BIDSO

BIOLOGICAL INSPIRED DETERMINING OF SALIENT OBJECTS

By

BUSHRA SABIR

RABIYA RASHID

IRFAN MAROOF

ALI KHURSHID

Submitted to the Faculty of Computer Science, Military College of Signals,

National University of Sciences and Technology, Rawalpindi,

in partial fulfillment for the requirement of a

B.E. Degree in Computer Software Engineering.

August 2010

ABSTRACT

Robotic vision is the main topic of research nowadays, in order to reduce the computation time while making visual decisions, it’s very important to visualize a scene according to human vision. Human visual system focus on certain regions/objects of particular interest while ignoring the surroundings, our project extracts those regions of interest according to visual senses/behavior of human being.

The purpose of this project is to design and develop a software system that takes dynamic and static scenes as an input and extract the salient regions based on human visual attention system. This project presents a model of visual attention that can be used as a building block for a real-time robotic vision system. Visual attention helps us decide what's important, and understanding how this works in humans may allow us to build smarter computer vision systems that know what to look for.

Region based approach is used to reduce the number of units that have to be processed by the attention systems. Top-down, bottom-up techniques and faster algorithms that are used that will help in increasing performance and efficiency of processing to integrate the attention procedure into a comprehensive biologically inspired vision system. Five features namely color; size, orientation, eccentricity and symmetry are used.

ACKNOWLEDGEMENTS

There is no success without the will of ALLAH. We are grateful to ALLAH, who has given us guidance, strength and enabled us to accomplish this task. Whatever we have achieved, we owe it to Him, in totality.

We are also grateful to our parents and family and well-wishers for their admirable support.

We would like to thank our supervisor Dr Naveed Rao and our co-supervisors Dr Irtiza Ali and Dr Zaheer Aziz for their help and motivation throughout the course of our project. Without his help we would have not been able to accomplish anything. They provided us with the opportunity to polish our technical skills and guided us in this area of learning.

Our project group would also like to acknowledge the administration of CS Dept, Military College of Signals for their support and assistance. We are also extremely thankful to all the instructors who imparted knowledge on us during the course of our engineering studies at Military College of Signals, NUST.

Contents

CHAPTER 1

INTRODUCTION

1.1 Introduction

As technologies has advanced a lot in past decades, it will be a matter of time when robots will act humanly on visual input and extract the salient objects based on biological visual attention. Visual attention helps us decide what's important, and understanding how this works in humans may allow us to build smarter computer vision systems that know what to look for. The algorithms which were previously used were based on spatial and frequency domain filters which were not efficient, they lagged in performance and used only a single technique either top-down or bottom-up. But our system will use region based approach to reduce the number of units that have to be processed by the attention systems. Our system will use top-down, bottom-up techniques and faster algorithms that will help in increasing performance and efficiency of processing to integrate the attention procedure into a comprehensive biologically inspired vision system. This project will be done with the collaboration of Paderborn University, Germany.

1.2 Scope And Objective

The purpose of this project is

1. This system used the region growing technique for segmentation as mentioned in paper.
2. Our project will determine a salient region with respect to five features namely color, size, eccentricity, orientation, and symmetry. It will use the global and local voting mechanism for finding salient region in the input.

3 Our system will integrate both top-down and bottom-up techniques for finding the focus of attention.

4 As computation time is the main issue, in segmentation early clustering of pixels must be done.

5 Motion detection has to be added. If there is a movement in the scene it must detect it.

6 Our project will keep human visual perception into consideration and hence, the clusters obtained by this technique is biological plausible

7 One of milestone is to use fast and robust algorithms for extracting features involved in visual attention from segmented region.

8 Methods for inhibition of return of regions must be used. It must have a capacity of around 15 regions.

9 Our software system will act as a module of visual attention and focusing of mobile objects like robots.

10 A simulation must be developed in gazebo(a simulation tool for robots) to show the validation of our system.

1.3 Goals

The main goals that we want to achieve from this project are

1. To integrate top-down and bottom-up approaches of visual attention in a single unit.
2. To perform efficient segmentation using new method of segmentation mentioned in paper [CSFRAM].

3. To used regions rather than individual pixels for computation of saliency in accordance to [EVAR] and [HSI].
4. To integrate five features of visual attention i.e. size, color, eccentricity, orientation and symmetry to compute saliency.
5. To make the saliency computation computationally less expensive to achieve better performance.

1.4 Problem

The main problem address is vision in robot is not like human vision, Camera act as eyes of robot. How robots will know what to focus and what not? Given an image, the robot will focus on the whole image that can:

1. Increases the computational effort
2. Is an inefficient way
3. Expensive and time-consuming.

1.5 Solution

As humans only focus on the salient regions/objects and surroundings seem blurred unless focused as in foveated image. So the solution to this problem is Robots should also focus only on salient regions in an image. This will

1. Increases efficiency
2. Decrease Computational effort
3. Less expensive and Faster

1.6 Applications

Our technique computes saliency of important regions according to the human visual attention so it can be used for;

1. Compression based on region of interest coding
2. For analysis of Ads/Banners in newspapers and websites
3. Intelligent Cameras for salient region location.
4. Web Image search engine to search for object given a sample image

1.7 Potential Difficulties and Limitations

1. Only monocular images are taken as input and stereo images are not considered this means that attention model is not concerned with depth factor.
2. Although the algorithms followed are computationally better as seen by comparing different other algorithms during literature review, but still it is not very fast.
3. Only RGB images are taken for processing.
4. Only extraction of the most salient object is in scope of our project, embedding it in a real robot and artificial techniques needed for robot to act/react according to visual attention is not in the scope of this project.
5. .GIF types of images are not supported by our system.

1.8 Intended Audience

This document is intended for

1. **Project Supervisor**: In order to be sure they are development of the project fulfills requirements provided in this document.
2. **Project Team:** In order to be sure they are developing the right project that fulfills requirements provided in this document.
3. **Project Panel:** In order to analyze and evaluate the progress of the project.

4. **Users**: In order to get familiar with the idea of the project and suggest other features that would make it even more functional.
5. **System administrators**: In order to know exactly what they have to expect from the system, right inputs and outputs and response in error situations.

1.9 Organization of Report

Chapter1: Introduction of the project and an overall functionality of the project.

Chapter2: Describes the literature review done for this project.

Chapter3: Describes the requirement specifications of this project.

Chapter4: Describes the overall design and architecture of this project

Chapter5: Describes the system development with all the details of the system functions and explains the way they have been implemented.

Chapter6: Presents the results of the model used for computing saliency.

Chapter6: Presents the results and analyze it with other techniques and methodologies available in market.

Chapter7: Presents the simulation developed of this system

Chapter8: Presents the conclusion and future work.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

We have gone through different papers and books to understand the problem completely and come up with a better solution. At first a brief overview of the color models is given and the requirement of color model for our system. Then available segmentation algorithms are discussed. In the end a little bit of introduction is given to bottom-up and top-down approaches.

2.2 Human Visual System

At first a brief overview of the human vision system is done that how human vision works and how the image is processed in human brain. It was concluded that the attention mechanism applies a filtration process in the visual input that selects only relevant and important portions from the viewed scene for high level analysis for brain. Human vision focuses on certain important areas either of their concern or those regions that are highly prominent either because of their color, shape, eccentricity or other features etc. For this purpose a search paper [VAJMW] has been studied.

2.3 Models of Natural Vision

This section reviews the concepts and theories that have influenced the computation models of visual attention existing today. There are some theories, like the feature integration theory and guided search theory that have a large following in the community of computational modelling. The model that has been used also combines concepts from these two theories.

2.3.1 Feature Integration Theory

The feature integration theory was proposed by Treisman and Gelade [TG80] in 1980. It has been one of the most inuential psychological models of human visual attention. It suggests that the human vision system can detect and identify separable features in parallel across a display and this early, parallel process of feature registration mediates between texture segregation and figure-ground grouping. They further conclude from their experiments that locating any individual feature or performing their conjunctions requires attention to be diverted to each relevant location.

According to this theory several primary visual features are processed and represented with separate feature maps in an early step of visual processing. These maps are later integrated in a saliency map that can be accessed in order to direct attention to the most conspicuous areas. Efforts can be found in experimental psychology to identify the features that stimulate the visual attention mechanism. Some of identified features in this regard include color contrast [LPA95], orientation [Ner04], motion (or a sudden change) [Ita01], eccentricity [BFR84], and symmetry [OH03]. A detailed review of the features involved in the human attention can be seen in [WH04]. Models regarding combination of the feature channels in the pre-attention phase are proposed in [Koc99] and [Ner04]. The operation is modelled as square of sum in [Koc99] while [Ner04] proposes that the features are combined in the visual cortex using a multiplication-style operation.

A search task is categorized into two kinds according to this theory, namely, feature search and conjunction search. Feature search can be performed fast and pre-attentively for targets defined by primitive features. Conjunction search is the serial search for targets defined by a conjunction of primitive features. It is much slower and requires conscious attention. Color, orientation, and intensity are proposed as primitive features for which feature search can be performed.

2.3.2 Guided Search:

The guided search model was introduced by Wolfe and colleagues in 1989 [WCF89] and then its revised version was presented in [Wol94]. This model is related to the process of visual search in which the main objective is reduction in response time to identify presence of a search target in a given scene. The model suggests that a number of features like color and orientation are computed and stored in maps in each of which presence of these features is encoded. The top-down influence of the target features controls the construction of these maps to apply bias to a particular category of feature values. A weighted summation method is used to integrate these maps together into a combined activation map. Peaks in this activation map are visited serially as targets of attention. Locations once visited are marked in an inhibition map that is used to avoid rapid revisiting of the same locations. This model, especially its part of map construction, has been an inspiration for many computational models of attention as its processes are clearly explained making them feasible for conversion into algorithms.

2.3.3 Fine-Grain Top-Down Attention

A majority of the existing attention models have demonstrated visual search as a primary area of application for their models. Most of these models have utilized manipulation on bottom-up saliency maps in order to let the search target pop-out quickly. Although such processing demonstrates visual search as an attentional behavior but in practical sense these models lag behind the performance of natural visual search significantly. This suggests that the top-down tasks of attention have a different nature and require a separate mechanism for computing saliency. There is another school of thought about the top-down influences that leads to fine-grain nature of this attentional pathway in natural vision, which appeals better than the strategy mostly followed by the contemporary attention models.

The models of human vision such as [LD04] suggest target related feature processing in the V4 area of brain. Similarly the models on feature and conjunction search, for example [LHG97], also presume excitation and inhibitions on particular feature magnitudes rather than whole channels. Results of psychophysical experiments reported by [Ham05], [Dec05], and [NI06b] also support the concept of search on particular feature values rather than excitation on a whole feature channel. The work of [Ham05] has shown that a population of neurons encoding the target color and/or orientation gets a gain while others get suppressed. According to [Dec05], each feature channel can adopt many values that are evaluated by a specialized layer of neurons in the human brain.

Recent psychological models of attention such as [HT06] and [Knu07] agree on the concept that top-down modulations of neural responsiveness are precise for the features upon which attention is to be diverted. Apart from the excitation of the neurons concerned with the stimulus, it has been reported that neurons tuned for non-target stimulus parameters exhibit a decrease in sensivity [RD03]. The experiments reported by [NI06b] explicitly declare fine-grain nature of top-down attention. These findings suggest that the top-down saliency mechanism constructs task dependant maps to allow quick pop-out of the target rather than using the bottom-up saliency maps. The model proposed here follows this newly discovered strategy in its top-down pathway.

2.3.4 Inhibition and Facilitation of Return

After focusing one salient object/location, the next important component of attention that gets activated is inhibition of return (IOR). This process enables the vision system to fixate on a variety of locations (or objects) in the scene, otherwise the gaze would stay fixed to one salient location. It is worth mentioning here that there has been a continuing debate on early selection and late selection. In the early selection models, such as [Bro], [Tre60], and [TG67], attention is diverted to a location without forming a semantic meaning to the contents. Hence attention and IOR work only on location basis. On the other hand, according to late selection [DD63] the contents of sensory data are analyzed semantically before attending, therefore objects may be identified and used as units to perform attention processes. In the work presented in this dissertation the object based attention is taken into consideration.

It has been established by experiments in psychophysics that inhibition takes place in terms of both location and object features [GE94] [WLW98]. Evidence is provided for inhibition in the immediate vicinity of the attended location and a U-shaped function has been reported which strongly suppresses the immediate surroundings of the attended location and gradually fades to no suppression after a limited diameter [CT03]. The work of [LPA95] discovers the idea of feature based inhibition in which inhibition on color of the recently attended object has been reported in human vision. It was further confirmed by experiments reported in [PKK97] that inhibition takes place in terms of object identity apart from the spatial inhibition of return. The psychological model of attention proposed in [Knu07] defines an explicit role of a working memory while processing for bottom-up as well as top-down visual attention.

Under some visual behaviors, such as search and track, bias has to be given to certain features and/or locations so that the next fixations are driven towards similar looking features or nearby positions. This component of the attention mechanism is called facilitation of return (FOR) [OMY05] [CC06].

2.4 Computational Models of Attention

This section reviews the existing computational models of attention in a categorized arrangement. Categorization is made based upon the basic principle used for computing the focus of attention. Existing models could be classified into four types of approaches, namely, connectionist, saliency based, rarity based, and object based. The proposed model combines attributes of the last three types of approaches in its current status. This review will also provide a basis to understand the innovations suggested in the used attention model.

2.4.1 Saliency-Based Models

The saliency-based models have their foundations in the feature integration theory. A prominent model of this category was presented in [IKN98] and then refined in [IK00]. It builds saliency maps for three features, namely, color channels, intensity, and orientations. Each feature is computed by a set of linear center-surround operations between fine and coarse scales analogous to visual receptive fields. These feature maps are combined into three conspicuity maps for intensity, color and orientation through across-scale addition. At any given time, the maximum in the resultant of saliency maps defines the most significant image location to which the focus of attention should be directed. This is done by a 2D layer of leaky integrate-and-fire neurons. This layer feeds into a biologically plausible winner-take-all (WTA) neural network. Shift of attention to the winner location causes a global inhibition of all WTA neurons and transient activation of local inhibition.

Many flavours of the above model can be found with different variations in methodology. The model described in [PSL02] uses the opponent color theory for constructing the feature map of color contrast using a computation scheme very similar to [IKN98]. It introduces new feature maps for edges and symmetry. It computes two color maps and the center-surround is implemented as the difference between fine and coarse scales of a Gaussian pyramid images. A total of 24 maps are computed and combined into four conspicuity maps. Unsupervised learning is used to determine the relative importance on different bases to generate a suitable salient region. The IOR process is implemented by masking the currently attended focus of attention for the next attention cycle. The model presented in [MCBT06] implements the opponent color theory by computing the color distance in Krauskop's color space. Contrast sensitivity functions are applied on the three color components in the frequency domain. The saliency of an achromatic structure is enhanced if this structure is surrounded by a high contrast in chromatic channels.

2.4.2 Rarity-Based Models

Models of this category concentrate on finding locations in the visual input that contain rarity with respect to a considered feature. The method of [Ste01] for color saliency picks a selected set of neighbourhood pixels around a target pixel and compares it with a similar pattern of neighbourhood at several test locations. The exclusiveness is computed by subtraction of color components of every corresponding pixel in the neighbourhood patterns around the target and each test location. A large value of this exclusiveness adds a score of saliency to the target. The sum of these scores after checking a number of test locations decides the final saliency value for the target. Another work presented in [Ahu96] generates the color contrast map according to rareness criteria on feature maps of intensity contrast, saturation contrast, and hue contrast. Intensity and saturation is convolved with a Laplacian of Gaussian kernel at each point. The circular nature of hue is normalized before applying the convolution. The orientation map is constructed using a rareness criteria using a Gabor kernel of four different angles on intensity, saturation, and hue and then picking the maximum as the resultant. The model of [AL06] also takes rarity in terms of visual features into account to identify salient regions in the scene.

2.4.3 Object-Based Models

In this category of attention modelling the computation of saliency is done on basis of higher level units instead of individual pixels. Objects are formulated by combination of features and clustering of points that belong together due to similarity of some attribute. The model presented in [SF03] computes object-based saliency depending on groupings. A grouping is considered to be a hierarchical structure of "objects and space"; hence it may be a point, an object, a region, or a structure of other groupings. The primary features are extracted exactly as done in [IKN98], but it constructs the intensity, color (red, green, blue, and (yellow), and orientation pyramids after applying a Gaussian filter and then a Gabor steerable filter on the five feature channels of intensity, red, green, blue, and yellow. The shift of attention is carried out by using an algorithmic approach with a coarse to fine strategy.

Some models have partially region-based components in their strategy. The attention model of [BMB01] uses a region-based approach for construction of maps for color and eccentricity. The conspicuity of a region in terms of color is calculated as the mean gradient along its boundary to the neighbour regions. The color gradient between the two regions is defined as the Euclidian distance between mean values of the color components in MTM color space. The eccentricity map is constructed using moments of segmented regions. The model proposed by [LLY+05] also utilizes a region-based method for the feature of color contrast and texture contrast. They include skin color and face existence as cognitive features for attention. A three step approach is used for color contrast in which the image is first clustered using a k-means algorithm. The biggest cluster having a large enough size is considered as background and then color difference of each cluster is computed in contrast to the background. The resultant map is scaled and truncated to remain within prescribed limits.

Formation of objects from raw pixel data is a significantly complex task. The model proposed in [SF03] remained till a theoretical proposition without going into details of implementing formation of the so-called groupings. Other models in this category used the early clustering approach but suffered from the computational complexity resulting in fairly long response time. The approach proposed in this dissertation is an effort to make advancements and innovations in the methodology of the early clustering paradigm in order to make object-based approach usable in real-time attention systems.

2.5 Selection Of Color Model

Segmentation of color images has to follow a more complex approach as compared to gray-scale segmentation. In gray-scale, only intensity information is available for which the computer can discriminate 256 levels. On the other hand colored images contain compound structure comprising of at least three components, leading to millions of colors per pixel.

These components have different meanings and roles in different models, called color-spaces, such as RGB, CYMK, YUV, HSI, and CIELAB etc. Each space has advantages and limitations in terms of segmentation and selection of the best color space is one of the major difficulties in color image segmentation and the decision is mostly made depending upon the requirements of target application.

2.5.1 Types Of Color Model

There are different types of color models available. Some of which are discussed here;

2.5.1.1 The RGB Color Model

In RGB model, each color appears in its primary spectral components red, green and blue. This model is based on Cartesian coordinate system. This model is used for display devices where colors are produced by emission of light.

2.5.1.2 The CMYK Color Model

CYMK is utilized in printing systems where colors are produced by inks that absorb certain wavelengths while reflecting some other. These spaces are suitable for synthesis but not appropriate for analysis of colors as some shade of the same color may have a fairly distant representation making it difficult to consider them as similar.

2.5.1.3 The CIELAB Color Model

The CIELAB space is a perceptually uniform representation of color but its values do not define absolute colors unless a reference white point is specified whose definition is assumed to follow a standard and is not explicitly stated [LJBV05].

2.5.2 Requirements Of Color Model For Our Project

Our requirement of color model for the purpose of visual attention is to obtain regions from such that the regions may have minimum influence of shades and shadows so that a uniformly colored images surface does not get split into many segments.

2.5.3 Why HSI is Selected

HSI (Hue- Saturation-Intensity) space, sometimes also called HSV (Hue-Saturation-Value) or HSL (Hue-Saturation-Luminance), has a good representation of the colors of human perception [CL94] and has good capability of dealing with highlights, shades, and shadows [TT92].

2.6 Selection of Segmentation Algorithm

In this section segmentation techniques and segmentation requirements (why segmentation is needed for our system), are discussed.

2.6.1 Different Segmentation Techniques

We include a review of colour segmentation approaches here because an innovative method for clustering is also developed during the course of this project. This review will help in developing an awareness about the existing techniques and judging the novelness of the proposed segmentation

2.6.1.1 Edge Based Technique:

The edge definition we use can be summarized as i) The gradient magnitude is a local maximum in the direction of the gradient, and ii) A normalized measure of edge strength, the gradient magnitude weighted by the scale, is locally maximum over scales. This definition is unbiased, in the sense that it is not parameterized for particular applications.

Edge based techniques such as [HB90] and [Sin99] have a common problem that they fail to take into account the correlation among the color channels and miss certain crucial information revealed by color [IPV00].

2.6.1.2 Graph Based Technique

Graph based image-segmentation is a fast and efficient method of generating a set of superpixels, also known as segments, from an image. They supersede old edge-based approaches as they not only consider local pixel-based features, but instead look at global similarities within the image. In the published literature, the graph-based algorithm has been run with a fixed, hand-tweaked set of parameters without any systematic optimization. Unfortunately, there exists no direct scientific metric of what makes a “good” image segmentation, other than the general idea that regions which are visually (or even contextually) similar should be grouped together to become part of the same segment. A frequent goal of image-segmentation is to determine the boundaries of each contextual object within the image. Thus, one of the best available methods of evaluating the quality of an image segmentation algorithm is by comparing it to some sort of ground truth, such as images which have been manually segmented by humans. We will follow this approach in this project by comparing the results of a graph-based segmentation algorithm to the corresponding human segmentations.

2.6.1.3 Model Based Technique

In model-based schemes image regions are modelled as random fields and the segmentation problem is posed as a statistical optimization problem. Applied to medical images, segmentation distinguishes between different organs, tissues, vessels, etc. However, the amount of the 3D data that is acquired with CT and MRI can be very large. This is the reason why it often takes medical doctors a lot of time to really understand the data and to analyze and interpret it correctly. Intelligent algorithms which automatically identify important structures in MRI images are intended to support medical doctors in their work.

Model-based techniques forgo computation time for the quality hence is usable only in those cases where computational complexity is not an issue.

2.6.1.4 Region Based Technique

As its name implies, region growing is a procedure that groups pixels or sub pixels into larger regions based on predefined criteria for growth. The basic approach is to start with a set of “seed” pixels and from these grow regions by appending to each seed those neighboring pixels that have predefines properties similar to the seed (such as specific ranges of color or intensity).

Selecting a set of one or more starting point often can be based on the nature of the problem. The selection of the similarity criteria not only depends on the problem under consideration but also on the type of the image data available. When the images are monochrome, region analysis must be carried out with a set of descriptors based on intensity levels and spatial properties.

2.6.2 Role Of Segmentation In Our Project

Segmentation is one of the most important steps in the point of view of machine vision systems. It clusters neighbouring pixels into groups or regions using some homogeneity criterion such as similarity of color. Segmentation plays the basic role in our project. The more fine the segmentation we do the better object we can recognize form an image/video. Our whole project is based on segmentation so it must be up to the perfection level. We want our segmentation algorithm to;

I. Produce regions that are as large as possible (i.e., produce as few regions as possible).
II. Produce coherent regions, but allow some flexibility for variation within the region.

2.6.3 Why We Are Using Region Growing

Region-based techniques are considered better when processing speed is a major issue as they provide acceptable quality within a reasonable computation time [CJSW01] [IPV00]. Region-growing methodology has a natural computational advantage over its split-and-merge counterpart in the same category. In terms of color quantization for segmentation, HSI space is considered most appropriate [LB01] and it has provision of overcoming the illumination effects such as shades and shadows.

This approach has specific advantages over boundary based methods:

a. It is guaranteed (by definition) to produce coherent regions. Linking edges, gaps produced by missing edge pixels, etc. are not an issue.
b. It works from the inside out, instead of the outside in. The question of which object a pixel belongs to is immediate, not the result of point-in-contour tests.

Region-based techniques work best on images with an obvious homogeneity criterion and tend to be less sensitive to noise [CJSW01]. There are two typical approaches available under this category. One is to use region split-and-merge as in [OPR78] and other is region-growing, for example [TB97]. The later usually applies a merging step to combine segments having further similarity of color features.

As the target application of our segmentation routine is a biologically inspired attention system, the factor of human perception will have a prominent influence on each step of the process. We would like to build homogeneous segments in a given scene that are potentially distinct regions for humans. Surveys on the color segmentation techniques, for example [SK94] and [CJSW01] show that the area-based segmentation gives a good optimization of segmentation quality and processing time. In the area-based category, the region growing procedure has the advantage of computational simplicity over the split-and-merge strategy. The same surveys establish the advantage of using hue for discounting illumination effects like shading, shadowing, and highlights according to human perception over its other counterparts. As our problem domain has a severe restriction of computation time, we infer to select a region growing method using HIS (hue, intensity, saturation) color space in the proposed segmentation technique. However we suggest introduction of some innovative enhancements in order to improve the quality of output without causing much escalation in the computation time.

2.7 Top-Down And Bottom-Up Approach

After extraction of feature maps next step in our work break down structure was to achieve top down and bottom-up saliency by considering behavior influence. We studied the human behavior while during visual attention and found out that human follow bottom up while exploring, follow top-down approach while searching and both while examining/tracking [LD04], [LHG97], [Ham05], [Dec05], and [NI06b] .The well-known model discussed in [IKN98] and [IK00] mainly deals with search behavior but uses bottom-up procedure for this purpose. The selective tuning model [RT06] remains in a behavior resembling explore as it does not apply top-down conditions to excite the target of search and lets the salient items pop-out during a process of bottom-up saliency and inhibitions. The models of [PSL02] and [MCBT06] are restricted only to explore while the model given in [SF03] discusses both explore and search behavior by integrating bottom-up and top-down biasing in the process of hierarchical selectivity. The model of [BMB01] considers three behaviors of explore, search, and detect changes while [FBR05] implements explore and search for dynamic scenes.

2.8 Chapter Summary

The procedures proposed by existing theories of the visual attention, are computationally expensive as the correlated processes carried out in the brain are significantly complex. Hence, there is a need to investigate ways to reduce the complexity of the computational model of attention and enable it to integrate with other vision algorithms so that the output of attention may become an effective contributor for improving performance of the vision system.

CHAPTER 3

REQUIREMENT ANALYSIS

3.1 Introduction:

This section of the document contains a structure for a software requirements specification (SRS) document. It describes the services and functions which the system should provide, the constraints under which the system must operate, overall properties of the system i.e. constraints on the system's emergent properties and definitions of other systems which the system must integrate with. There are misunderstandings between customers, those developing the system requirements and software engineers developing or maintaining the system. So this section of document will help in clarifying the requirements of the proposed system. Some of the intended audiences of this section include system customer, project manager, system engineer, system test engineer and system maintenance engineer.

This chapter includes the functional requirements that we inferred from the scope and the objectives of the project. It also discusses the non-functional requirements that must be achieved for better performance of the system.

3.2 Overview

This project is being done with the collaboration of Paderborn University, Germany. We basically wanted to develop a visual system for a robot system. The algorithms which were previously used were based on spatial and frequency domain filters which were not efficient, they lagged in performance and used only a single technique either top-down or bottom-up. But our system will use region based approach to reduce the number of units that have to be processed by the attention systems. Our system will use top-down, bottom-up techniques and faster algorithms that will help in increasing performance and efficiency of processing to integrate the attention procedure into a comprehensive biologically inspired vision system.

3.3 Business Context

As technologies has advanced a lot in past decades, it will be a matter of time when robots will act humanly on visual input and extract the salient objects based on biological visual attention. So now there is a need to make an effective visual system so that robots can work exactly as humans

3.4 Specific Requirements

In this section functional and non-functional requirements are discussed.

3.4.1 Functional Requirement

1. The system should have a graphical user interface that should provide it with the platform to upload any image/video in the supported format (BMP, DIB, JPEG, JPG, JPE, PNG, PBM, PGM, PPM, SR, RAS, TIFF, TIF, AVI, 3GP, other video formats).

a. **Description:** Interface must be there ad should support as many types of images as it can. It will make our system user friendly.
b. **Criticality:** Interface is not a building block of our system.
c. **Technical-Issues**: .GIF is not supported.
d. **Risks:** As the interface is being made in QT and we are not very familiar with the environment

2. On successful upload the system should display a segmented image and feature maps as an output in another window.

a. **<u>Description</u>:** As the user upload the image and it is in a proper format system should display a segmented image and all the five feature maps i.e. color, size, eccentricity, orientation, symmetry.

b. **<u>Criticality:</u>** This step s very critical as the top-down and bottom-up maps depend on segmented image and feature maps.

c. **<u>Technical-Issues</u>**: Image can be noisy so it can give wrong results.

d. **<u>Risks</u>:** if the image is divided into wrong segments it will give poor results.

3. It should make proper Top-down and Bottom-up maps.

a. **<u>Description</u>:** Proper processing must be done on the previously produced feature maps to form the top-down and bottom-up maps.

b. **<u>Criticality</u>:** the processing on the maps must be proper because the focus of attention depend solely on these maps.

c. **<u>Technical-Issues</u>:** it can produce wrong results if proper weights are not assigned.

d. **<u>Risks</u>:** If the feature maps are giving wrong values then obviously it will affect the results of top-down and bottom-up maps.

4. If motion is there in the input video then it should give priority to moving object.

a. **<u>Description</u>:** if there is an object moving in the uploaded video then obviously our focus of attention will be that moving object.

b. **<u>Criticality</u>:** This requirement is critical the moving body must be detected.

c. **<u>Technical-Issues</u>**: As we are not dealing with stereo images, so depth factor is not considered. All the objects are at equal distance for the viewer. Actually that's not the case. In reality if there is a moving object far away in the scene humans might not notice but as the objects are at equal distance from viewer our system focuses on the object even if it is far away in the scene.

d. **<u>Risks</u>:** If there is an error in reading a frame of a video this motion detection might not work properly because it depends on the consecutive frames.

5. Once the feature maps are created the user must be provided with the following options;

1. **<u>Search:</u>**

   This feature should offer user to perform top-down approach towards computation of saliency. It should let the user to search any image in the uploaded image/video.

   i. **<u>Priority</u>:** the priority of this feature is high and it is must to achieve it.

   ii. **<u>Precondition</u>:** Image/video is successful fully uploaded and search option is selected.

   iii. **<u>Functional Requirements:</u>**

   1) The system displays an image.

2) User selects an object from the image

3) Then the system should display the image/frame with focus of attention highlighted (by marker).

iv. <u>Constraint/Assumption</u>:

1. Assuming there are no administrative right issues and system is not corrupt.

2. <u>Explore:</u>

This feature should offer user to perform bottom-up approach towards computation of saliency. It should be able to highlight the most salient objects in an image with respect to human visual behavior with a marker.

i. <u>Priority:</u> Priority of this feature is high and it is must to achieve it.

ii. <u>Precondition:</u> Image/video is successful fully uploaded and Explore option is selected.

iii. <u>Functional Requirements:</u>

1) The system should display the image/video with highlighted (by marker) focus of attention.

iv. <u>Constraint/Assumption:</u>

1. Assuming there are no administrative right issues and system is not corrupt.

3. <u>Examine or Track:</u>

This feature should offer user to perform both top-down and bottom-up approaches towards the computation of saliency. Precondition for this feature is that object to examined/tracked should be within the given image/frame. If video is uploaded than the object should be tracked though out the video. If image then the object is examine within that image.

i. **Priority:** Priority of this feature is low and it is an optional feature if time permits.

ii. **Precondition:** Image/video is successful fully uploaded and Examine/Track option is selected.

iii. **Functional Requirements**:

1) The system display an image

2) User selects an object to be tracked or examined.

3) The system should display the image/frame with focus of attention i.e. the target by highlighting it by marker.

iv. **Constraint/Assumption:**

1. Assuming there are no administrative right issues and system is not corrupt

3.4.2 Non-Functional Requirements

1. Performance:

The system should have a mechanism for maintaining a reasonable level of performance. This is particularly important if a real heavy image/video is uploaded.

2. Reliability:

The system should have a low failure rate and a high level of service availability. As this system is for real time robot so it must be reliable.

3. **Efficiency:**

Algorithms used must be so efficient that it should take as less time as possible because it is a real time system.

4. **Usability:**

As it is a desktop application it's interface must be such that anyone could use it without any prior training.

5. **Robustness:**

The system should be robust and able to generate similar result on the transformed image as of original image.

3.5 System Requirements

This part of the document describes the relationship between the system, its components and the external environment of the system. We are using debian as an operaying system and netbeans for developing in c++.

3.6 Hardware Requirements

Presently, following hardware components and interfaces are needed but more can be required if advance level support/functionality is desired of the system

1. Core 2 Duo computer with at least 1 Gb Ram.

3.7 Software Requirements

The functional modules of our system require following software to properly work.

1. Netbeans IDE, Visual Studio
2. Debian (Linux) as an operating system
3. Opencv
4. Gazebo and player for simulation

3.8 Chapter Summary

This chapter briefly described the functional requirements of our system. These functional and non-functional requirements are extracted from the scope of our project.

CHAPTER 4

SYSTEM DESIGN AND ARCHITECTURE

4.1 Introduction

This document contains a structure for a software design specification. Most software projects fail because of the flaws in the design, so the design phase can be referred to as a very crucial stage in the software development lifecycle. The purpose of writing this document is to describe the design of the system in detail. This document contains a detailed description of the design of the system, and helps in clearing any doubts which might have been left in the specification of the requirements of the system. Misunderstanding between the users of the system and those developing it can further be clarified if the user can see the design of the system. So this document will help in clarifying the requirements as well as design of the proposed system. Some of the intended audiences (readers) of this document include system customer, project manager, system engineer, system test engineer and system maintenance engineer.

4.2 System Overview

This part of the document describes the design relationship between the system, its components and the external environment of the system. There is a segmentation module which gives segmentaed image as an output. This segmentated image is used by the feature map module so make feature maps according to color, size, orientation, eccentricity and symmetry. These maps are used by top-doen and bottom-up modules to get the focus of attention.

4.3 System Architecture

Architectural Diagrams shows the basic architectural layout of the system being designed. In our system, first of all, the user inputs an image/video. Then image is divided into segments. Then this segmented image is used to form the feature maps. With the help of these maps the saliency is computed.

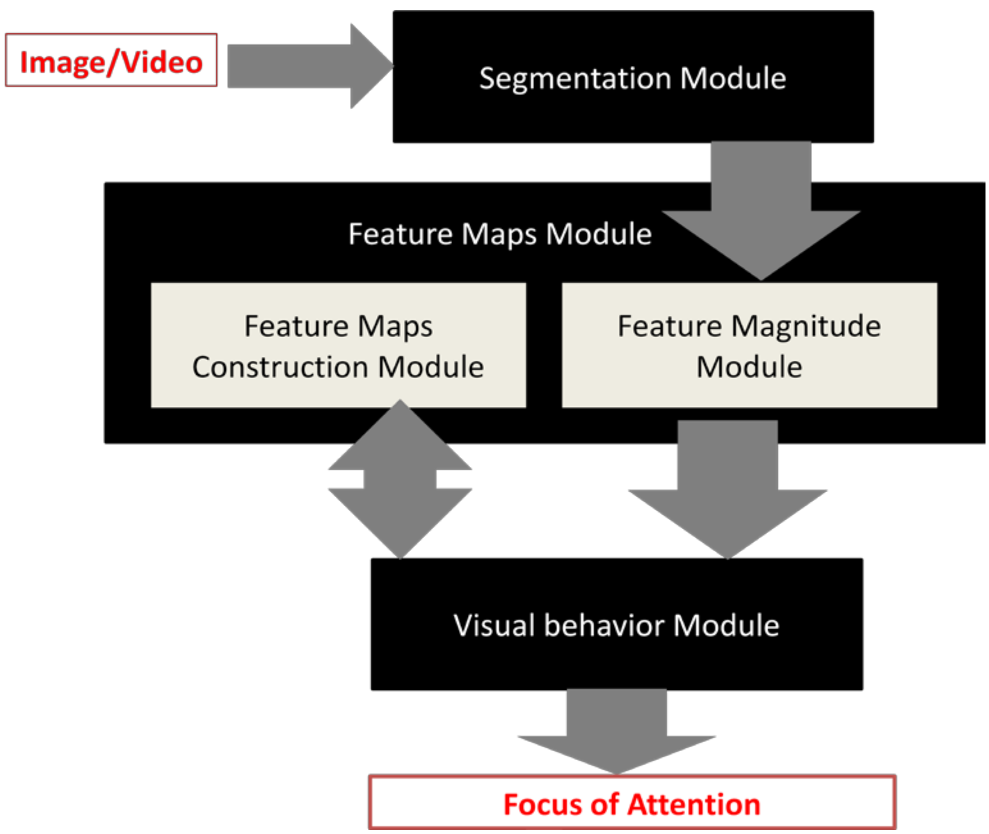


Fig 4.1 Architecture Design of BIDSO

4.4 Decomposition Description

Under this heading, the content matter will describe the decomposition description of the different components of the system. In BIDSO, there are five main components which are discussed in detailed as follows:

4.4.1 Segmentation Module

The segmentation algorithm works in two phases of operation. In the first phase, only those chromatic pixels are selected as seeds for region growing that have high amount of saturation and intensity. This condition is limited only to high intensity for achromatic seeds. It facilitates to begin the region growing from prominent portions of objects and helps to position the seeds at central locations so that areas of their corresponding regions are evenly spread around them. Another advantage of this step is that it allows those regions to grow first that have a potential probability of attracting visual attention otherwise there are chances that such regions get merged into other unattractive segments and get neglected in the attention process. In the second phase, the restrictions on saturation and intensity of the seeds are lowered in order to allow the left over areas to get segmented. Values of the thresholds are also relaxed in such a way that the remaining pixels get a higher chance of joining some segmented region. Seeds with black color are not allowed to grow in the first phas. Similarly gray seeds are avoided in the first phase because each color turns into gray at low saturations. These two colors get their opportunity in the second phase.

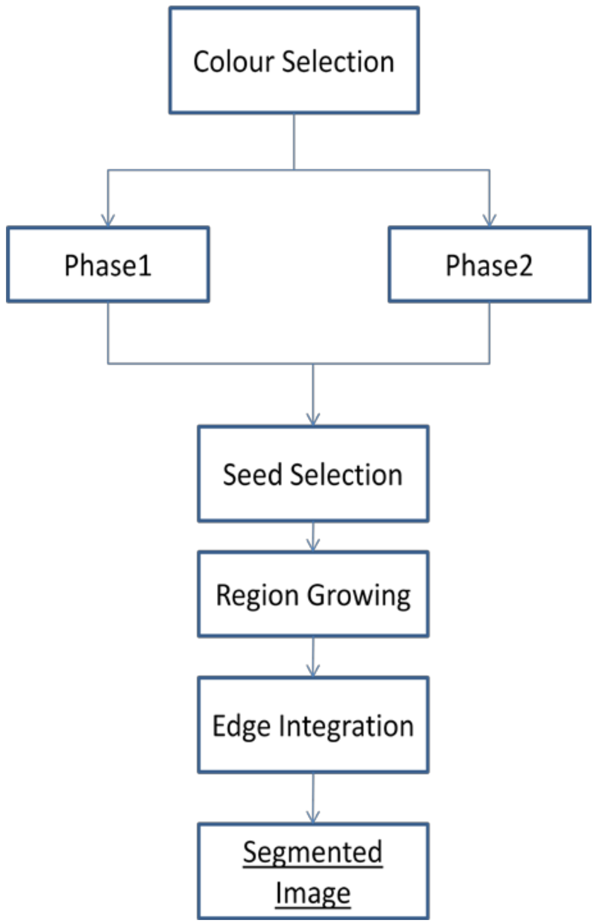


Fig 4.2 Architecture Design of Segmentation Module

4.4.2 Feature Maps Construction

This is the next phase of BIDSO. In this we construct feature maps and pop-out the most prominent region according to the selected feature. Features we are considering are color, size, symmetry, orientation, eccentricity.

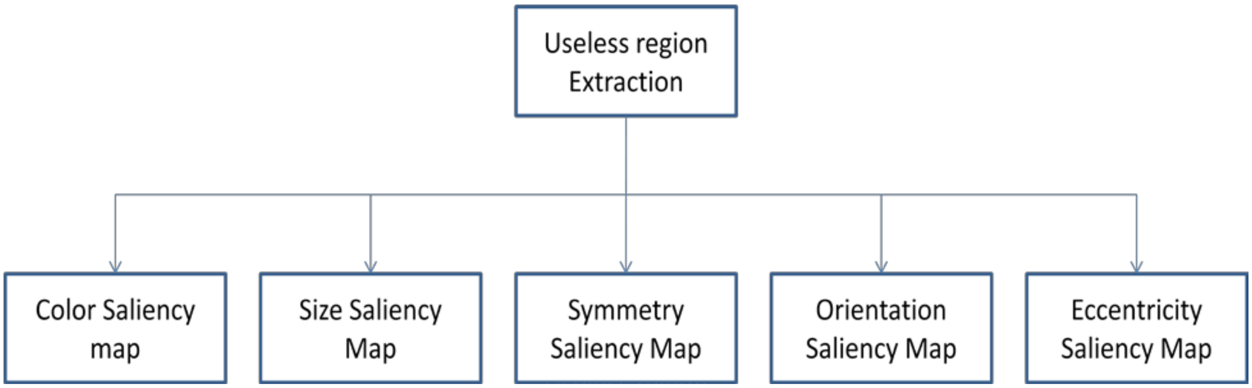


Fig 4.3 Architecture Design of Feature Map Module

4.4.3 Bottom-Up Fusion/Explore Map

The resultant bottom-up saliency map is obtained by summation of all feature maps of this pathway after applying the weights to the feature maps related to the active behaviour of attention obtained. If there is some motion captured priority is given to moving object.

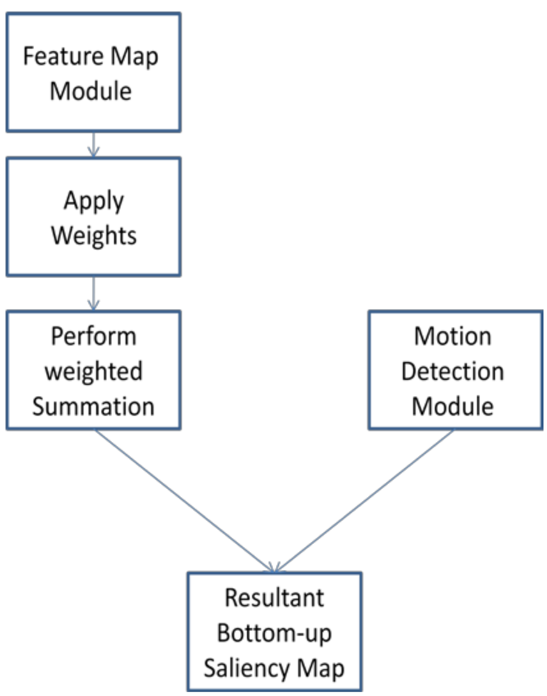


Fig 4.4 Architecture Design of Bottom-up/Explore Module

4.4.4 Top-Down Fusion/Search Map

The resultant top-down saliency map is constructed through summation after applying the weights to the feature maps related to the top-down pathway.

Fig 4.4 Architecture Design of Top-down/Search Module

4.5 Use-Case Diagram

Use Case diagram explains how the possible users of the system would interact with the system and what all are the functions they can perform according to their roles in the system.

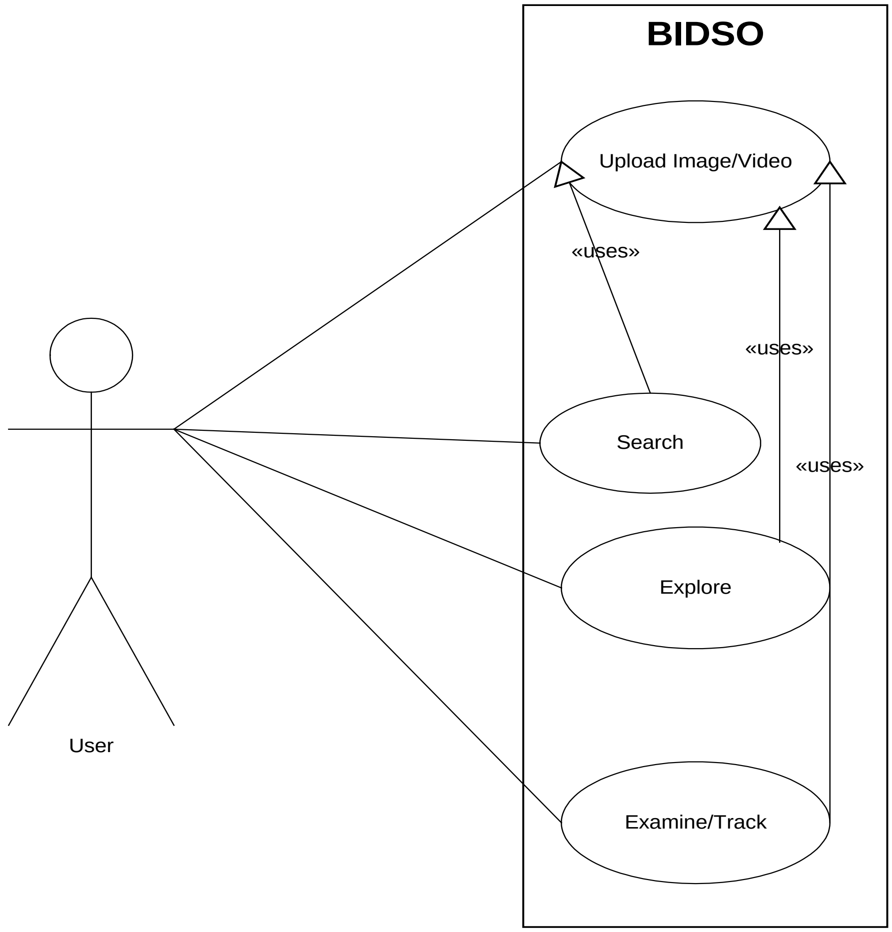


Fig 4.5 Use-case diagram of our system

4.5.1 Fully Dressed Use-cases and Sequence Diagrams

4.5.1.1 ***Use case UC1: Upload***

1) **Primary Actor**: User

2) **Preconditions**: input image is in format supported

3) **Success Guarantee (Post conditions):** Feature magnitude calculation and segmentation frames/image is returned.

4) **Main success scenario / Basic flow:**

   1. User uploads the image or video in the format supported(BMP, DIB, JPEG, JPG, JPE, PNG, PBM, PGM, PPM, SR, RAS, TIFF, TIF, AVI, 3GP, other video formats)

   2. System performs segmentation and saves the region list obtained from segmentation.

   3. System returns segmented image/frame to the user.

   4. System determines the feature magnitudes (color, size, eccentricity, symmetry, orientation) and saves the feature maps.

   5. System displays the segmented image to the use.

5) **Extensions/ Alternative flow:**

   1a. Invalid format entered:

   1. Give error message and ask user to upload the image/video again

2a. Due to memory issues feature magnitudes are not saved:

1.Generates an error message “memory leak-out problem, contact administrator”.

4a. Due to memory issues region list is not saved:

1. Generates an error message “memory leak-out problem, contact administrator”.

6) **Special requirements**: Image/video should be noise free and Camera calibrated (if you want to read video from camera)
7) **Technology and Data Variations List**: Read image/video from camera if required.

Sequence Diagram of UC1: Upload

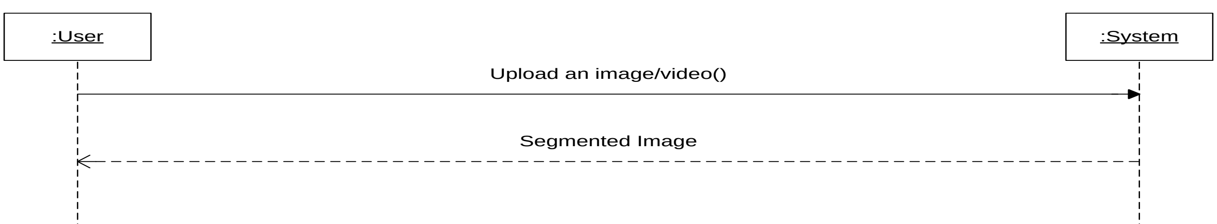


Fig 4.5.1.1 Sequence diagram of UC1

4.5.1.2 Use case UC2: Search

1) **Primary Actor**: User

2) **Preconditions**: Region list and Feature magnitudes are available and user selects the search option.

3) **Success Guarantee (Post conditions):** Outputs the focus of attention (image required by the user) and intermediate top down saliency maps only on user request.

4) **Main success scenario / Basic flow:**

   1. User select search option for the first time for image uploaded.

   2. Select an image window is displayed by the system.

   3. User selects an object within that image.

   4. System determines the final top down saliency map.

   5. System returns the output image with Focus of Attention marked by a marker.

5) **Extensions/ Alternative flow:**

   1a. if the search option is not selected for the first time the same image

   1.Step-2 in main flow will be skipped will be skipped.

5a. Due to memory issues region list is not saved

1.Generates an error message "memory leak-out problem, contact administrator".

6a. System asks whether the user want to see intermediate top down saliency maps and final saliency map or not.

1.If the user want to see the intermediate top-down feature maps

i. System displays the top-down intermediate Saliency map and final saliency map

ii. System displays the image with highlighted focus of attention.

2. If the user doesn't want to see intermediate maps and final saliency maps

i. System displays the image with highlighted focus of attention.

6) **Special requirements**: No special requirement

7) **Technology and Data Variations List**: No technology and data variations

Sequence Diagram UC2:Search

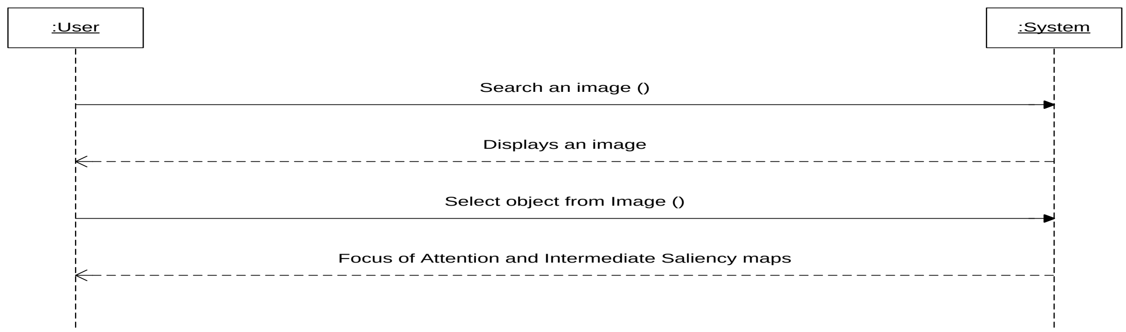


Fig 4.5.1.2 Sequence diagram of UC2

4.5.1.3 Use case UC3: Explore

1) **Primary Actor**: User
2) **Preconditions**: Region list and Feature maps are available and user selects the Explore option.
3) **Success Guarantee (Post conditions):** Outputs the focus of attention (image required by the user). Also outputs the final bottom up saliency map and intermediate bottom up saliency maps only on user request.
4) **Main success scenario / Basic flow:**
    1. Use selects the Explore option.
    2. System computes the bottom up saliency map.
    3. System determines the final bottom up saliency map and saves the map.

4. System returns the output.

5) **<u>Extensions/ Alternative flow:</u>**

3a. Due to memory issues region list is not saved:

1. Generates an error message “memory leak-out problem, contact administrator”.

4a. System asks whether user the user want to see intermediate bottom up saliency maps and final bottom up saliency map or not.

1. If the user want to see the intermediate bottom-up feature maps
   - **i.** System displays the intermediate bottom-up saliency maps and final bottom-up saliency map.
   - **ii.** System displays the image with highlighted focus of attention.

2. If the user don’t want to see intermediate maps and final saliency maps
   - i. System displays the image with highlighted focus of attention.

6) **Special requirements**: No special requirement

7) **Technology and Data Variations List**: No technology and data variations

Sequence Diagram of UC3: Explore

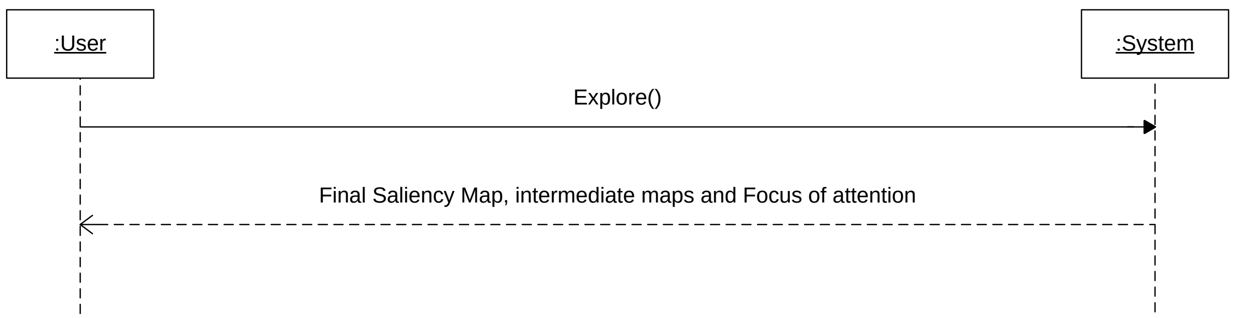


Fig 4.5.1.3 Sequence diagram of UC3

4.5.1.4 Use case UC4: Examine/Track

1) **Primary Actor**: User
2) **Preconditions**: Region list and Feature maps are available and user selects the Examine/Track.
3) **Success Guarantee (Post conditions):** Outputs the focus of attention (image required by the user). Also outputs the final bottom-up and top-down saliency map and intermediate bottom-up and top-down saliency maps only on user request.
4) **Main success scenario / Basic flow:**
   1. Examine/Track option is selected.
   2. System computes the bottom-up saliency map.

3. System computes the top-down saliency map.
4. System determines the final bottom-up and top-down saliency map saves the maps.
5. The system returns the output.

5) <u>Extensions/ Alternative flow:</u>

4a. Due to memory issues region list is not saved:

1. Generates an error message “memory leak-out problem, contact administrator”.

5a. System asks whether user the user want to see intermediate top-down and bottom-up saliency maps and final top-down and bottom-up saliency map or not.

1. If the user want to see the intermediate top-down and bottom-up feature saliency maps

- **i.** System displays the intermediate bottom-up and top-down saliency maps, final bottom-up and top-down saliency map.
- **ii.** System displays the image with highlighted focus of attention.

2. If the user don’t want to see intermediate maps and final saliency maps

- **i.** System displays the image with highlighted focus of attention.

6) Special requirements: No special requirement

7) **Technology and Data Variations List**: No technology and data variations

Sequence Diagram of UC4: Examine/Track

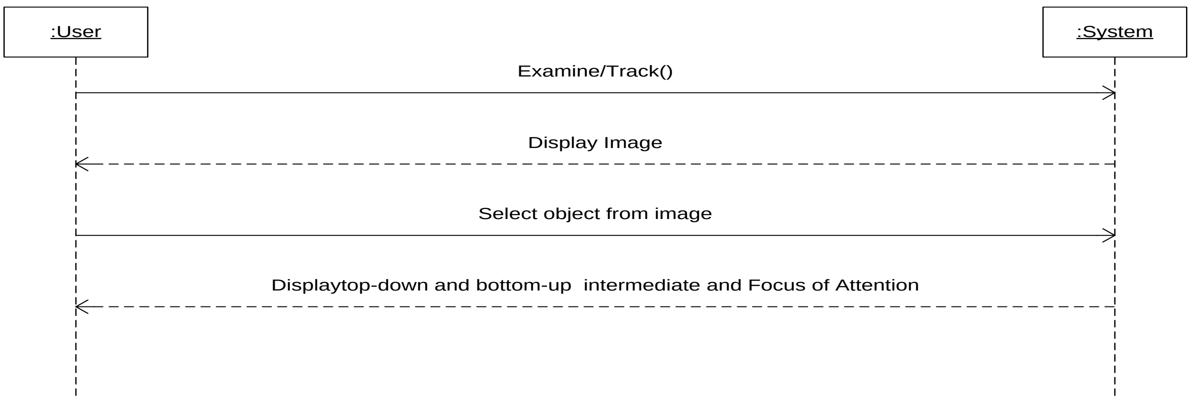


Fig 4.5.1.4 Sequence diagram of UC4

Sequence Diagram of BIDSO

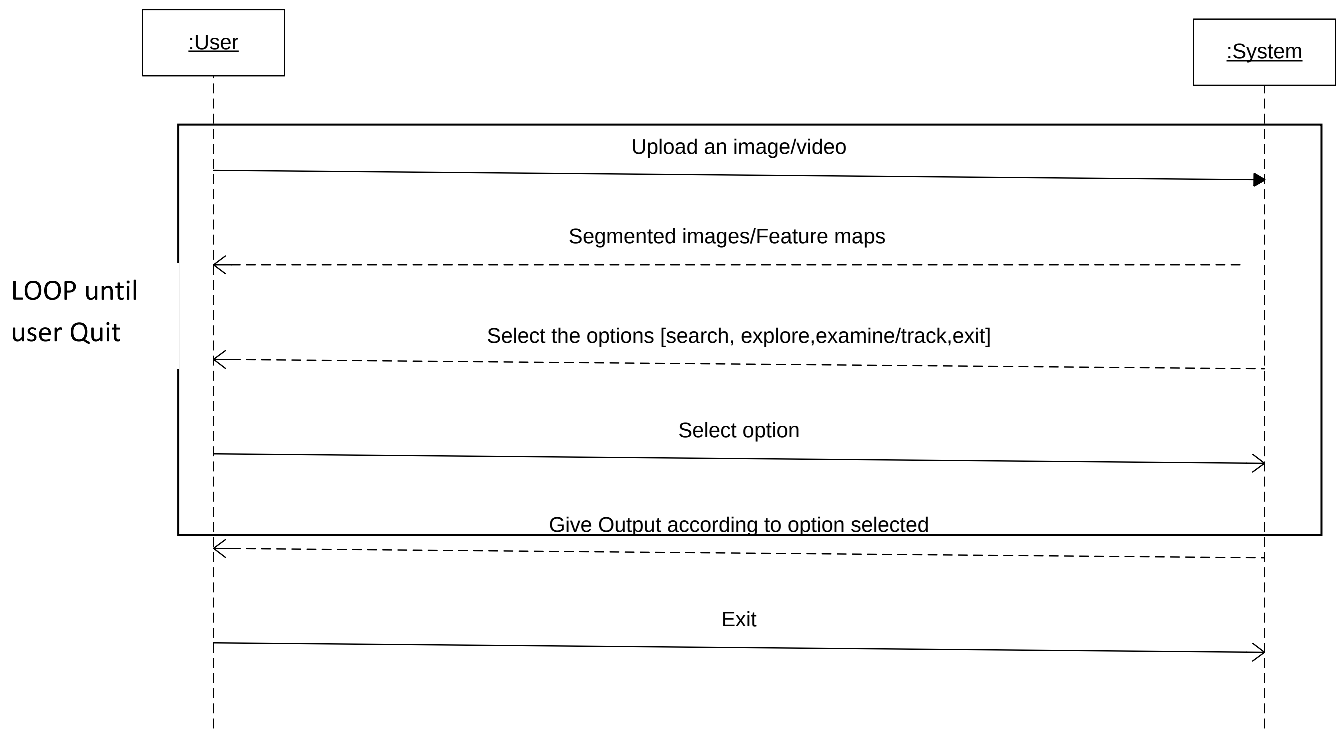


Fig 4.5.1.4 Sequence diagram of our system

4.6 Class Diagram

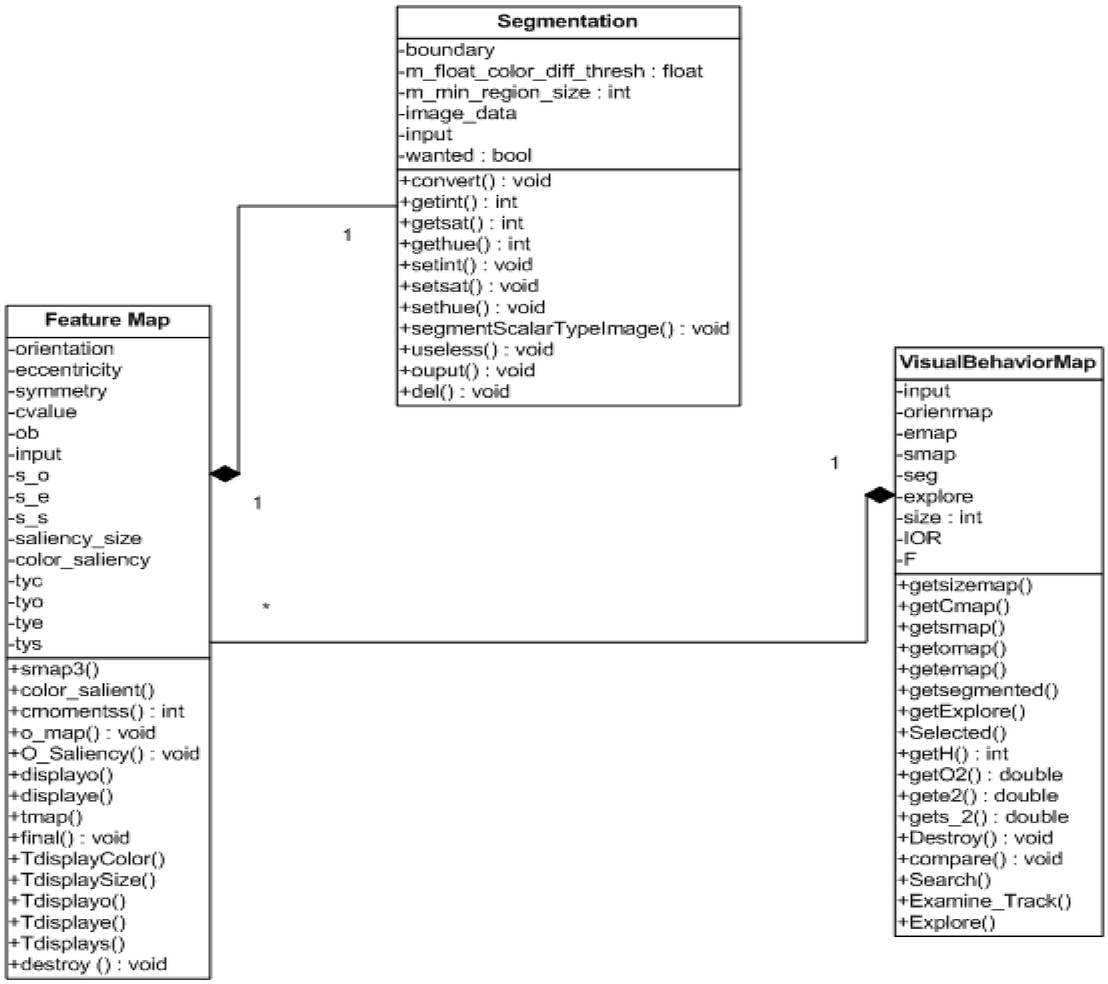


Fig 4.6 Class diagram of our system

4.7 Data-Flow Diagram

This diagram shows how the data is flowing through different modules of our system.

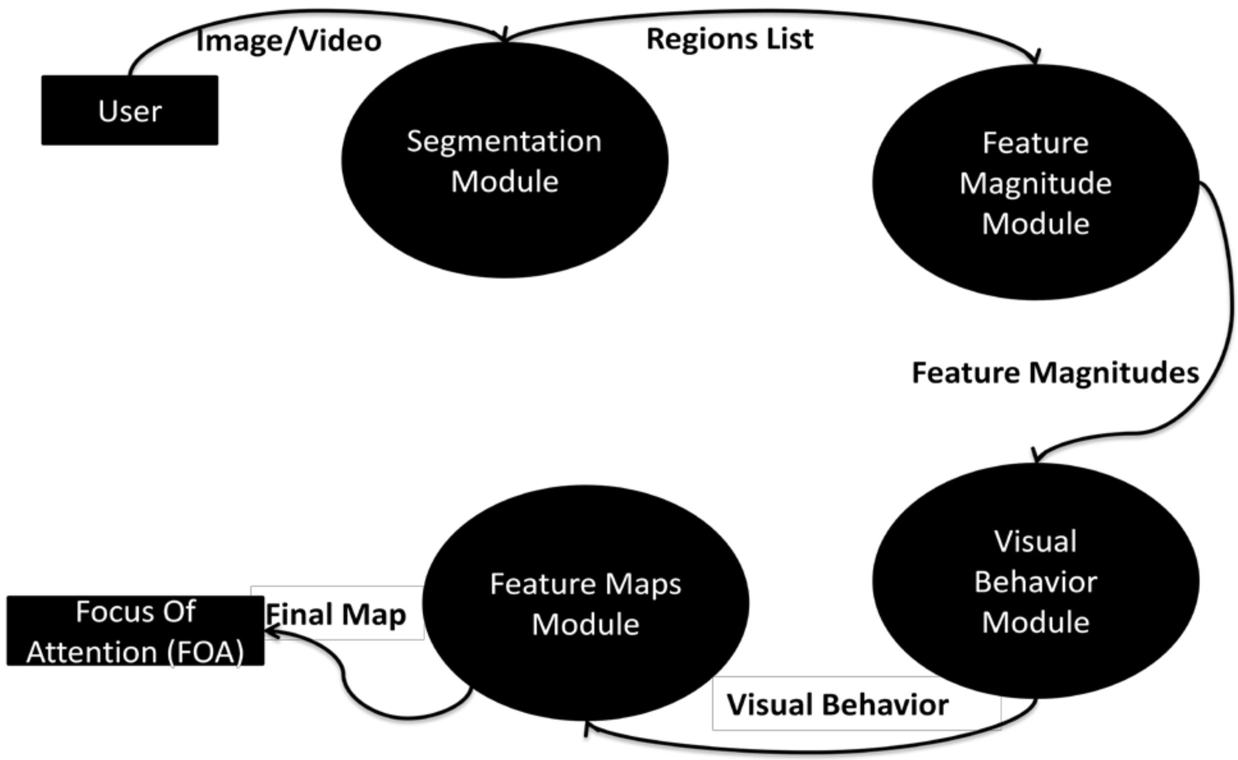


Fig 4.7 DFD level 1 of our system

4.8 Chapter summary

The overall architecture of the proposed attention model has been sketched with description of its constituting building blocks. The main advancement as compared to the existing attention models is the integration of bottom-up and top-down pathways into a single architecture and inclusion of explicit influence of visual behaviors in internal steps of the model. The behavior dependant functionality of different modules has been explained in detail.

CHAPTER 5

SYSTEM IMPLEMENTATION

5.1 Introduction

This section of the document contains the implementation details of the system. It describes the functionality of the system and also a complete explanation of the system from the implementation point of view. The basic idea is to make the reader familiar with the implementation details of the system so that he can have some idea about the actual working of the system.

5.2 Overview

Basically the purpose of our project is to find the focus of attention. We are developing this system for a robot so that it can see like human beings. There are three ways by which human focuses. The human being:

i. Focuses on the most prominent object in the scene.

ii. Focuses on the object he wants to find

iii. Focuses on the object which is moving

For most prominent object we are using the bottom-up technique, for guided search we are using top-down technique and for moving object we are using motion tracking algorithm.

5.3 Implementation

In this section the implementation algorithms and techniques which we have used are discussed.

5.3.1 Segmentation Module

In this section the parts of the segmentation modules are discussed.

5.3.1.1 Seed classification

The segmentation algorithm works in two phases of operation.

In the first phase, chromatic pixels are picked as seeds for region growing that have high amount of saturation and intensity. It facilitates to begin the region growing from prominent portions of objects and helps to position the seeds at central locations so that areas of their corresponding regions are evenly spread around them.

In the second phase, the restrictions on saturation and intensity of the seeds are lowered in order to allow the left over areas to get segmented. Values of all concerned thresholds are also relaxed in such a way that the remaining pixels get a higher chance of joining some segmented region. Seeds with black color are not allowed to grow in the first phase due to convergence of all hues and unpredictable behaviour of saturation at extremely low intensity. Similarly gray seeds are avoided in the first phase because each color turns into gray at low saturations. Hence they can swallow major parts of neighbouring regions with chromatic colors due to overlap of hues at this saturation level. These two colors get their opportunity in the second phase.

Color Classification:

The color of seed pixel is classified at two stages.

First Stage: In the first stage, the purpose is to select an appropriate procedure according to the nature of the seed color. At this point the seed is evaluated to see if it is white, black, gray, or with a chromatic color because the process of region growing will be different for each of these seeds.

Second Stage: The second stage categorization is needed only for the seeds with chromatic colors where class of the seed's hue has to be determined in order to activate a suitable set of thresholds.

.

5.3.1.2 Integrated Edge and Region Homogeneity Check

The algorithm integrates the testing of the color homogeneity between the seed and rest of the region pixels with the check to determine if the region border has been reached. This allows the regions to grow with suitable tolerance to illumination and other colors variations and, at the same time, distinguishes the fine edges between them. In order to achieve this we establish four different sets of checks one of which will be activated according to the category of region seed.

Check for Black:

For region growing around the black pixels the check is the simplest for it. We just need to check that if given the color of the neighbour pixel the intensity of the pixel must be less then the threshold of intensity of black. i.e.

intensity(pixel) < minimum intensity of black color (5.1)

Check for Gray:

For growing region around gray we need to perform three checks.

1. Firstly the neighbour pixel being tested should be gray, i.e., its saturation has to be below the saturation for gray.

2. The second test is on the intensity difference between the seed color and the neighbour color which should be under the allowed intensity difference between the seed and other pixels of a gray region.

3. Thirdly, the intensity difference between neighbor pixel and its neighbor must be less than threshold value. This is to check whether there is an edge between pixels or not.

Check for White:

Checks for growing region around white are same as that for Black. We just need to check that if given the color of the neighbour pixel the intensity of the pixel must be greater then the threshold of intensity of white. i.e.

$$\text{intensity(pixel)} > \text{intensity of white color} \tag{5.2}$$

Check for Chromatic:

There are a total of 9 checks for chromatic pixel

1. The three are meant for stopping the growth of a region if the neighboring pixel is black, gray or white. These checks are important as the gray areas can have arbitrary values in their hue channel while the white and the black regions may contain subjective values for both hue and saturation, hence gray pixels can easily get swallowed when a neighbouring chromatic region is growing.

2. The three checks are to allow small differences between hue, saturation, and intensity components of the seed color and those of the neighbour.

3. Then three checks inspect if two adjacent neighbour pixels in the given region form an edge in terms of hue, saturation, or intensity.

5.3.1.3 Region Construction

Region construction is carried out using a usual region growing procedure in which all pixels found eligible to join the region in an 8-connected neighborhood are labeled with the region identity and pushed in a stack. Later the stack is popped and the same procedure is repeated for the popped pixel considered as part of the region. This process continues until the stack gets empty.

5.3.1.4 Filtration of Useless Areas

The regions which are much less or greater as compared to the size of the image are discarded. The regions that are greater than 40% of the image are treated as background. While the regions that are less than 10% of the image are considered as segmentation error. We have used the factor of perimeter which is max for the regions of moderate size and small for small and big regions.

5.3.2 Feature Maps

This is the second phase of our project. We will construct the maps on the basis of five features namely color, size, orientation, eccentricity, symmetry. This is to find out which is region is prominent according these features.

5.3.2.1 Color Map

Color map is constructed using the theory discussed below;

5.3.2.1.1 Color Contrast in Color theory

Apart from psychology, valuable information can also be found in the literature on color theory about the attributes of colors that contribute in making an object visually prominent or receding. In terms of color saliency, other methods of artificial visual attention have concentrated only on those attributes of colors that were reported in psychology and many important aspects described for this purpose in the color theory have been neglected. Artists practice these aspects for creating effects of contrast, visual advancement, and activeness in their illustrations. Johannes Itten was one of the first experts of color theory who described methods for color combinations offering contrast. He has defined different situations in which the human vision finds contrast in a colored scene. According to his research, the contrast can occur due to presence of objects posing high difference of intensities, saturation, and/or hue. Other reported causes include presence of opponent colors and co-occurrence of warm and cool colors [Itt61].

Another relatively modern source of theoretical concepts on colors is available in [Mah96]. We combine these concepts with those of Itten's and formulate a set of points that are feasible for computation. Another important issue is to decide that which color will receive the benefit of saliency in presence of a contrast. The summarized points with the mention of the saliency winning color in each situation are listed below:

1. **<u>Contrast of Saturation</u>**: A contrast is produced by low and highly saturated colors. The value of contrast is directly proportional to the magnitude of the saturation difference. Highly saturated colors tend to attract attention in such situations unless a low saturated region is surrounded by highly saturated one.

2. **<u>Contrast of Intensity</u>**: A contrast will be visible when dark and bright colors co-exist. The greater is the difference in intensity the higher is the effect of contrast. Bright colors catch the eye in this situation unless the dark one is totally surrounded by the bright one.

3.**Contrast of Hue:** The difference of hue angles on the color wheel contributes to creation of contrast. High difference will obviously cause a more effective contrast. Due to the circular nature of hue, the highest difference between two hue values can be 180 degree.

4.**Contrast of Opponents**: The colors that reside on the opposite sides of the hue circle produce a high amount of contrast. This naturally means that the difference of the hue angles should be close to 180. The colors residing in the first half of the hue circle, known as the active color range, will dominate on the rest of the passive ones.

5.**Contrast of Warm and Cool :** The warm colors namely red, yellow, and orange are visually advanced. These colors are present in the first 45 of the hue circle. Warm and cold colors create a contrast in which warm colors remain dominant.

6.**Accent Colors**: The color of the object covering a large area of the scene will become the ground color (trivial for attention). Colors covering a small relative area, but offering a contrast, are called accent colors. Accent colors get the benefit of contrast in terms of attracting visual attention.

7.**Dominance of Warm Colors**: The warm colors dominate their surrounding whether or not a contrast in the environment exists.

8.**Dominance of Brightness and Saturation**: Highly bright and saturated colors are considered as active regardless of their hue values. Such colors have more chances of attracting attention.

5.3.2.1.2 Color Map Construction

For color map construction we check the above mentioned rules to compute the saliency of a region. Voting style mechanism is used to find the prominent region.

1. **Step 1:** The first step collects votes for each Ri from other regions that possess opponent hues. Two hues are said to be opponent if they lie at the opposite sides on the color wheel. In other words the hue difference should be close to 180 degrees. It is a generalization of the red-green and blue-yellow opponents as used by existing methods.

2. **Step 2:** The second step collects votes from regions that are at large hue distances from Ri.

3. **Step 3:** In the third step we extend the contrast of warm and cool to contrast of active and passive colors. A color is considered as active if its hue is in the first half of the color wheel.

4. **Step 4:** The fourth step conducts voting for contrast of saturation. Regions possessing highly different saturations in local and global context will add to the saliency of Ri.

5. **Step 5:** The fifth step collects votes for Ri from regions having highly different intensity (contrast of intensity).

6. **<u>Step 6:</u>** The warm colors consisting of the red, orange, and yellow ranges are given an extra vote to strengthen their color saliency in the sixth step.

7. **<u>Step 7</u>**: Finally, the seventh step supports saliency of highly bright and saturated colors.

5.3.2.2 Size Map

Although prominence due to size may be suppressed in presence of high contrast in other visual attributes such as color, saliency in terms of area can play a useful role in situations where a target of attention does not surface due to other features. The size based saliency mainly contributes in suppressing large sized background regions and unnoticeable small sized regions. The contrast of size with respect to the neighborhood needs to be computed to find objects having exclusive size. Figure 3.1 provides examples of two scenarios where the only obvious feature to determine saliency is the size of objects and such uniquely sized objects are the obvious attractors of attention. The exclusiveness of size with respect to the neighborhood and the global context is determined using a voting style mechanism. A given region Rj with a size similar to Ri will not contribute to the size saliency of Ri, one with significantly different size will give a partial supporting vote, and if such Rj surrounds Ri then the contribution is a full support because a situation similar to accent color is developed.

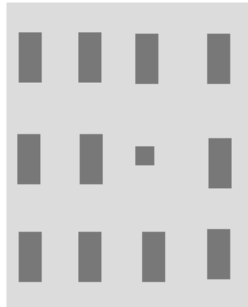


Figure 5.1 Size saliency due to uniqueness

5.3.2.3 Orientation, Eccentricity And Symmetry Map

5.3.2.3.1 Orientation, Eccentricity And Symmetry Calculation:

We use a traditional technique involving moments for finding the orientation and eccentricity of regions. Theses feature values are used to determine saliency with respect to these features. Three types of discrete two dimensional moments are computed for each Region as follows;

$$M^{i}_{1,1} = \sum (x-x')(y-y') \quad (5.3)$$

$$M^{i}_{2,0} = \sum (x-x')^2 \quad (5.4)$$

$$M^{i}_{0,2} = \sum (y-y')^2 \quad (5.5)$$

Here (x', y') is the center of the region. Now the orientation o^i and eccentricity e^i are computed as

$$o^i = 1/2 \tan^{-1} (2 M^i_{1,1}/(M^i_{2,0} - M^i_{0,2}))$$

$$e^i = ((M^i_{2,0} - M^i_{0,2})^2 + 4 M^i_{1,1})/ (M^i_{2,0} + M^i_{0,2})^2$$

5.3.2.3.2 Map Construction

After having computed the values of feature magnitudes for the rest of the three features, the remaining task in building the saliency maps is to determine exclusiveness with respect to these features in local as well as global context. As the prominence of a region mainly arises from the rarity of its feature values [Ner04], hence the saliency of an object with respect to a particular feature decays strongly when another object with a similar value of that feature exists is the near surrounding. A weaker decline of saliency occurs in case of similarity existing outside a certain radius.

5.3.3 Visual Behaviors

The focus of attention is based purely on the visual behavior. There are three behaviors available search, explore and detect changes. So the focus of attention while searching a predefined object will be entirely different than from those under free viewing. In order to make our model biological, we have to integrate these behaviors while finding the focus of attention.

We implement the influence of active visual behavior using a set of behavior-dependant weights for each individual behavior in which every set member is associated to a specific feature channel. We are using different behavior dependent weights. We use four categories, namely inactive, low, medium, and high for this purpose.

For an inactive the channel becomes totally inactive and hence is non-contributing. Therefore its value is zero. We assign low= 1 meaning that no extra emphasis will be given on the related channel when computing the resultant saliency. The value medium is taken equal to the sum of all low weights at that particular time in order to make influence of the involved channel higher than the others. If some channel has to be given higher weight in presence of a channel with a medium importance then the category high will be used in which the weight will be again sum of all weights counting the involved channel as a medium one. Hence, for our model with five channels in hand, we can have medium= 5 and high= 13.

5.3.3.1 Search Behavior

For visual behavior of search top-down methodology is used. In other words it is a task dependent behavior. In this top-down pathway the color channel is given more emphasis over the other four channels because color is the most stable feature while searching an object whereas other features like size, eccentricity, symmetry, and orientation can significantly vary when a vision system goes around in the environment or the objects themselves rotate and translate in three dimensions. Therefore color will have a medium weight while other channels will be given low weights under search behavior.

5.3.3.2 Explore Behavior

When the system works under explore behavior i.e. task independent, the bottom-up methodology is used. It sets the top-down channel inactive. The orientation channel is reported as one of the confirmed feature channels and the other shape-based features like eccentricity and symmetry are listed as probable channels in human vision [WH04], hence we assign them medium weights.

5.4 Chapter Summary:

The prominent contributions of the work presented in this chapter are the innovations in the computation schemes of feature maps for color contrast and symmetry, proposal for determining size contrast as a formal feature map, and inclusion of eccentricity map together with orientation map as feature channels in the process of visual attention. Hence, efficient algorithms for construction of five feature maps are proposed that are able to be integrated into a region-based model of visual attention and, in turn, into other intelligent vision systems.

CHAPTER 6

TESTING AND RESULTS

6.1 Testing

Software Testing is the process of executing a program or system with the intent of finding errors. It involves any activity aimed at evaluating an attribute or capability of a program or system and determining that it meets its required results. Software is not unlike other physical processes where inputs are received and outputs are produced. Where software differs is in the manner in which it fails. Most physical systems fail in a fixed (and reasonably small) set of ways. By contrast, software can fail in many bizarre ways. Detecting all of the different failure modes for software is generally infeasible.

6.1.1 Types of Testing

Following techniques have been followed for testing;

Black box testing

It is the type of testing in which the system is tested on inputs, expected results and system outputs without going into internal functionality of system.

Unit Testing

Unit testing is a method by which individual units of source code are tested to determine if they are fit for use i.e. they adhere to all the functional requirements. It is performed to ensure that each module of this system is working fine. First of all unit testing is performed on the segmentation module. There is no quantitative way of checking whether segmented image is correct. The results are compared with other algorithms using the same segmentation technique.

Unit testing was performed for the feature magnitudes module. The results were compared with the calculations done manually.

Unit testing was done for bottom-up and top-down modules. For the given input, the output was checked with the results of human beings for that input image.

Integration Testing

Integration testing is the phase of software testing in which individual software modules are combined abs tested as a group. Integration testing was performed by checking that the compatibility of each module with each other.

System Testing:

System testing is the phase of software testing in which overall system is tested that is adheres to the functional requirement.

White box testing techniques

The white box testing technique deals with the internal logic and structure of the code. The test written based on the white box testing strategy in cooperate coverage of the code write, branches, paths, statements and internal logic of the code. Insertion of print statements was done to find the results and according to manually expected results, moreover we checked each branch of loops in detail.

6.2 RESULTS

The proposed attention model is implemented as a complete software using object oriented programming techniques with C++. This chapter presents the output generated by the model on various test cases selected to judge its performance. The experiments with the model were carried out for three attentional behaviors, namely explore, search, and examine, for which details are described in sections 6.3, 6.4 and 6.5 respectively.

6.2.1 Experimentation Platforms

Experiments were performed to test the capabilities of the proposed model for different visual behaviors using three experimentation platforms. The first platform is an evaluation framework for single images. Images can be loaded as input for the model and a behavior can be activated through the options provided in a graphical user interface. Using this software the saliency maps can be seen at each attempt of attention and the fixated regions get marked by prominently colored rectangles. Figure 6.1 shows a screenshot of the graphical user interface of this system in which saliency maps and fixated locations for a sample image are also visible.

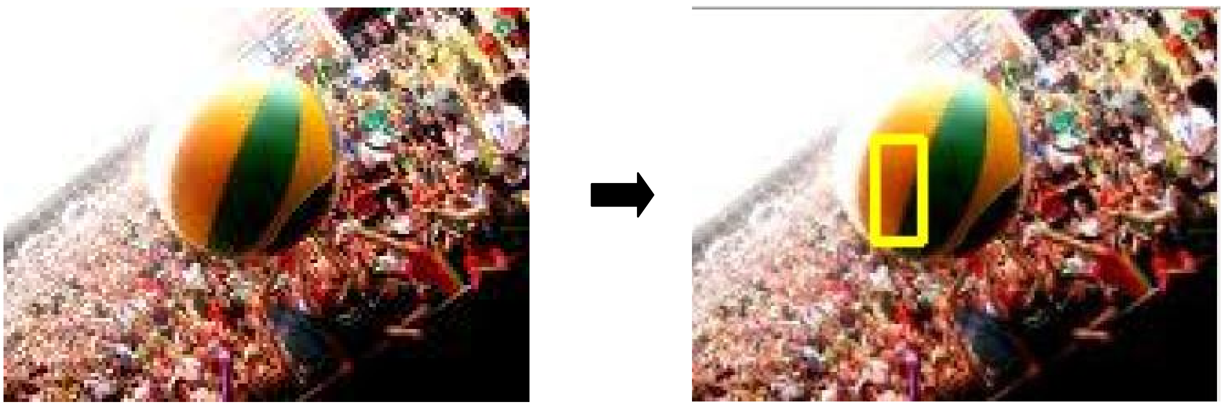
Figure 6.1 Graphical user interface of the implementation of the proposed attention model. Processing on a sample input is shown with the intermediate results.

6.2.2 Exploration Results

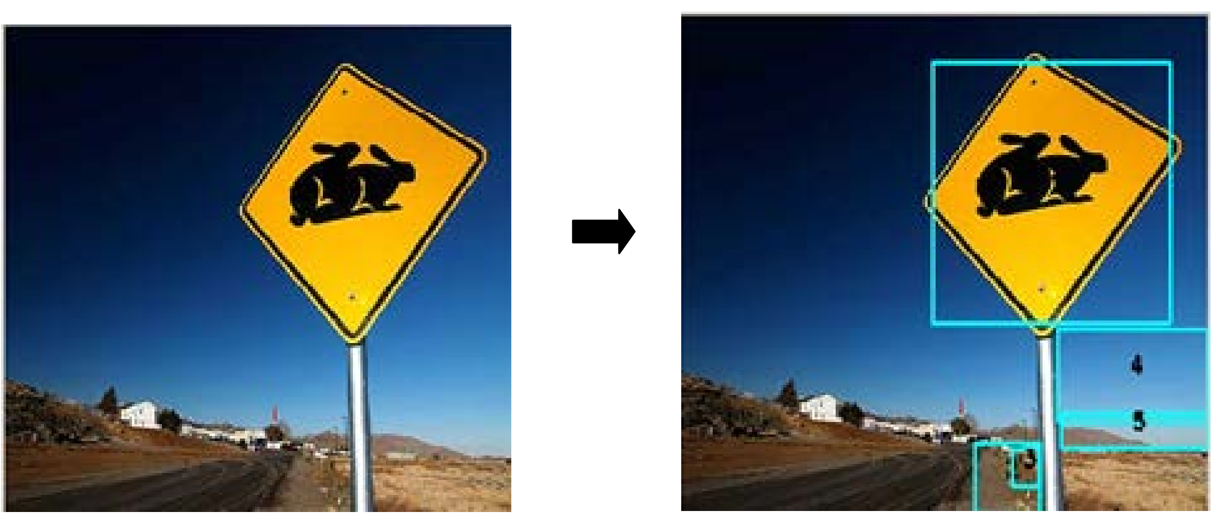
Under the explore visual behaviour the vision system performs free viewing with no top-down task given to it. The requirement is to identify those locations in the given scene that would be salient for the natural/human vision. The developed attention model was tested with static images to verify its ability to locate salient regions and then its capability to work in dynamic environments was tested using videos.

A variety of visual scenes including snapshots taken from camera of robot head, synthetic environment of the simulation framework, and other images collected from image databases on internet were used to experiment with the proposed attention model. Figure 6.2 presents results of attention on images having conjunction of different features in synthetic and natural scenes.

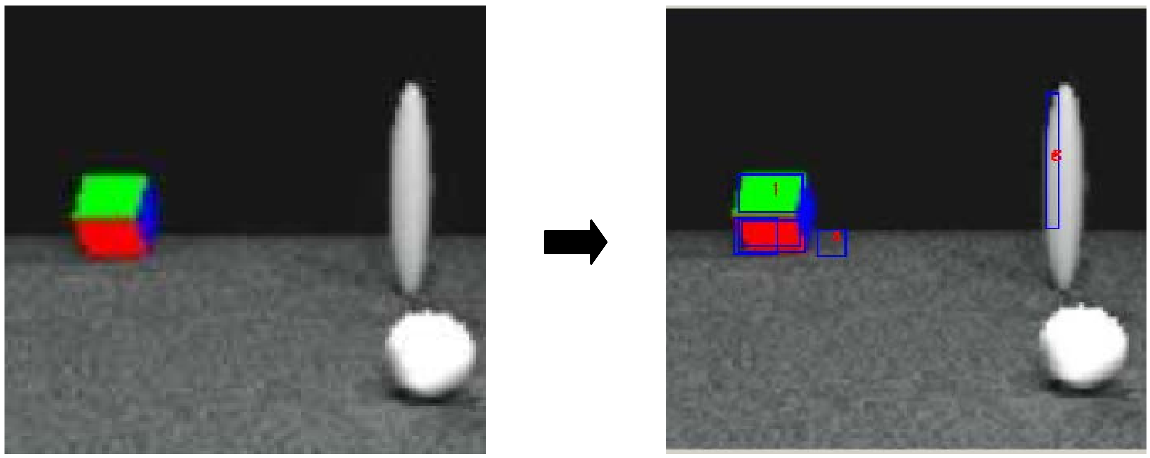
(a) shows the image of a football and in the background there is crowd. Definitely this football will grab your attention. The input image in part (b) is a traffic scene in which the traffic signs offer a high bottom-up saliency. (c) offers rendered 3D objects having saliency with respect to different features with a very simple background.



(a)



(b)



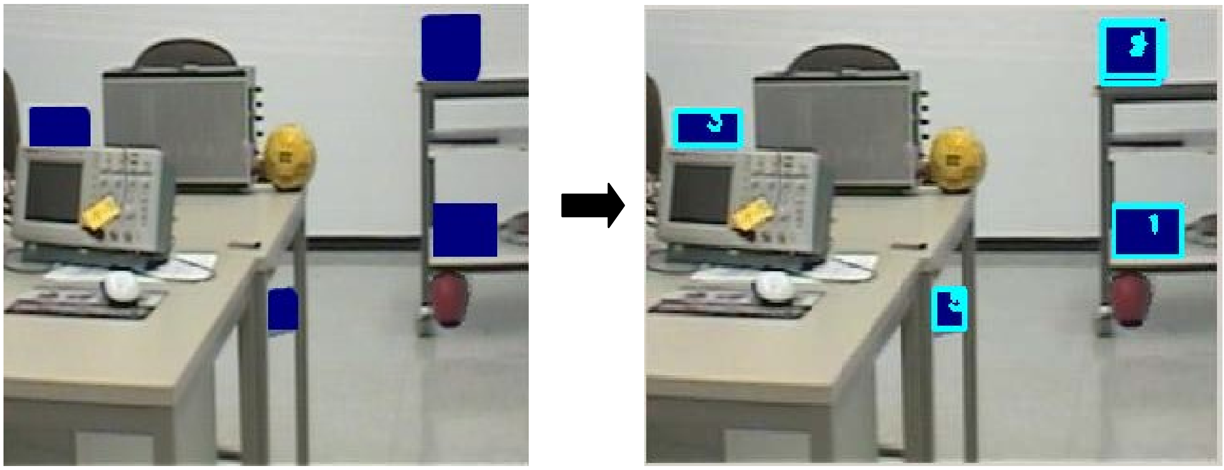
(c)

Figure 6.2 shows the input image used for Explore results and their corresponding results

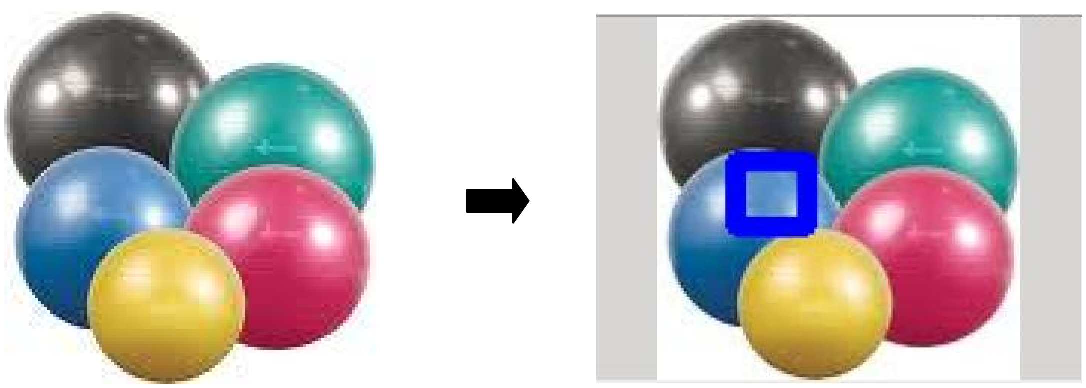
6.2.3 <u>Search Results</u>

For experiments to test functioning of the proposed model under the visual behaviour of search, description of the objects to be searched is given to the attention model in form of images containing the isolated targets over a blank background. Such descriptions of the target may be considered as the top-down conditions for the attention mechanism. In the current status, the system is able to work with single regions at a time rather than composite objects hence the system picks the largest foreground region from the given image of the search target.

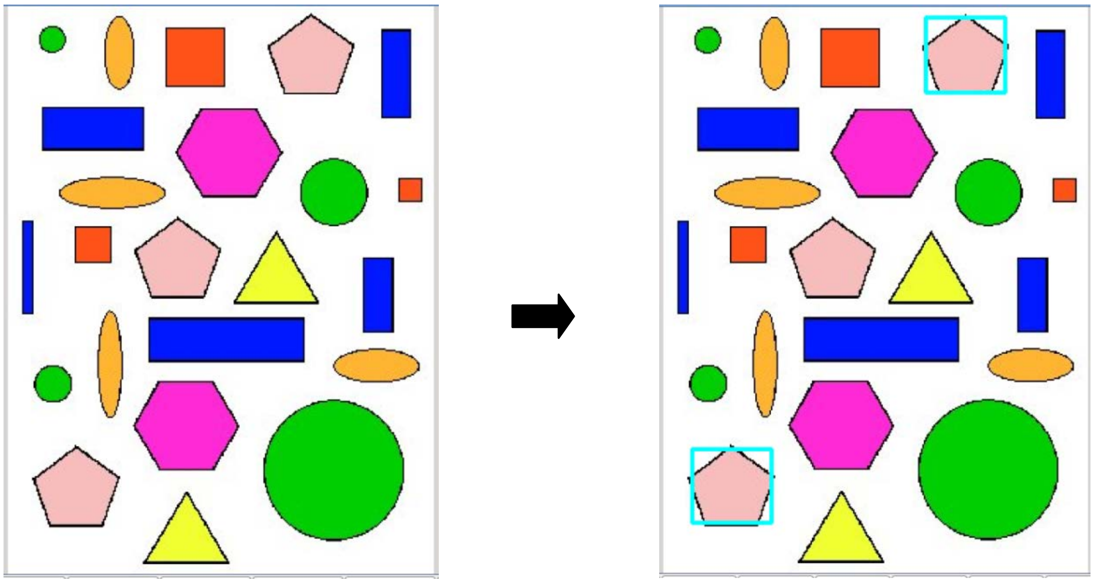
The first scenario of experiments was the search in static scenes in which the attention mechanism was allowed to mark as many occurrences of the target as possible. These experiments tested the ability of the system to select all relevant locations. (a) reflects this scenario with the search field as a still scene having four occurrences of the target (a dull blue box with some texture) in the scene. Results of the first five fixations (t = 1 to t = 5) by the attention system are reported. (b) subfigure shows the results if it was asked to find the blue ball. (c) shows the results when pink pentagon was selected.



(a)



(b)



(c)

Figure 6.3 shows the input images for search behavior and their corresponding results

6.2.4 Track/Examine

Tracking behavior is also incorporated in BIDSO. Tracking is different from Search in a way that in search the position of the target object is unknown in the scene while in tracking the position of the target object is known right now and it needs to be tracked in the next frames. This tracking behavior is only visible in videos.

6.3 Summary

This chapter has presented results obtained from the attention model. The software to work with static images allows viewing the results along with the intermediate processing being carried out. The results for visual behaviors of explore, search, and examine has been obtained on this platform. The output under all scenarios is promising and shows success of the proposed methodology in advancing the state of the art in this area of research. A formal evaluation of the results and comparison with other existing attention models is performed in the next chapter.

CHAPTER 7

RESULT ANALYSIS

7.1 Introduction:

This chapter presents the results of the system for evaluating its performance. Currently, a standardized set of images for testing is not available; self created samples are used to verify the ability of the model to identify saliency with respect to different visual features. Output of the proposed model is evaluated using experiments related to bottom-up as well as top-down aspects of attention. Output of the proposed model is compared with the attention models of [IKN98], [AL06], and [BMB01] in order to quantify the achievements gained through the model. The software of the models of [IKN98], [BMB01] and [AL06] was obtained from their websites. Secondly, the region-growing segmentation algorithm has also been tested with other available segmentation algorithms.

7.2 Validation Of Results:

In order to verify the ability of the proposed model to determine saliency with respect to the individual features, input images each containing salient objects in context of only one feature were used as input for the model. Figure 6.1 shows the input images with respect to feature namely color, size, eccentricity, orientation, symmetry respectively.

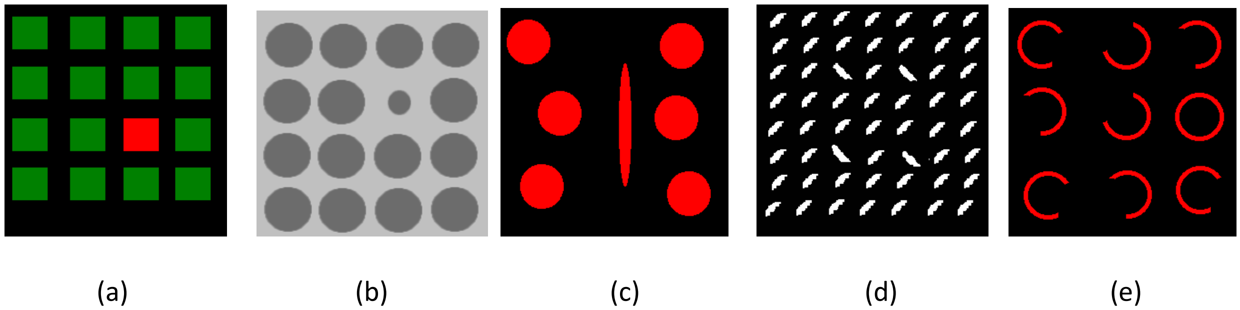


Figure 7.1 Shows the input images (a) color (b) size (c) eccentricity (d) orientation (e) symmetry

Figure 6.2 displays the corresponding saliency maps and the second row shows the foci of attention. It can be clearly seen that the outstanding object due to each feature was marked by the system hence the system's response to individual features is valid.

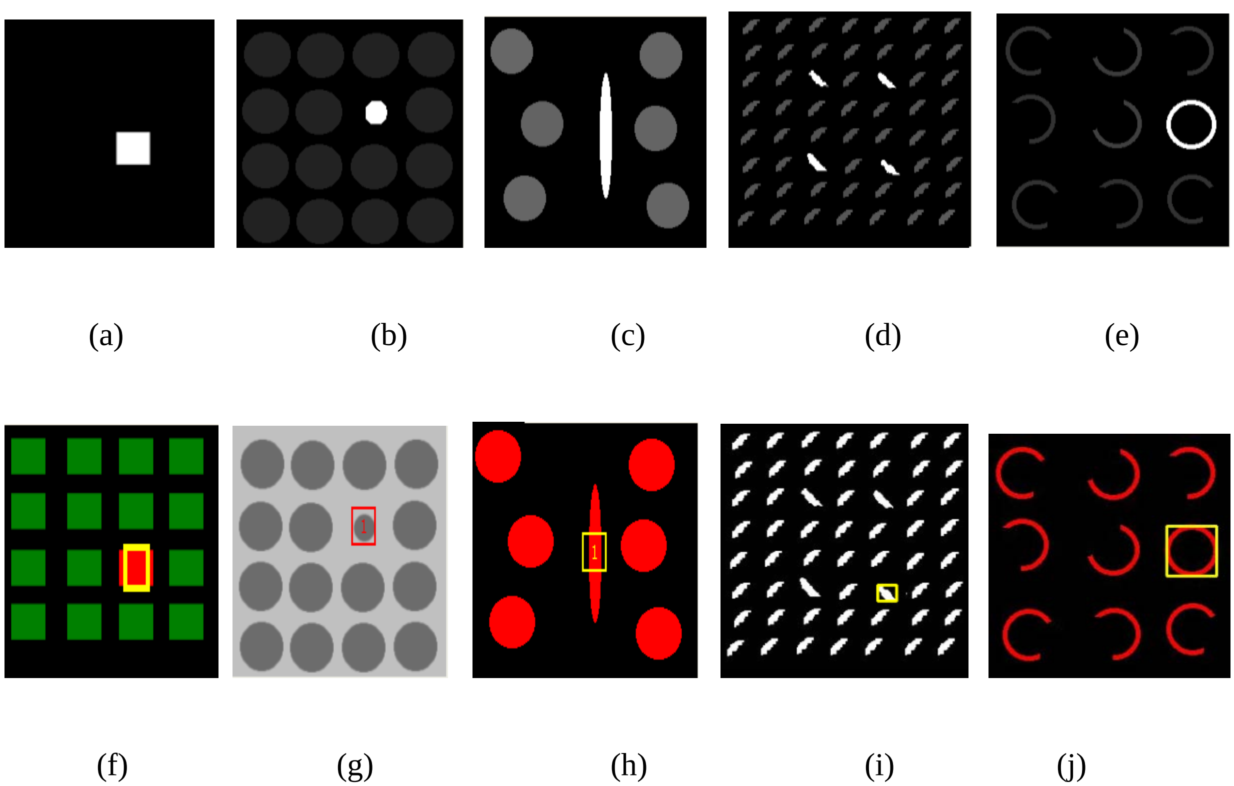
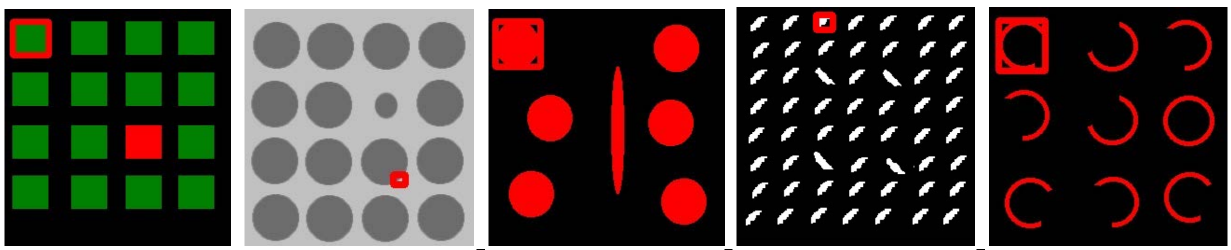


Figure 7.2 Row 1 shows the corresponding saliency maps, Row 2 shows the focus of attention

7.2.1 Comparison With [AL06]:

Figure 6.3 shows the output of the samples taken as input. The model proposed in [AL06] does not perform well in the benchmark images because it concentrates mainly on global rarity while ignoring local feature-based saliency.

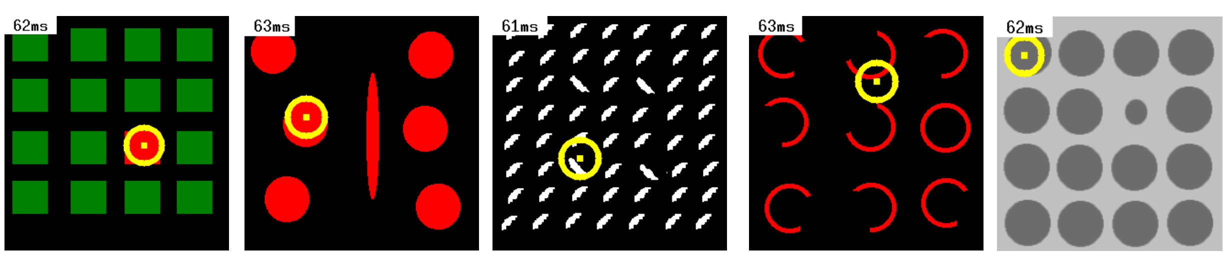


(a) (b) (c) (d) (e)

Figure7.3 Shows the focus of attention for algorithm [AL06]

7.2.2 Comparison With [IKN98]:

The model of [IKN98] does not compute the feature channels of eccentricity, symmetry, and size hence it was unable to pick the correct objects from the input samples given in figures 6.4 (b), (d), and (e).

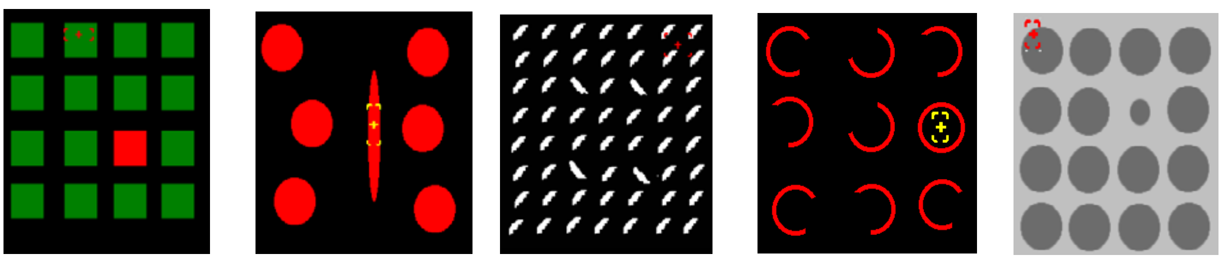


(a) (b) (c) (d) (e)

Figure 7.4 Shows the focus of attention for algorithm [IKN98]

7.2.3 Comparison With [BMB01]:

It is observable that the model of [BMB01] is not successful in identifying color saliency when the regions are separated by some other regions (black background between the colored boxes in figure 6.5(a)) because it computes contrast on the region edges only and ignores the global context. The contrasts due to orientation and size are also not considered in this model, hence, it could not identify the saliency correctly for the image given in figures 6.5(c) and (e)



(a) (b) (c) (d) (e)

Figure 7.5 Shows the focus of attention for algorithm [BMB01]

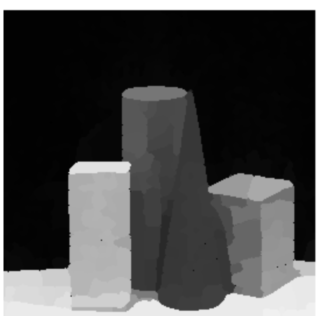
7.2.4 Segmentation Comparison with scale space method [DN04]:

The approach used of color segmentation was tested using many artificial and real life images. The results are very encouraging and the segmentation output was found suitable according to the requirements of the region-based attention model. We have also compared our results with some existing segmentation methods that use computationally heavy statistical methods and produced fairly good results for general-purpose segmentation. A qualitative comparison can be done by observing these results.

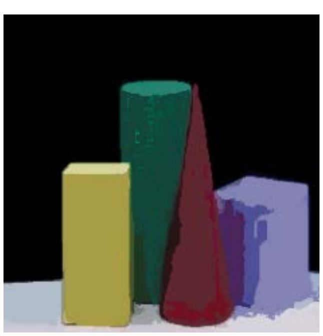
On the other hand the scale space method [DN04] handles these situations in a better way but it is over segmenting in chromatic regions. Both of the competitive methods are unable to separate the yellow colored melon overlapping the similarly colored banana in the fruit image.



(a) (b) (c)



(d) (e) (f)



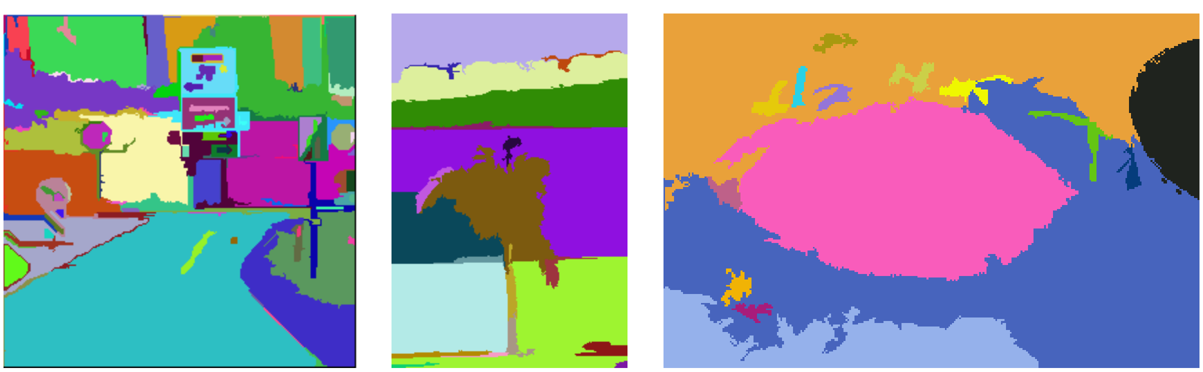
(g) (h) (i)

Figure 7.6 Top row: input images. Second row: results of a scale space method [DN04]. Bottom row: results of the method used.

7.2.5 Segmentation Comparison with graph-based method[FH04]:

The graph-based method [FH04] has performed very well with the chromatic colors but has flaws in the achromatic areas. For example, it splits the uniform black background of the image in the second column into many regions while it merges the white border line of the road into the gray road in the traffic scene.

(a) (b) (c)



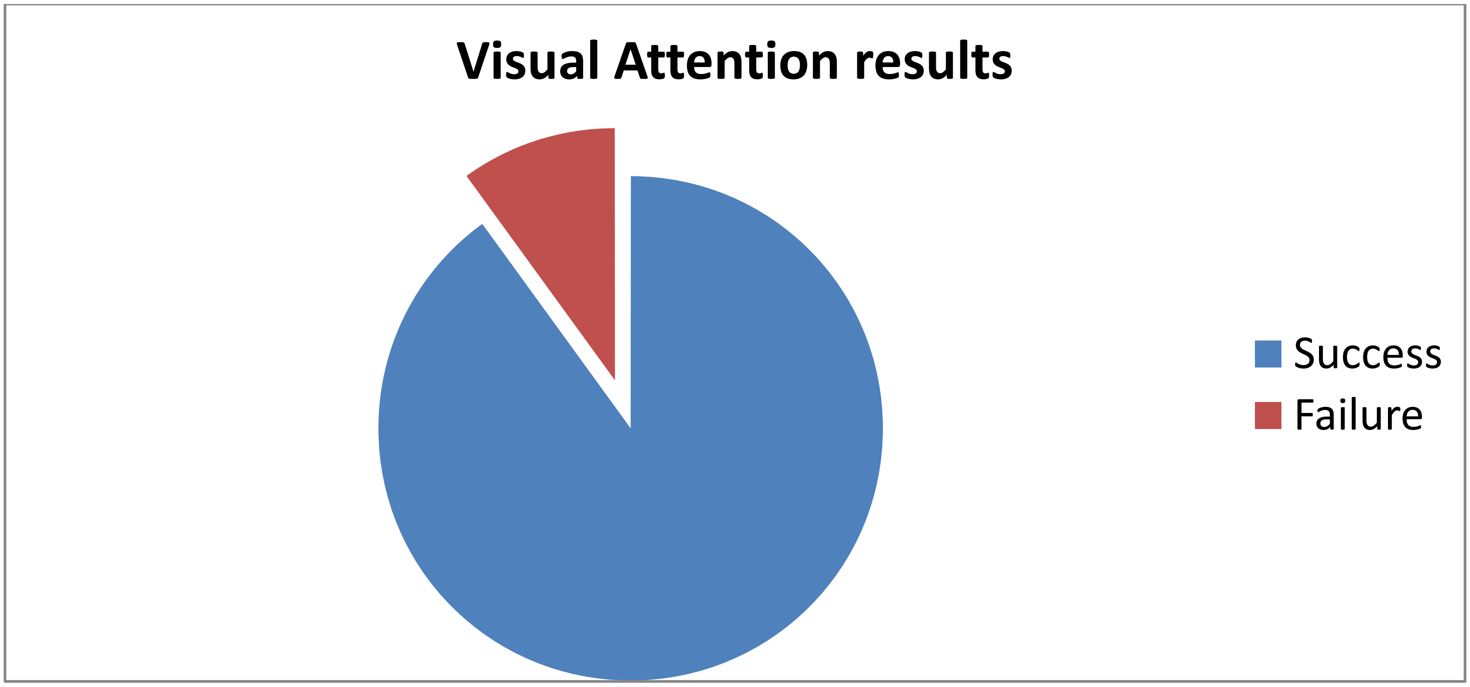
(d) (e) (f)

(g) (h) (i)

Figure 7.7 Top row: input images. Second row: results of a graph-based method [FH04]. Bottom row: results of the method used.

7.2.6 Validation by Human Subjects

51 human subjects were considered at random with data sample of 30 images out of which 41 people focus of attention matched to the results of our system, 9 people read heading rather than visual artifacts.



7.3 **Chapter Summary:**

In this chapter testing has been done and the results are compared with the algorithms for which software are publically available. It is shown that the model used has given valid results and are better than the other algorithms available. The ability of the proposed method in performing well both for chromatic and achromatic color can be very useful in situations when the mobile vision system goes through low light areas where color distinction based upon hue becomes difficult.

CHAPTER 8

SIMULATION

8.1 Introduction

A simulation has been created to validate the results. In this project real robot is simulated using Player and Gazebo. In this chapter, basic introduction of the tools used and the results are discussed.

8.2 Player:

Player is a device server that provides a powerful, flexible interface to a variety of sensors and actuators (e.g., robots). Because Player uses a TCP socket-based client/server model, robot control programs can be written in programming language and can execute on any computer with network connectivity to the robot. In addition, Player supports multiple concurrent client connections to devices, creating new possibilities for distributed and collaborative sensing and control.

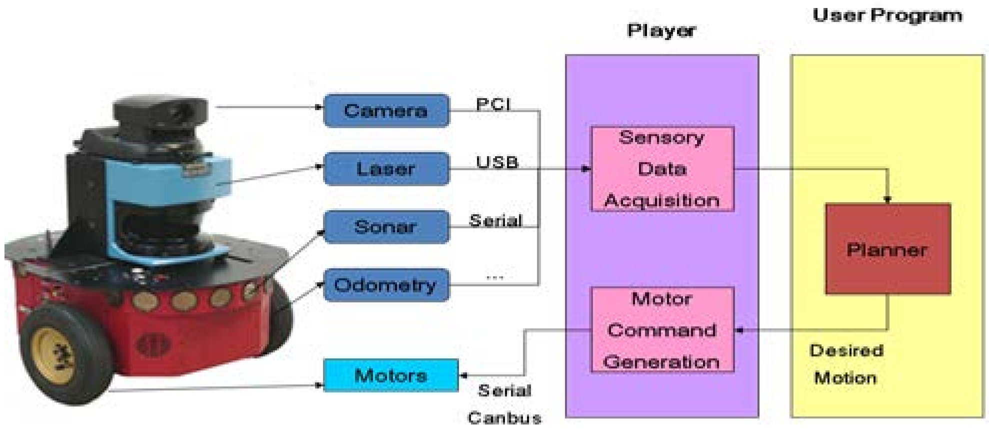


Figure 8.1 Shows the structure of player

8.3 Stage

Stage simulates a population of mobile robots, sensors and objects in a two-dimensional bitmapped environment. Stage is designed to support research into multi-agent autonomous systems, so it provides fairly simple, computationally cheap models of lots of devices rather than attempting to emulate any device with great fidelity. We have found this to be a useful approach.

8.4 Gazebo:

Gazebo is a simulator for a small group of robots in a 3D environment. Similarly to Stage, a 2D environment simulator, Gazebo can simulate a population of robots, objects and sensors. There are a few differences between the two simulators. **Gazebo** is designed for a small number of robots while Stage can handle hundreds of robots. In addition, Gazebo has a higher precision then Stage. Since both Gazebo and Stage are Player compatible, client programs written using one simulator can usually be run on the other one with some or no modifications.

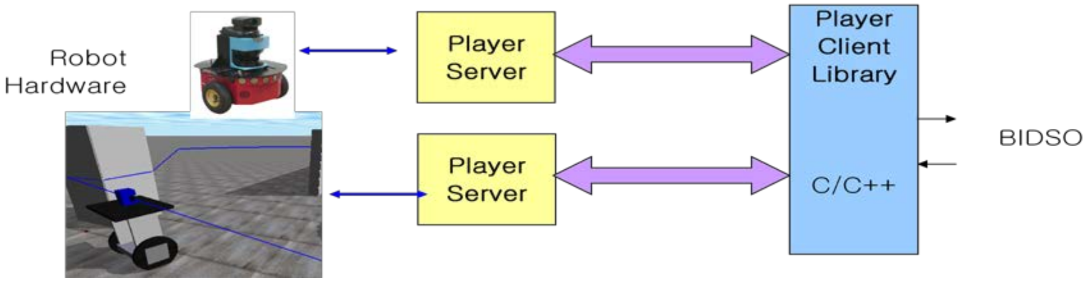


Figure 8.2 Shows the structure of Gazebo

8.5 <u>Results</u>:

<u>Gazebo World File:</u>

This is the environment file that is built in gazebo. Robot has a camera. It captures the scene with its camera and this scene in the form of image goes to the client side that is the user program.

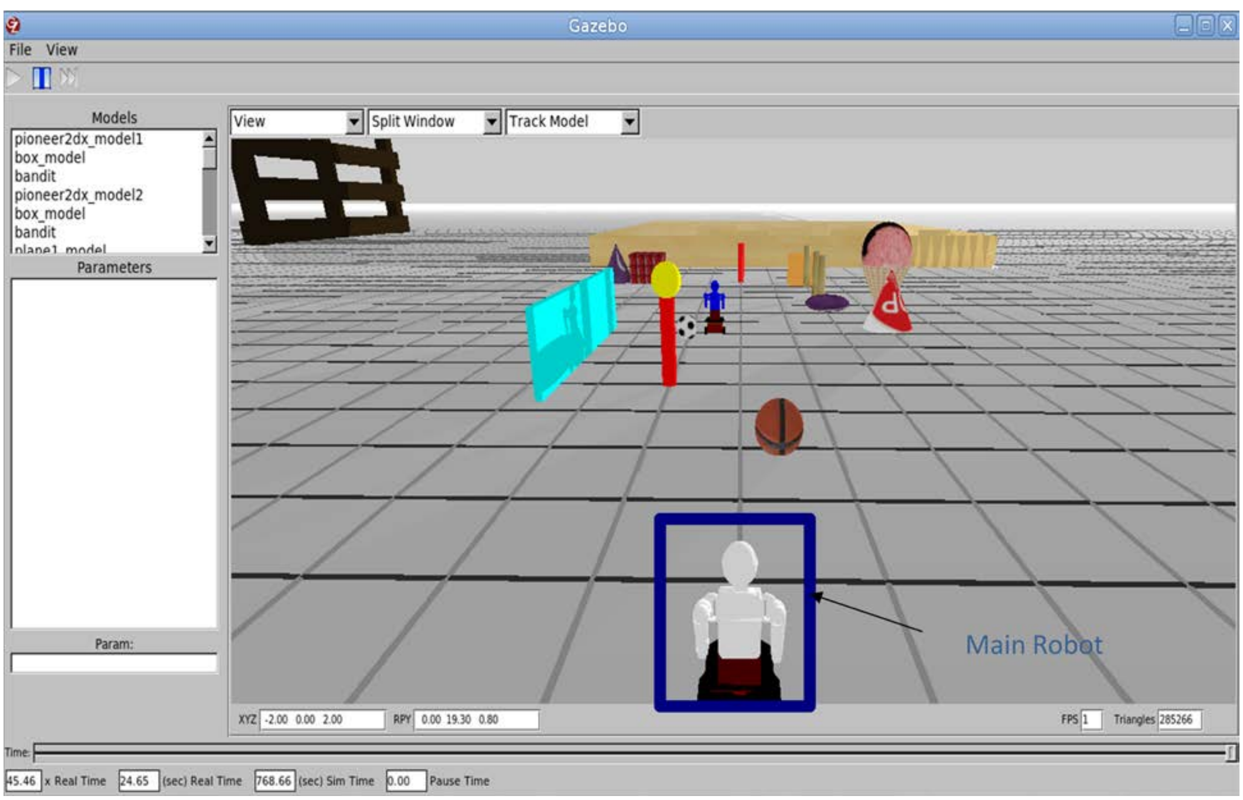


Figure 8.3 shows the environment in gazebo

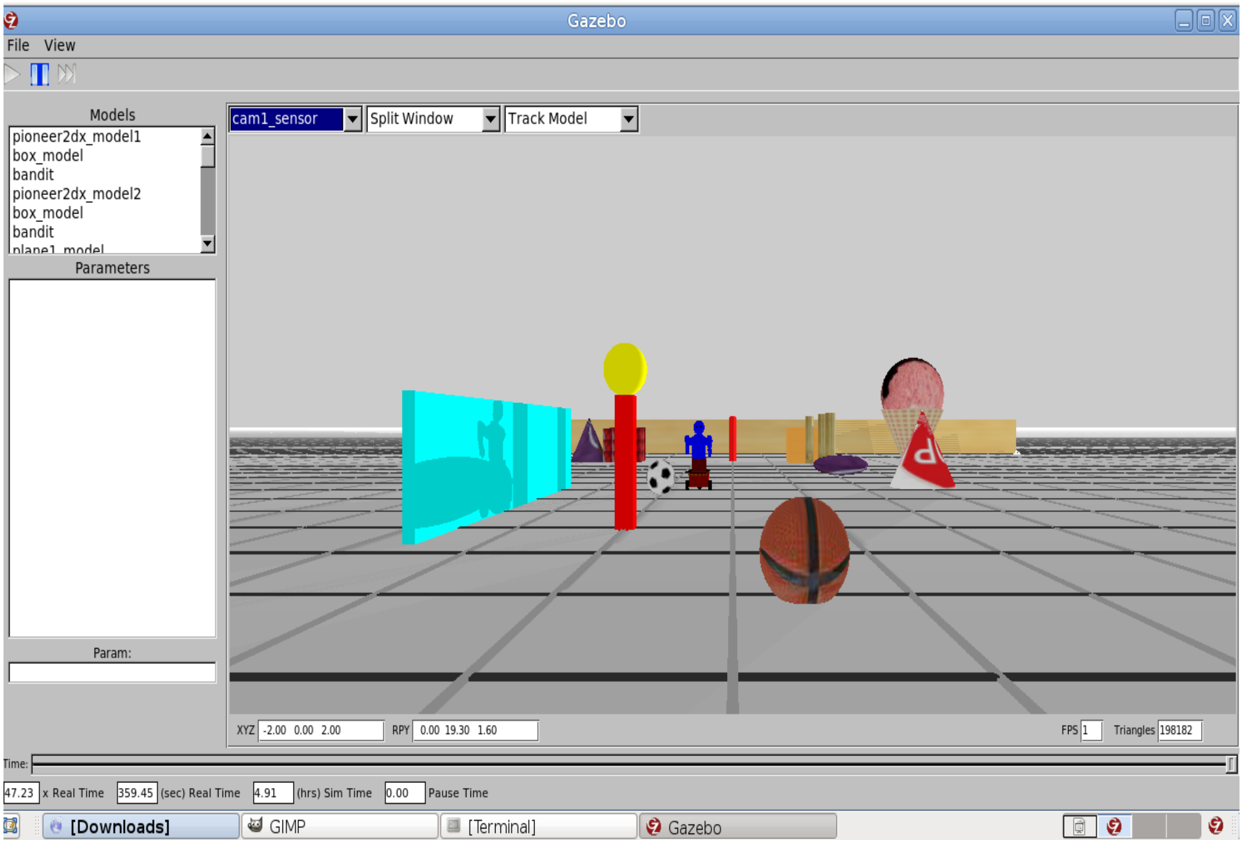


Figure 8.3.1 shows the camera view of robot in gazebo

<u>BIDSO and Player integration with Gazebo</u>

The code is integrated with Gazebo using player. First of all the configuration file (.cfg) is executed. The drivers of the simulation are provided with the ports to transfer data from gazebo to the user program.

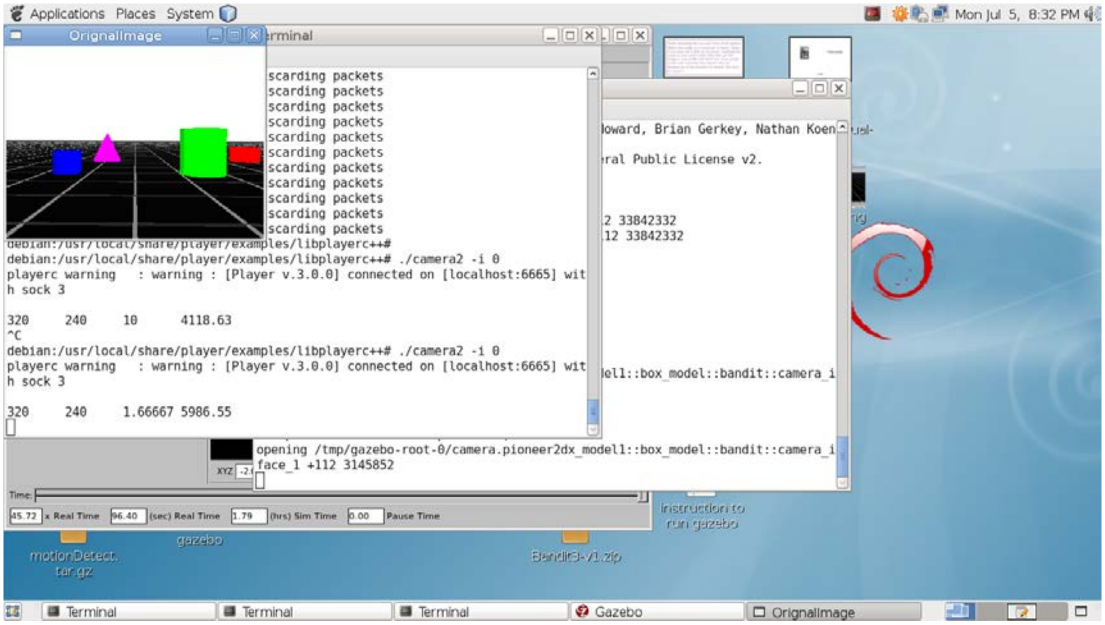


Figure 8.4 Showing integration of code with gazebo

Bottom-up/Explore

The input to the code is the image taken from environment in gazebo. The result of bottom-up/Explore is shown in the figure 8.5

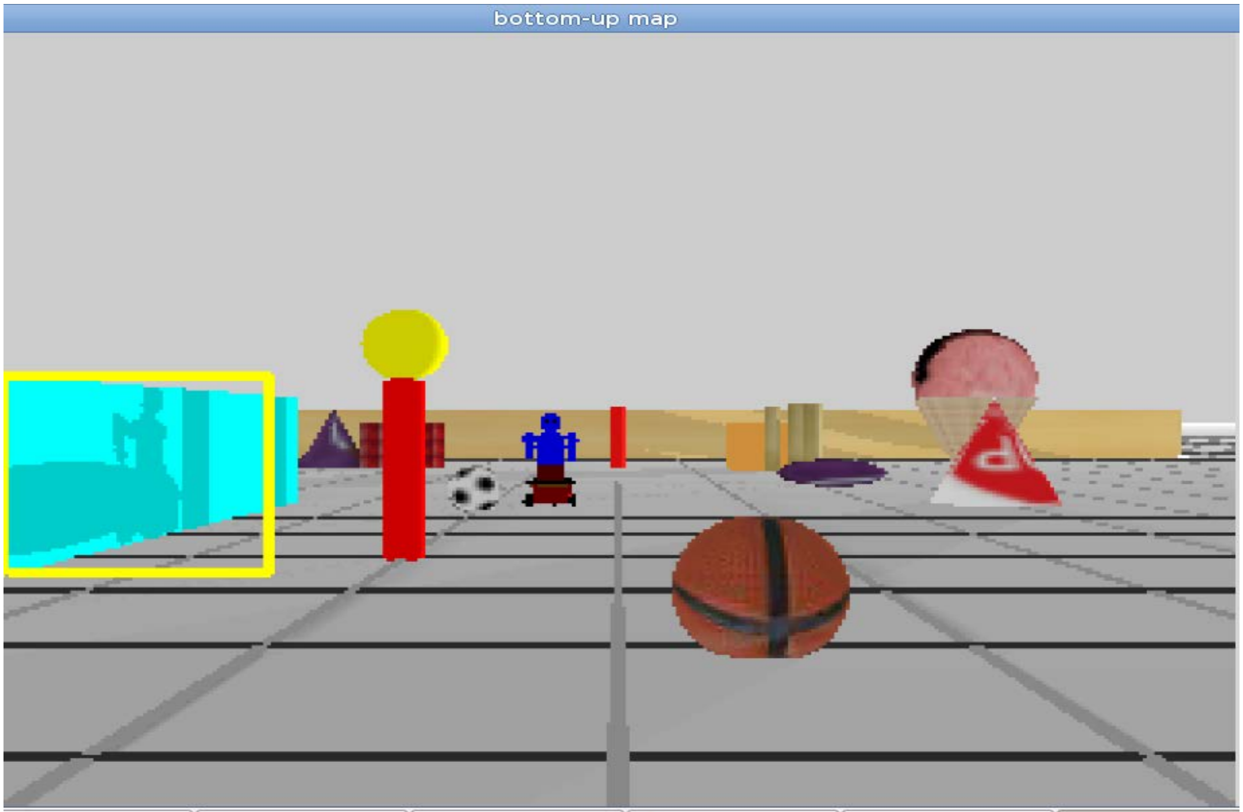
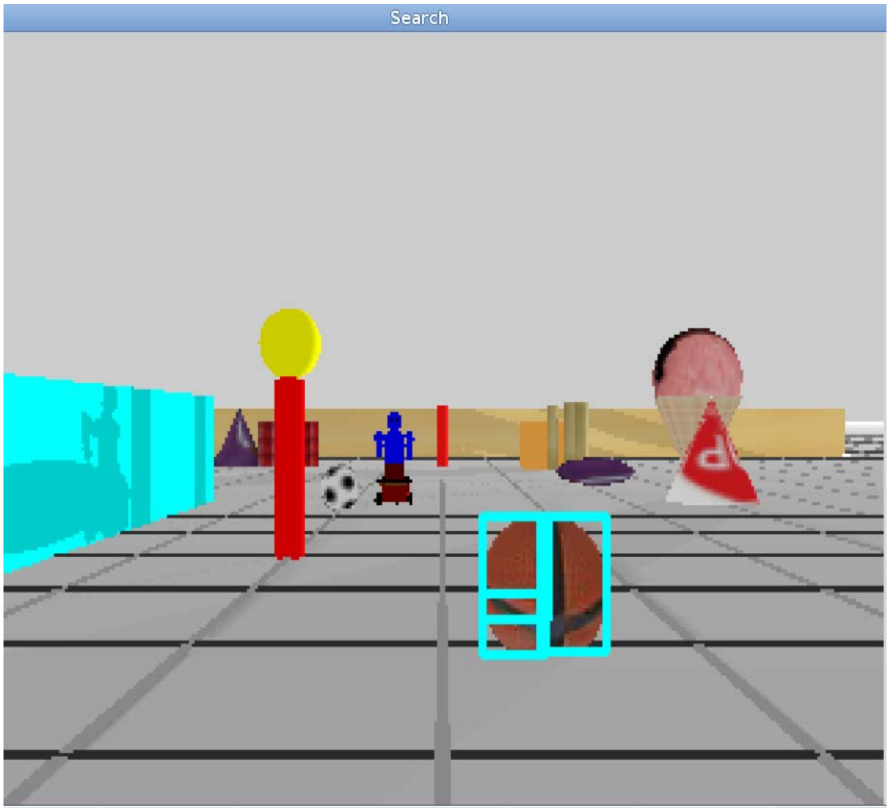


Figure 8.5 Showing results for Bottom-up/ Explore

Top-Down map when a cone is selected

The input to the code is the image taken from environment in gazebo. The result of top-down/Search is shown in the figure 8.6 User selects the basket ball from the right image is the screenshot. And the left image shows the result.



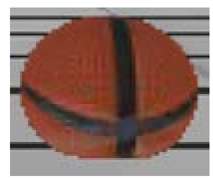


Figure 8.6 Showing results for Top-Down/Search **Search Target**

8.6 Chapter Summary

This chapter has presented results obtained from the attention model after running it on three different experimentation platforms. The software to work with static images allows viewing the results along with the intermediate processing being carried out. The results for visual behaviors of explore, search, and examine has been obtained on this platform. Working of attention in explore and search behaviors has been experimented using the simulation framework.

CHAPTER 9

CONCLUSION AND FUTURE WORK

9.1 Conclusion

BIDSO is a desktop application that does all processing in run time. Our aim was to develop such a system that can be used as a building block for the real robots. The main objective achieved is to know the important regions with respect to human vision system in an image/video by using fast algorithm of segmentation, considering five feature channels and using both top-down and bottom-up techniques. These many channels have never been considered before in any of the attention channels. Methods for applying inhibition of return and determining pop-out in the region-based paradigm were also specially developed. The model integrates the bottom-up and top-down pathways in a single architecture which was also not done before by any other model.

BIDSO can explore the salient object, search a given object and track an object movement. It can be used for applications like Compression, Attractive Ads/ Banners, Medical Images. Like humans BIDSO focus only on salient regions in an image/video so it increases efficiency, decrease computational effort, is less expensive and faster.

9.2 Discussion

We have done early segmentation which helps in handling the global contrast, helping in finding shape based features, therefore increasing the efficiency. It lags behind the pixel based approaches in context of robustness because our model completely depends on the segmentation and segmentation is completely effected by noise and distortion.

We have used region growing and it has more explorative capability as compared to pixel based, because units of processing are bigger clusters in the region-based approaches that allow computation of contrast at a broader (or global) level. On the other hand pixel based approaches may remain on close vicinity because unit of processing is a single pixel or a small cluster of pixels.

The performance of our model is better than other models in many aspects but there is a limitation that output purely depends on the segmentation of the given input and the region based attention lag behind in robustness against distortions and noise in the input images/videos.

9.3 Outlook

In our model we have proposed early clustering of pixels. Here the main question is the format of clustering because the region segmentation is a solution for computer processing of images and details of clustering in human vision are not clearly understood yet. A study of clustering mechanism in human vision will be a contribution to this field. Involvement of knowledge in the attention processing may also lead to obtaining saliency in terms contrast of known and unknown objects. Computation of other features such as texture, and depth in case of stereo images can also be an included.

9.4 Future Work

The main aim of our project is to design and develop a software system that takes dynamic and static scenes as an input and extract the salient object based on human visual attention system. We have achieved our main goal. Future work in this project can further make it more applicable in real time situations. Further work can be done on the attention mechanism that it should be implemented in autonomous robotic machines that should be able to perform all required visual behaviors without human intervention. The system should be able to switch between behaviors autonomously depending upon the situation and requirements of the active task. In real-world scenarios, attention is performed in a three dimensional space. An improvement in the current state of the art will be to determine saliency in 3D rather than two dimensional maps.

BIBLOGRAPHY

[Ahu96] Ahuja, N.: A Transform for Multiscale Image Segmentation by Integrated Edge and Region Detection. In: Transactions on Pattern Analysis and Machine Intelligence 18 (1996), pp. 1211-1235

[AL06] Avraham, T.; Lindenbaum, M.: Esaliency - A Stochastic Attention Model Incorporating Similarity Information and Knowledge-Based Preferences. In: WRUPKV-ECCV 06. Graz, 2006

[AM07b] Aziz, M. Z.; Mertsching, B.: Color Saliency and Inhibition in Region Based Visual Attention. In: WAPCV 2007. Hyderabad - India, 2007, pp. 95-108

[CJSW01] Cheng, H. D.; Jiang, X. H.; Sun, Y.; Wang, J.: Color image segmentation: Advances and Prospects. In: Pattern Recognition 34 (2001), pp. 2259-2281

[CL94] Carron, T.; Lambert, P.: Color Edge Detector Using Jointly Hue, Saturation, and Intensity. In: IEEE Conference on Image Processing. Austin, USA, 1994, pp. 977-1081

[DC04] Deng, H.; Clausi, D. A.: Unsupervised Image Segmentation Using a Simple MRF Model with a New Implementation Scheme. In: Pattern Recognition 37 (2004), pp. 2323-2335

[FH04] Felzenszwalb, P. F.; Huttenlocher, D. P.: E_cient Graph- Based Image Segmentation. In: International Journal of Computer Vision 59 (2004), pp. 167-181

[FBR05] Frintrop, S.; Backer, G.; Rome, E.: Goal-Directed Search with a Top-Down Modulated Computational Attention System. In: DAGM 2005, LNCS 3663, 2005, pp. 117-124

[Ham05] Hamker, F. H.: Modeling Attention: From Computational Neuroscience to Computer Vision. In: WAPCV 2004, LNCS 3368, 2005, pp. 118-132

[HB90] Haddon, J. F.; Boyce, J. F.: Image Segmentation by Unifying Region and Boundary Information. In: Transactions on Pattern Analysis and Machine Intelligence 12 (1990), pp. 929-948

[IK00] Itti, L.; Koch, C.: A Saliency Based Search Mechanism for Overt and Covert Shifts of Visual Attention. In: Vision Research (2000), pp. 1489-1506

[IKN98] Itti, L.; Koch, U.; Niebur, E.: A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. In: Transactions on Pattern Analysis and Machine Intelligence 20 (1998), pp. 1254-1259

[IPV00] Ikonomakis, N.; Plataniotis, K. N.; Venetsanopoulos, A. N.: Color Image Segmentation for Mutimedia Applications. In: Journal of Intelligent and Robotic Systems 28 (2000), pp. 5-20

[LD04] Lanyon, L.; Denham, S.: A Model of Object-Based Attention That Guides Active Visual Search to Behaviourally Relevant Locations. In: WAPCV 2004, LNCS 3368, 2004, pp. 42-56

[LHG97] Laar, P.V.D.; Heskes, T.; Gielen, S.: Task-Dependent Learning of Attention. In: Neural Networks 10 (1997), pp. 981-992

[LJBV05] Lopez, F.; J.M.Valiente; Baldrich, R.; Vanrell, M.: Fast Surface Grading Using Color Statistics in the CIE Lab Space. In: IbPRIA 2005, LNCS 3523 (2005), pp. 666-673

[MCBT06] Meur, O. L.; Callet, P. L.; Barba, D.; Thoreau, D.: A Coherent Computational Approach to Model Bottom-Up Visual Attention. In: Transactions on Pattern Analysis and Machine Intelligence 28 (2006), pp. 802-817

[Ner04] Neri, P.: Attentional Effects on Sensory Tuning for Single-Feature Detection and Double-Feature Conjunction. In: Vision Research (2004), pp. 3053-3064

[OPR78] Ohlander, R.; Price, K.; Reddy, D. R.: Picture Segmentation Using Recursive Region Splitting Method. In: Computer Graphics and Image Processing 8 (1978), pp. 313-333

[PSL02] Park, S. J.; Shin, J. K.; Lee, M.: Biologically Inspired Saliency Map Model for Bottom-up Visual Attention. In: BMCV 02, LNCS 2525, 2002, pp. 418-426

[SK94] Sharbek, W.; Koshan, A.: Color Image Segmentation: A survey, Technical Report 94 - 32. Berlin - Germany : Fachbereich 13 Informatik, Technical University Berlin, October 1994