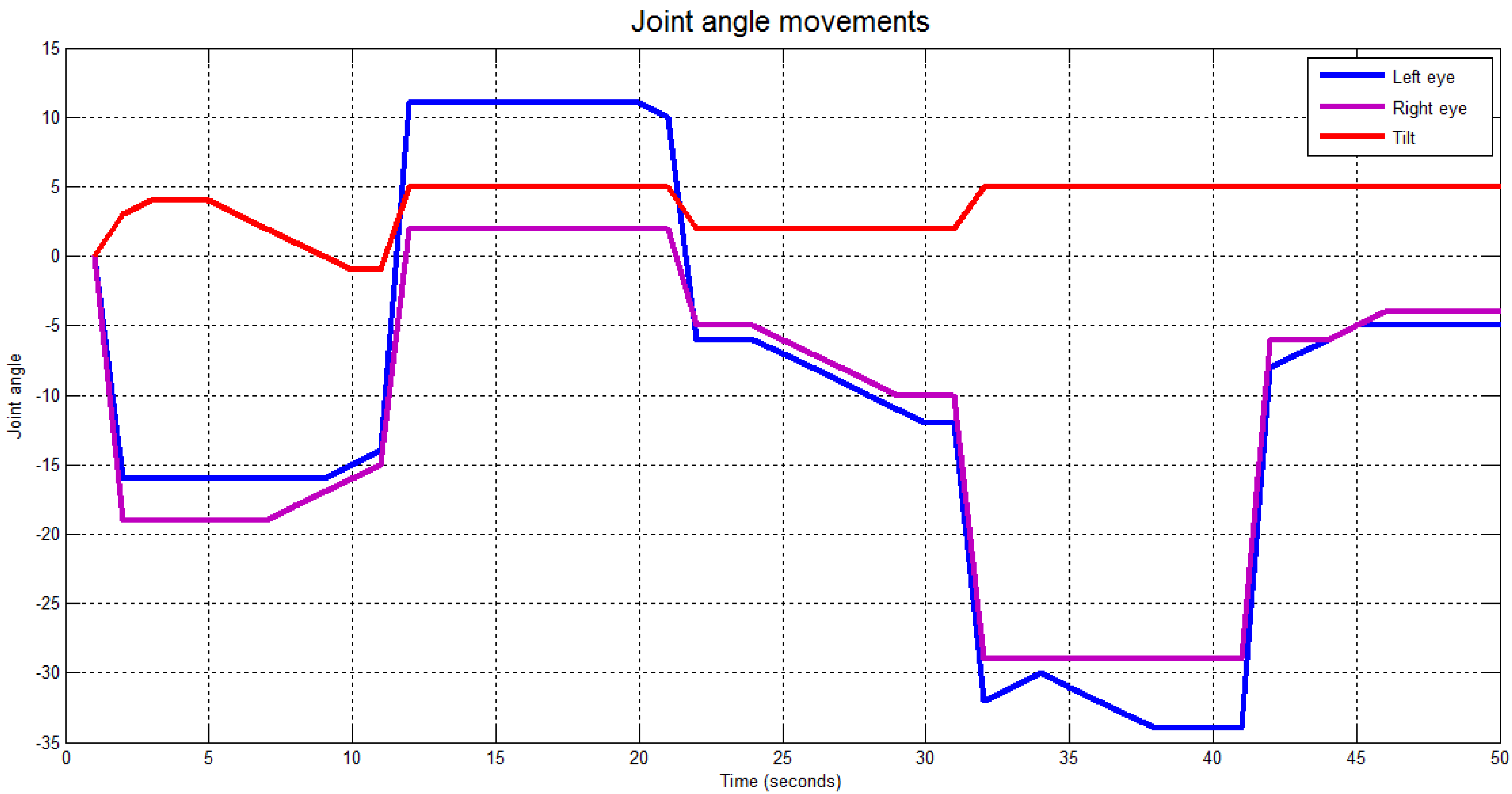


Figure 5.7: Joint angle movements of focusing to calendar

5.2. Results of the object searching mode

Results are presented of the experiment performed using the object searching in domestic environment as shown from figure 5.8. The user given command to search object a feedback whether the movement is **“near”**, **“middle”**, and **“far”**. Results are presented of the experiment performed using the objects searching in domestic environment as shown in below figures.

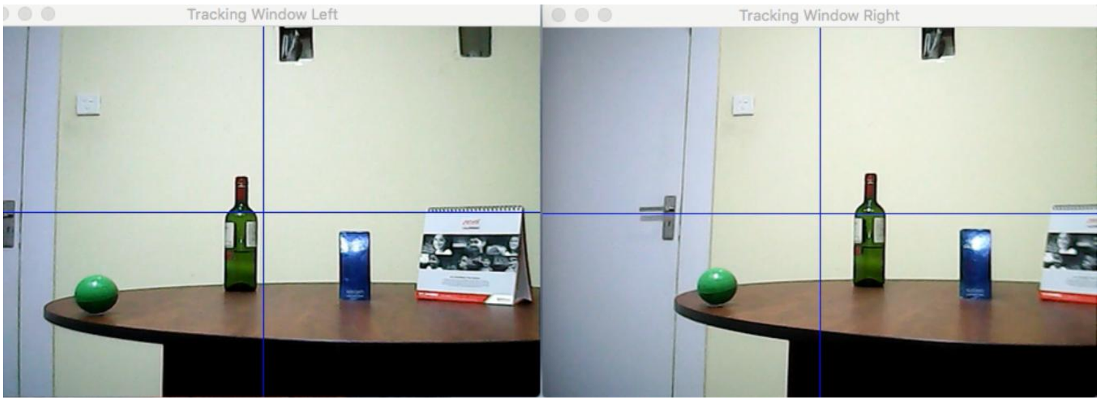
Figure 5.8: Object searching environment in domestic scenario

5.2.1. Case I: Results of the object searching mode (initially focus to ball)

Following results are presented the experiment performed of the robot eye using the web camera view to search object a feedback whether the movement is **“near”**, **“middle”**, and **“far”** in domestic environment as shown from Figure 5.9(a) to 5.9(d).

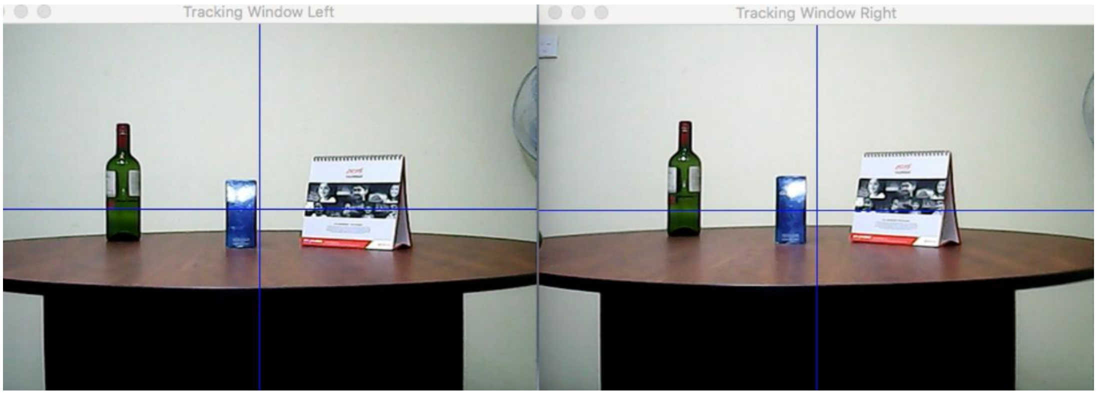
Left eye Right eye

Figure 5.9 (a): Results obtained for searching mode web camera view in domestic environment (initially focus to ball)



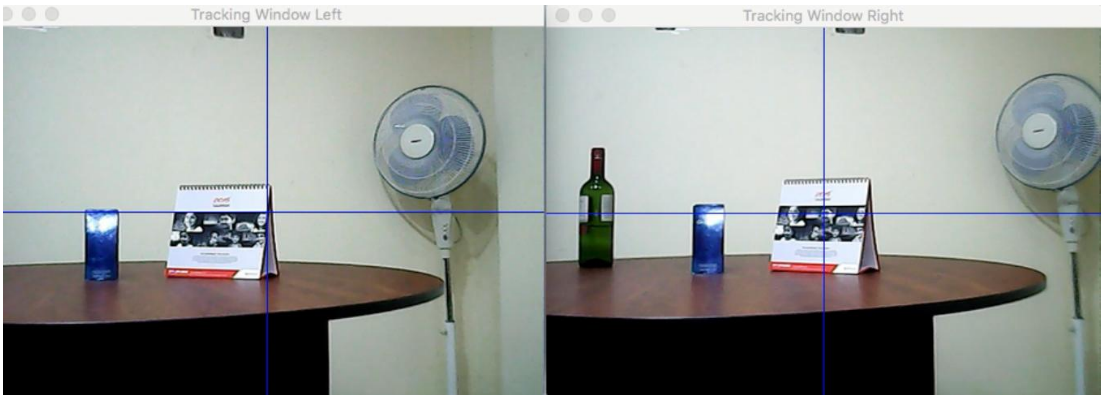
Left eye Right eye

Figure 5.9 (b): Results obtained for searching mode web camera view in domestic environment (searching movement – **“near”**)



Left eye Right eye

Figure 5.9 (c): Results obtained for searching mode web camera view in domestic environment (searching movement – **“middle”**)



Left eye Right eye

Figure 5.9 (d): Results obtained for searching mode web camera view in domestic environment (searching movement – **“far”**)

Following results are presented the experiment performed of the robot eye using the environmental view to search object a feedback whether the movement is **“near”**, **“middle”**, and **“far”** in domestic environment as shown from Figure 5.10(a) to 5.10(d).

Figure 5.10 (a): Results obtained for searching mode environmental view in domestic environment (initially focus to ball)

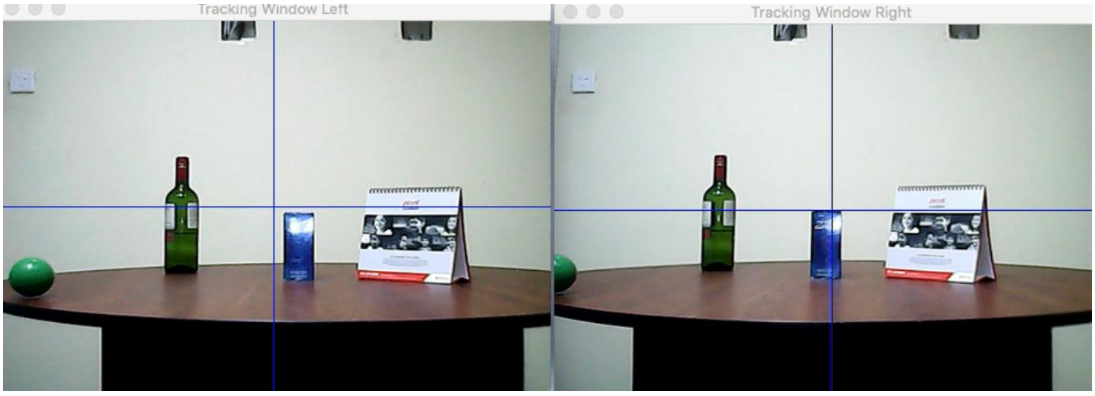
Figure 5.10 (b): Results obtained for searching mode environmental view in domestic environment (searching movement – **“near”**)

Figure 5.10 (c): Results obtained for searching mode environmental view in domestic environment (searching movement – **“middle”**)

Figure 5.10 (d): Results obtained for searching mode environmental view in domestic environment (searching movement – **“far”**)

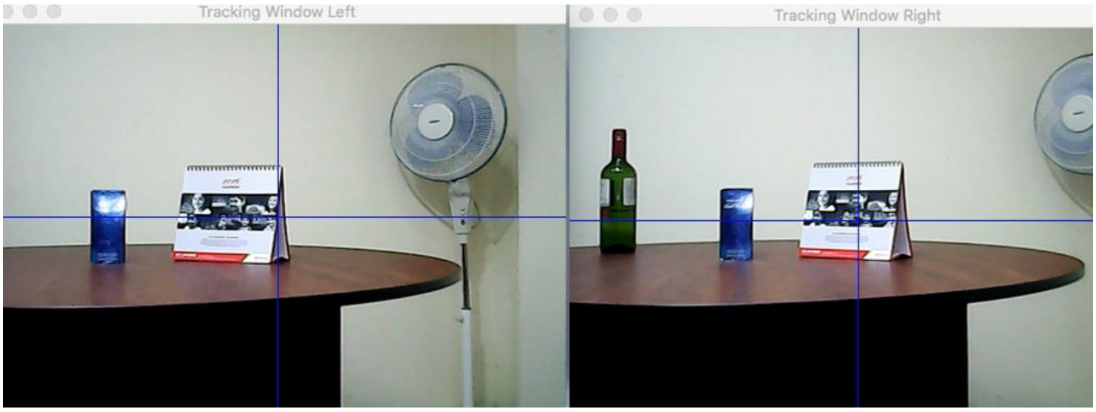
5.2.2. Case II: Results of the object searching mode (initially focus to blue box)

Following results are presented the experiment performed of the robot eye using the web camera view to search object a feedback whether the movement is **“near”**, **“middle”**, and **“far”** in domestic environment as shown from Figure 5.11(a) to 5.11(d).



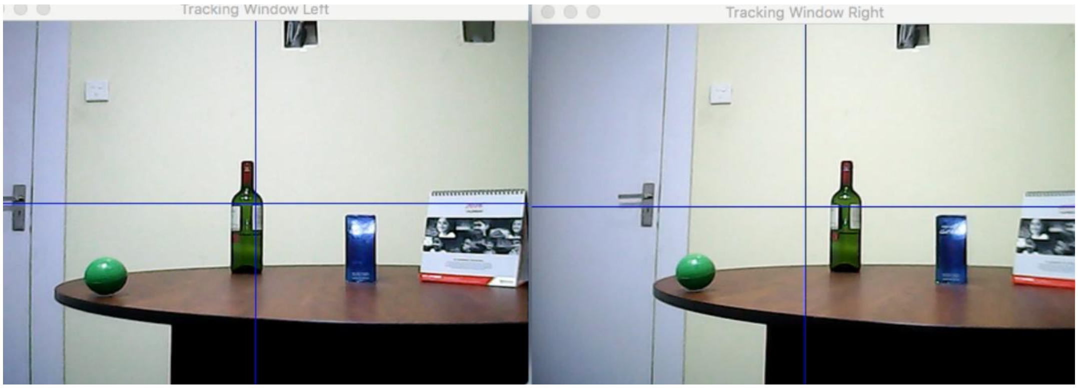
Left eye Right eye

Figure 5.11 (a): Results obtained for searching mode web camera view in domestic environment (initially focus to blue box)



Left eye Right eye

Figure 5.11 (b): Results obtained for searching mode web camera view in domestic environment (searching movement – **“near”**)



Left eye Right eye

Figure 5.11 (c): Results obtained for searching mode web camera view in domestic environment (searching movement – **"middle"**)

Left eye Right eye

Figure 5.11 (d): Results obtained for searching mode web camera view in domestic environment (searching movement – **"far"**)

Following results are presented the experiment performed of the robot eye using the environmental view to search object a feedback whether the movement is **"near"**, **"middle"**, and **"far"** in domestic environment as shown from Figure 5.12(a) to 5.12(d).

Figure 5.12 (a): Results obtained for searching mode environmental view in domestic environment (initially focus to blue box)

Figure 5.12 (b): Results obtained for searching mode environmental view in domestic environment (searching movement – **“near”**)

Figure 5.12 (c): Results obtained for searching mode environmental view in domestic environment (searching movement – **“middle”**)

Figure 5.12 (d): Results obtained for searching mode environmental view in domestic environment (searching movement – **“far”**)

5.3. Relevant practical scenarios in the object manipulation of service robot

Robot assistance in activities of daily living, especially for the elderly, should be executed in an intuitive and natural way [28]. This requires a robot can proactively understand the user's intention and automatically provide the desired service. Moreover the robot should be able to predict human intention and then choose the most reasonable actions to cooperate with humans. So satisfied assistance is based on accurate and real-time intention recognition technologies [29]. For example, a smart wheelchair could detect human's sight paths to adjust its direction and speed in a more cooperating and a less frustrating way. Surveillance robots have memory of the past sub-plans to help them recognize intention. Recognizing system in a fighter distinguishes the most dangerous hypotheses in a very short time to assist the pilot's manipulation [30]. Intention recognition generally relies on action recognition, voice recognition, biological motion recognition, plan recognition, and activity or behavior recognition. These features could be extracted by using various image processing, gesture detecting, and motion calculating technology and object affordance. The affordance is a relation of action/activity/intention and a specific object used to predict the next action/activity [31, 32].

Video and audio-driven human perception has attracted lots of interests in many research fields such as surveillance, security, and abdominal detection and many related applications are emerging in our daily lives [33]. In addition, many researches have dealt with camera sensor network for driving advantages in detecting many humans since multiple cameras schemes have defmite advantages for the detection of many humans with their enlarged sensing ranges and breaking through losing targets caused occlusion [34]. However, most of the previous studies on this topic mainly focus on the single modality such as multiple human tracking, face identification, or abnormal detection, respectively, and it has limitation brought to share meaningful human information between devices. In the daily environment, to generate more meaningful information that can be of substantive support to intonation consumer, intelligent devices require both capabilities of reliable human recognition with multimodalities and data association with each other among devices [35, 36].

An intelligent service robot is a machine that is able to perceive the environment and use its knowledge to operate safely in a meaningful and purposive manner. Intelligent service robots are being developed as a solution for the widening gap between supply and demand of human caregivers for elderly/disabled people [37]. These service robots are intended to be operated by non-expert users in human populated environments. Therefore, human friendly interactive features are preferred for domestic service robots.

Verbal communication is one of the mostly used communication modalities by the humans in order to communicate with companions. Therefore, human like verbal communication abilities are favored for domestic service robots with human friendly interactive features. The natural verbal communications phrases and utterance often include uncertain terms such as “little”, “far”, “large” and “high” related to spatial information, size/length of an item, usual processes etc. [38]. These uncertain terms are sometimes referred as fuzzy linguistic information, qualitative terms and fuzzy predicates. The quantitative meaning of uncertain information depends on various factors such as environment, context and experience. As an example, the quantitative meaning of distance related uncertain term, “far” when discussing about an arrangement of objects on a table is relatively very smaller than the “far” when discussing about locations of different cities [39]. Therefore, service robot must possess human like cognitive ability in understanding uncertain information in order to provide better interaction and service for the human users. Experiments have been carried out in an artificially created domestic environment. According to the experimental results, the behavior of the motion intention switcher is effective in identifying the actual intention of the user and switching the intention of the robot. Furthermore, the proposed method comforts the user since the proposed method is beneficial in reducing the number of user instructions/steps required in order to navigate a robot to a goal position compared to the existing approaches. Moreover, this performance improvement ascertains the applicability of the proposed concept for enhancing the human friendless of service robot [40].

In this context, service robots designed specially for domestic usage can serve with a higher integrity. Domestic service robots have come under close scrutiny among researchers. When collaborating with humans, robots should be able to clearly understand the instructions conveyed [41] by the human users. Voice interfaces are frequently used as a mean of interaction interface between users and robots as it requires minimum amount of work overhead from the users. However, the information conveyed through the voice instructions are often ambiguous and cumbersome due to the inclusion of imprecise information [42]. The voice instructions are often accompanied with gestures especially when refereeing objects, locations, directions etc. in the environment. However, the information conveyed solely through these gestures is also imprecise. Therefore, it is more effective to consider a multimodal interface rather than a unimodal interface in order to understand the user instructions [43].

Moreover, the information conveyed through the gestures can be used in order to improve the understanding of the user instructions related to object placements. This paper proposes a method to enhance the interpretation of user instructions related to the object placements by interpreting the information conveyed through voice and gestures. Furthermore, the proposed system is capable of adapting the understanding according to the spatial arrangement of the workspace of the robot [44]. Fuzzy logic system is proposed in order to evaluate the information conveyed through these two modalities while considering the arrangement of the workspace. Experiments have been carried out in order to evaluate the performance of the proposed system. The experimental results validate the performance gain of the proposed multimodal system over the unimodal systems.

In domestic environments it is necessary for the robot to be able to understand the spatial terms which are associated with object manipulation. Most common requirement would be to pick or place items on a surface like a table. Specially when referring to an area on a table there are certain set of verbal terms that are associated [45]. These terms usually doesn't have strict boundaries. Terms like "Left", "Right" and "Center" can be pointed out as examples. Even though in the system identifies the objects which are there in a certain area on a surface, it lacks the capability to pin point a location when placing an item. It will be tedious task if the user has to recursively segment the areas into small portions in-order to get a higher accuracy [46].

CHAPTER 6: DISCUSSION AND CONCLUSION

6.1. Problems encountered and action taken

When the environmental light condition is poor the web camera automatically increases the exposure level by increasing the shutter open time. Shutter open time has inverse relationship to the frame rate of the camera. Therefore, when the light condition gets poor, the frame rate of the camera decreases. Low frame rate slows down the vision feedback and reduces the performance of the system.

This will lead to undesirable problems such as difficulty of detect moving objects. In order to maintain the environmental light condition at a sufficient level to avoid dropping of the frame rate. Therefore environmental light condition was maintained at the desired level by using artificial lighting which also helps to improve the light condition. Sometimes due to background objects, noises are detected in the image processing and this will lead to falls triggering of the system.

When the objects are in a dark color background the camera's exposure level will be increased and due to that the lighter color objects are seen as more whitish. In such occasions object detection was difficult. The use of a white artificial background provides a remedy for this problem also.

Replacing web cameras with high frame rated cameras and using a PC with high performance are recommended for the future developments to have higher performance. The motors and mounting hubs are products from the same manufacturer and they are designed to be compatible with each other, it was able to identify some backlashes while connecting the mounting hubs with motors.

Due to those backlashes, smoothness of movements got disturbed giving rise to vibrations. Backlashes associated with mounting hubs were eliminated by adding some fittings to the hub connecting points but inherent backlashes of gearboxes couldn't be avoided. The only available remedy that can be proposed to avoid this problem is replacing the motors and gear boxes with higher quality products even though they may be very expensive than the motors that are currently used.

Some challenges associated with the mechanical eye design included incorporating servos. Those servos are not symmetrical, which meant that in order to incorporate those in a design, most surrounding components could not be symmetrical either. Interestingly, as this adds complexity to the mechanical components, it also adds cost to manufacture them, meaning that incorporating servo motors instead of symmetrical but more expensive components could prove to actually indirectly be the more expensive option.

In an effort to standardize parts of the design, the original design of all layers of the eye robot was the same, meaning that each layer had the same shape. This is an improvement because it costs less to make multiples of a single part than to make multiple different parts.

The tracking of the various objects which have different shapes, sizes and colors is tested and implemented successfully. From this test scenarios can conclude that the CAMShift algorithm is superior to other algorithms in order to track and search the position of the respective objects.

6.2. Conclusion

In this project a methodical approach has been followed to the design and develop an interactive robotic eye for adapting robot attention to user command request about the distance of an object based on the visual attention of the robot. In a human robot interaction, the humans may use command request, which focus or search object a feedback whether the movement is “near”, “middle”, and “far”. The actual quantitative meaning of those terms depends on spatial arrangement of the domestic environment where the attention is focused on. Therefore, spatial information of the environment is analyzed to adapt robot’s perception about the distance of an object, which is in its vision field. The concept has been tested for different test cases by changing the visual attention of the robot eye. A research has been carried out to identify the natural human perceptions about the vision field through the domestic environments in those testing scenarios. The process includes the mechanical and electrical designs, in the design process close attention has been paid to the human bio-mechanics to realize a design that reaches anthropomorphism to a closer degree.

The proposed method is capable to mimics key visual functions of the human brain promises to robot eye maneuver quickly and safely through adaptive vision field through the domestic environments. The motion of changing adaptive vision field which used a focus or search objects in domestic environment more human-like manner using depth map analysis. The tracking and searching of the various objects which have different shapes, sizes and colors is tested and implemented successfully. From this test scenarios can conclude that the CAMShift algorithm is superior to other algorithms in order to track and search the position of the respective objects. Also the proposed robotic eye is designed in such a way that it can be used as a platform for facilitating further developments in integrating more interactive features to robotic eye.

6.3. Further developments

The robot could be significantly improved with redesign. After trials with robot eye for service robot to determine whether it is feasible to use the robot for its intended purpose, the overall design, materials, and manufacturing techniques should be evaluated and optimized for the quantity of robots that are expected to be made.

The outward appearance of the robot eye needs to be improved upon. The eye is too big and somewhat unequal, and its appearance may keep it from being accepted by more human-like. The eye robot could become more interactive if it had more sensors in it. Sensors of some sort would allow it to sense physical interaction.

Adding of a proper outer cover can be used to enhance the appearance of the interactive features. It is better to design the outer cover using molding techniques with low weight materials. And also there are constraints in the head design due to movements of the links and actuators which should be considered before modeling the outer cover.

The robotic eye can be used as a research platform for implementing various control strategies such as artificial intelligent systems and machine learning techniques. These can be implemented in further developments.

REFRENCES

[1] John J. Craig, Introduction to Robotics: Mechanics and Control: Pearson Education, Incorporated, 2005.

[2] T. Fong, I. Nourbakhsh, and K. Dautenhahn, "A Survey of Socially Interactive Robots: Concepts, Design, and Applications," The Robotics Institute, Carnegie Mellon University, Pittsburgh, Technical Report CMU-RI-TR-02-29, 2002.

[3] F.L. Lewis, D.M. Dawson, and C.T. Abdallah, Robot Manipulator Control: Theory and Practice, 2nd ed.: CRS Press, 2003.

[4] R.Beira, M.Lopes, M.Prac, Jos´e Santos-Victor, Alexandre Bernardino, "Design of the Robot-Cub (iCub) Head," in IEEE International Conference on Robotics and Automation, Florida, 2006, pp. 94-100.

[5] V. Ahlstrom and K. Longo, "Human factors design standard," Federal Aviation Administration William J. Hughes Technical Center, 2003.

[6] M. Ramachandran, A. Veeraraghavan, and R. Chellappa, "A fast bilinear structure from motion algorithm using a video sequence and inertial sensors," IEEE Trans. Pattern Anal. Mach. Intell., vol. 33, no. 1, pp. 186–193, Jan. 2011.

[7] K. Kral, "Behavioral-analytical studies of the role of head movements in depth perception in insects, birds and mammals," Behav. Process, vol. 64, pp. 1–12, 2003.

[8] J. Rajruangrabin and D. O. Popa, "Robot head motion control with an emphasis on realism of neck–eye coordination during object tracking," Journal of Intelligent & Robotic Systems, vol. 63, no. 2, pp. 163–190, 2011.

[9] M. Antonelli, A. P. del Pobil, and M. Rucci, "Bayesian multimodal integration in a robot replicating human head and eye movements," in Proc. IEEE

International Conference on Robotics and Automation (ICRA'14), 2014, pp. 2868–2873.

[10] X. Kuang, M. Gibson, B. E. Shi, and M. Rucci, "Active vision during coordinated head/eye movements in a humanoid robot," Robotics, IEEE Trans. on, vol. PP, no. 99, pp. 1 –8, 2012.

[11] R. Diankov and J. Kuffner, "Openrave: A planning architecture for autonomous robotics," Robotics Institute, Pittsburgh, PA, Tech. Rep. CMU-RI-TR-08-34, 2008.

[12] Vega J., Perdices E., Canas M., "Robot Evolutionary Localization Based on Attentive Visual Short-Term Memory," *Article on www.mdpi.com/journal/sensors,* pp.1268-1299, 2013.

[13] Shi Y., Liang B., Wang X., Xu W., Liu H., "Study on Intelligent Visual Servoing of Space Robot for Cooperative Target Capturing," *IEEE International Conference on Information and Automation,* pp.733-738, 2012.

[14] Bekele E., Lahiri U., Davidson J., Warren Z., Sarkar N., "Development of a Novel Robot-Mediated Adaptive Response System for Joint Attention Task for Children with Autism," *IEEE International Symposium on Robot and Human Interactive Communication,* pp.276-281, 2011.

[15] Suzuki S., Kurunoma K., HarashimaF., "Development of gaze detection system for multiple users-toward realization of an intelligent service interface," *ICROS-SICE International Joint Conference,* pp.3245-3250, 2009.

[16] Wang H., Liu Y., Wang Z., "Adaptive Visual Servoing Using Angle and Distance," *Proceeding of the 7th World Congress on Intelligent Control and Automation,* pp.479-484, 2008.

[17] Park H., Nho M., Kim H., Hyun P., “An eye-in-Hand Device for Hand Gesture-Based Human-Robot Interaction”, *9th International conference on Ubiquitous Robots and ambient Intelligence (URAI),* pp.106-107, 2012.

[18] Charles R., Brett P., Holroyd A., Sidner L., “Recognizing Engagement in Human-Robot Interaction”, *Human-Robot Interaction, Osaka,* 2009.

[19] Jen C., Hsia C., Chang H., Chang W., Hsu H., Che Y., Li C., Husan M., “An Efficient Object Recognition and self-Localization System for Humanoid Soccer Robot”, *SICE Annual Conference,* pp.2269-2278, 2010.

[20] Wang Y., Lang H., Clarence W., “An Autonomous Mobile Grasping System using Visual Servoing and Nonlinear Model Predictive Control”, *8th IEEE International Conference on Control and Automation,* pp.86-91, 2010.

[21] Saldien J., Goris K., Selma Y., Werner V., Dirk L., “On the Design of the Huggable Robot Probo”, *Journal of Physical Agent,* pp.3-11, Vol.2, 2008.

[22] Yang W., Zhu K., Ruan Q., Han D., “Moving Target Tracking and Measurement with a Binocular Vision System”, *15th International conference on Mechatronics and Machine Vision in Practice,* pp.85-91, 2008.

[23] Itseez. (2016, April) OpenCV (Open Source Computer Vision). [Online]. http://opencv.org/

[24] (2016, April) ARDUINO. [Online]. http://arduino.cc/en/main/software

[25] C. Zhang, Y. Qiao, E. Fallon, and C. Xu. An improved camshift algorithm for target tracking in video surveillance. In Proceedings of 9th. IT and T Conference, number 12, 2009.

[26] G. Xiang and X.Wang. Real-time follow-up head tracking in dynamic complex environments. Journal of Shanghai Jiaotong University (Science), 14(5):593–599, 2009.

[27] F. Dadgostar, A. Sarrafzadeh, and S. P. Overmyer. Face tracking using mean-shift algorithm: A fuzzy approach for boundary detection. In J. Tao, T. Tan, and R. W. Picard, editors, ACII, volume 3784 of Lecture Notes in Computer Science, pages 56–63. Springer, 2005.

[28] Loconsole, Claudio, et al, "A new Kinect-based guidance mode for upper limb robot-aided neuro rehabilitation," Intelligent Robots and Systems (IROS), IEEE/RSJ International Conference on, pp. 1037-1042, 2012.

[29] Zhou, Shandan, et al, "A context-aware reminder system for elders based on fuzzy linguistic approach," Expert Systems with Applications vol. 39, no. 10, pp. 9411-9419, 2012.

[30] Krauthausen, Peter, Learning Dynamic Systems for Intention Recognition in Human-Robot-Cooperation. KIT Scientific Publishing, 2012.

[31] Tahboub, Karim A, "Intelligent human-machine interaction based on dynamic bayesian networks probabilistic intention recognition," Journal of Intelligent and Robotic Systems, vol. 45, no. 1, pp. 31-52, 2006.

[32] Jacob, Mithun G, and Juan P. "Context-based hand gesture recognition for the operating room," Pattern Recognition Letters, vol. 36, pp. 196-203, 2014.

[33] Koppula, Hema S, and Ashutosh S, "Anticipating human activities using object affordances for reactive robotic response," Robotics: Science and Systems, Berlin, 2013.

[34] Sun, Y, Shaogang R, and Yun L, "Object–object interaction affordance learning," Robotics and Autonomous Systems, 2013.

[35] Q. Nguyen and J. Choi, "Multiple Sound Sources Localization with Perception Sensor Network," in Proc. of IntI. Symp. on Robot and Human Interactive Communication (RO-MAN), pp. 418-423, 2013.

[36] O. Popoola and K. Wang, "Video-Based Abnormal Human Behavior Recognition-A Review, " IEEE Trans. Syst., Man, Cybern., Part C:Appl. Rev., Vol. 42, No. 6, pp. 865-878, 2012.

[37] A. G. B. P. Jayasekara, K. Watanabe, K. Kiguchi, and K. Izumi, "Interpretation of fuzzy voice commands for robots based on vocal cues guided by user's willingness," in 2010 IEEE/RSJ Int. Conf. Intelligent Robots and Systems, 2010, pp. 778–783.

[38] A. G. B. P. Jayasekara, K. Watanabe, and K. Izumi, "Understanding user commands by evaluating fuzzy linguistic information based on visual attention," Artificial Life and Robotics, vol. 14, no. 1, pp. 48–52, 2009.

[39] M. A. V. J. Muthugala and A. G. B. P. Jayasekara, "Interpretation of uncertain information in mobile service robots by analyzing surrounding spatial arrangement based on occupied density variation," in 2016 IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS). IEEE, 2016, pp. 1517–1523.

[40] M. A. V. J. Muthugala and A. G. B. P. Jayasekara, "Mirob: An intelligent service robot that learns from interactive discussions while handling uncertain information in user instructions," in 2016 Moratuwa Engineering Research Conf. (MERCon). IEEE, 2016, pp. 397–402.

[41] J. Forlizzi and C. DiSalvo, "Service robots in the domestic environment: a study of the roomba vacuum in the home," in Proceedings of the 1st ACM

SIGCHI/SIGART conference on Human-robot interaction. ACM, 2006, pp. 258–265.

[42] H. Osawa, J. Orszulak, K. M. Godfrey, M. Imai, and J. F. Coughlin, “Improving voice interaction for older people using an attachable gesture robot,” in 19th International Symposium in Robot and Human Interactive Communication. IEEE, 2010, pp. 179–184.

[43] P. Tsarouchi, S. Makris, and G. Chryssolouris, “Human-robot interaction review and challenges on task planning and programming,” International Journal of Computer Integrated Manufacturing, vol. 29, no. 8, pp. 916–931, 2016.

[44] G. Gemignani, M. Veloso, and D. Nardi, “Language-based sensing descriptors for robot object grounding,” in Robot Soccer World Cup. Springer, 2015, pp. 3–15.

[45] M. A. V. J. Muthugala and A. G. B. P. Jayasekara, “Enhancing human robot interaction by interpreting uncertain information in navigational commands based on experience and environment,” in 2016 IEEE Int. Conf. Robotics and Automation (ICRA), 2016, pp. 2915–2921.

[46] P. H. D. A. S. Srimal and A. G. B. P. Jayasekara, “A multi-modal approach for enhancing object placement,” in National Conference on Technology Management. IEEE, 2017.