# Chapter 1

# Introduction

## 1.1 Overview

Intelligent robots have begun to appear in a variety of forms, including humanoid, to help with complex tasks such as cleaning, cooking, health care, surveillance or simply in soccer competitions trying to kick a ball. But why build a robot in anthropomorphic form?. Most of the environments are naturally designed to accommodate humans, and humanoid robots can effectively exploit the existing infrastructure. Any activity that can be performed by a person, such as climbing a ladder and squeezing through a man hole, can (in principle at least!) be mimicked by a humanoid robot. The humanoid form also facilitates human-machine interaction, allowing people to communicate and work co-operatively with their robotic helpers in a completely intuitive manner. Several companies have anticipated the trend in home automation and announced the development of small scale humanoid robots. An early commercial model, HOAP-1, was developed by Fujitsu and aimed purely at the research and development audience. Sony's more recent Qrio is described as an entertainment robot, although plans for commercialization have yet to be announced. In contrast, the wheeled humanoid robot from Mitsubishi known as Wakamaru is intended to provide practical domestic services such as security, health care, information and tele-presence, and will soon be available to domestic buyers. Larger scale humanoid robots are also in development for commercial interests, most notably Asimo from Honda Motor Co., and may represent the first practical universal aids. The figure 1.1 shows the robots HOAP-1, Qrio, Wakamaru and Asimo.

The development of autonomous humanoid robots require advances in diverse fields including mechanics, control, vision, sensing and artificial intelligence. To approach this task pragmatically, humanoid systems are usually divided into smaller sets of related skills that can be developed with reasonable independence. The functional building blocks that enable a humanoid robot to

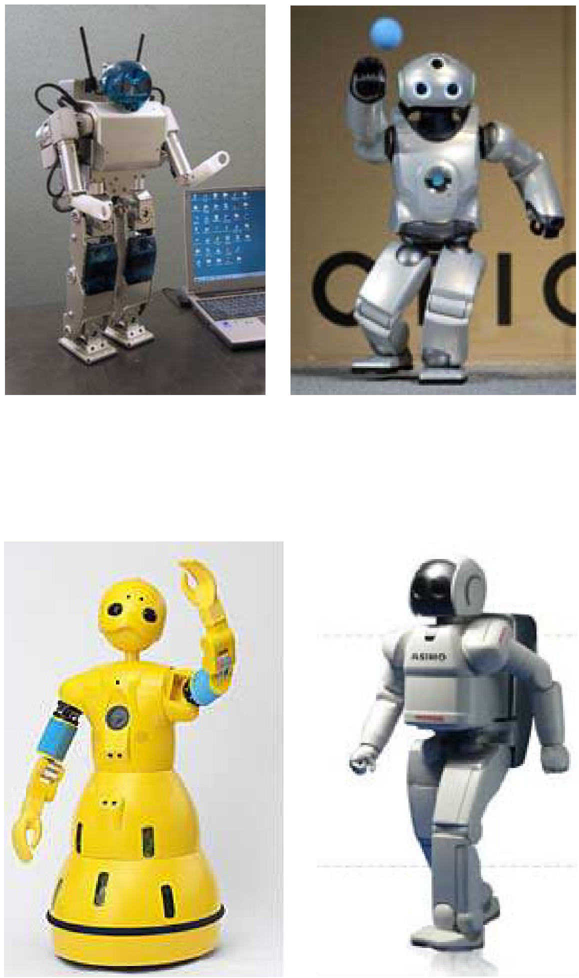

Figure 1.1: The robots *Top left*: HOAP-1, *Top right*: Qrio, *Bottom left*: Wakamaru and *Bottom right*: Asimo.

perform tasks can be broadly divided into the areas of robotic vision and gait control.

Robotic vision is an essential building block for autonomous operation. Complex humanoid tasks require a more sophisticated approach involving the maintenance of a consistent world model, and associated high-level interpretation to drive planning, prediction and tracking of objects. For humanoid robots, interaction with objects or humans are quite important to learn behavior. For such interaction, vision plays an important role in gathering information or in directing attention or gaze on the object or human. In this role, depth information is very important in two ways. First, depth information enables the robot to achieve interaction in 3-D space. Second, depth is a good information source for recognition of stationary objects.

Computer vision research has a long history among many fields of Artificial Intelligence. Moreover, stereo vision is one of the classical problems in the computer vision field and large amount of research approaches have been followed. In this book we present the construction of an active vision system to get depth information of an object which can move to variable velocities for the visual behavior of the humanoid robot and describe the intelligent techniques used to improve its performance. This research is a part of a project to build a humanoid robot called “Dany Walker” [214]. The Dany Walker (Figure 1.2) project has an ambitious goal to build a intelligent biped humanoid with two legs which have six degrees of freedom (DOF) being developed at the Freie Universität Berlin. The active vision head has mostly the same mechanism and controls as Dany walker’s system, but implements only the head. The scope of this book does not include behavior of the legs and gait control, but only the active vision system.

In the book two different active vision system were developed, one considered as previous step called "Monocular robotic Head" that had as objective to carry out the permanent object tracking but without depth information and the "Stereo robotic Head". During the book the terms active vision system and robotic head with their respective restrictions will be used to refer to the same system.

Intelligent techniques [209], an innovative approach to constructing systems, has just come into the limelight. It is now realized that complex real-world problems require intelligent systems that combine knowledge, techniques, and methodologies from various sources. These intelligent systems are supposed to possess humanlike expertise within a specific domain, adapt themselves and learn to do better in changing environments, and explain how they make decisions or take actions. In confronting real-world computing problems, it is frequently advantageous to use several computing techniques synergistically rather than exclusively, resulting in construction of complementary hybrid intelligent systems.

Intelligent techniques consist of several computing paradigms, including neural networks, fuzzy set theory, approximate reasoning, and derivative-free op-

Figure 1.2: The Dany Walker humanoid.

| Methodology | Strength |
|---|---|
| Neural Network | Learning and adaptation |
| Fuzzy systems | Knowledge representation via if-then rules |
| Genetic algorithm | Systematic random search |
| Conventional AI | Symbolic manipulation |

Table 1.1: Intelligent techniques.

timization methods such as genetic algorithms and simulated annealing. Each of these constituent methodologies has its own strength, as summarized in Table 1.1. The seamless integration of these methodologies forms the core of soft computing; the synergism allows soft computing to incorporate human knowledge effectively, deal with imprecision and uncertainty, and learn to adapt to unknown or changing environment for better performance. In general, Intelligent techniques do not perform much symbolic manipulation, so we can view it as a new discipline that complements conventional artificial intelligence (AI) approaches, and vice versa.

The active vision system was decomposed into four layers as object localization, tracking, control and depth measurement. We adopted the developmental approach, which is based on intelligent techniques, for the robot vision.

Recent developments in neural networks and fuzzy logic have changed the robot vision field dramatically. During the past few years there has been a large and energetic upswing in research efforts aimed at synthesizing fuzzy logic with neural networks. Neural networks provide algorithms for learning and are modeled after the physical architecture of the brain. Fuzzy logic deals with issues such as reasoning at the semantic or linguistic level and is based on the way brain deals with inexact information. Consequently, the two technologies complement each other. A variety of fuzzy-neural network models have been used in computer vision.

Neural networks (NNs) have been used to model the human vision system. They are biologically inspired and contain a large number of simple processing elements that perform in a manner analogous to the most elementary functions of neurons. Neural networks learn by experience, generalize from previous experiences to new ones, and can make decisions. Neural network models are preferred for image-understanding tasks because of their parallel-processing capabilities as well as learning and decision-making abilities. Robot vision deals with the recognition of various objects in a scene. It includes image processing and tracking. Often, the ultimate aim in developing a computer vision system is to perform tasks that are normally performed by a human vision system. Neural network models are also known as connectionist models or parallel distributed processing (PDP) models. Neural network models provide an alternative approach to implementing image enhancement techniques. Researchers such as Grossberg [210] considered properties of the human vision system and proposed neural network architectures for brightness perception under constant and variable illumination conditions. Neural network models based on Gabor functions [211] have been used for texture segmentation. The usage of Gabor functions has evolved because the receptive fields of neurons in a visual cortex are known to have shapes that approximate two-dimensional Gabor functions. Neural networks represent a powerful and reasonable alternative to conventional classifiers. Neural network models with learning algorithms such as backpropagation are being used as supervised classifiers, self-organizing neural networks and *learning vector quantization networks* with learning algorithms such as competitive learning or adaptive resonance theory (ART) are being used as unsupervised classifiers in the segmentation or localization of complex colors. Feed-forward networks with backpropagation-learning algorithms have been used in many pattern recognition applications.

The past few years have witnessed a rapid growth in a number of applications of fuzzy logic. Fuzzy logic techniques represent a powerful alternative to design smart engineering systems. Many rule-based expert systems are being used in practice. There are many advantages to using fuzzy inference systems. Fuzzy systems are linguistic, not numerical, making it similar to the way humans think. Fuzzy systems map input variables to output variables, and this mapping is defined in terms of linguistic rules. These systems simplify knowledge acquisition and representation. Fuzzy logic systems are robust and cheaper to make than conventional systems because they are easier to design. Fuzzy logic techniques have been used in many image recognition problems, such as the detection of edges, feature extraction, signal estimation, classification and especially *clustering*, which is used not only to segment image and categorize data, but are also useful for data compression and model construction.

Tracking is a standard task of computer vision with numerous applications in navigation, motion understanding, robot control, surveillance and scene monitoring. In an image sequence, moving objects are represented by their feature points detected prior to tracking or during tracking. Feature points may have local image properties assigned to them. In many applications, e.g., surveillance and scene monitoring, objects may temporarily disappear, enter or leave the view field. In some tasks, such events are of particular interest, while in others they are treated as admissible but disturbing. The character of motion and the merit of tracking quality also vary from task to task. Two main classes of tracking methods are traditionally distinguished: the optical flow based and the local feature based techniques. The algorithms based on optical flow are expensive to be implemented in real time tracking. Powerful algorithms based in feature techniques or similar region are the *CAMSHIFT, Kalman filter* and *Particle filter*.

During the past few years, there has been a large and energetic upswing in research efforts aimed at synthesizing fuzzy logic with neural networks. This combination of neural networks and fuzzy logic seems natural because the two approaches generally attack the design of "intelligent" systems from different angles. Neural networks provide algorithms for learning, classification, and optimization, whereas fuzzy logic deals with issues such as reasoning on a higher (semantic or linguistic) level. Consequently, the two technologies complement each other [212]. By integrating neural networks with fuzzy logic, it is possible to bring the low-level computational power and learning of neural networks into fuzzy logic systems. The synergism of integrating neural networks with fuzzy logic systems into a functional system with low-level learning, high-level thinking, and reasoning transforms the burden of the tedious design problems of the fuzzy logic decision systems to the training/learning of connectionist neural networks. There are many ways to synthesize fuzzy logic and neural networks. An architecture that fuses both systems is the **ANFIS** (adaptive network based in a fuzzy inference system) which is a class of adaptive networks that are functionally equivalent to fuzzy inference systems According to the network characteristics and learning algorithm the ANFIS results attractive for estimate and prediction applications.

Application of fuzzy inference systems to automatic control was first reported in Mamdani's paper [213] in 1975, where, based on Zadeh's proposition, a fuzzy logic controller (FLC) was used to emulate a human operator's control of a steam engine and boiler combination. Since then, fuzzy logic control has gradually been recognized as the most significant and fruitful application for fuzzy logic and fuzzy set theory. In the past few years, advances in microprocessors and hardware technologies have created an even more diversified application domain for fuzzy logic controllers, which ranges from consumer electronics to the automobile industry. Indeed, for complex and/or ill-defined systems that are not easily subjected to conventional automatic control methods, FLCs provide a feasible alternative since they can capture the approximate, qualitative aspects of human reasoning and decision-making processes. However, without adaptive capability, the performance of FLCs relies exclusively on two factors: the availability of human experts, and the knowledge acquisition techniques to convert human expertise into appropriate fuzzy if-then rules and membership functions. In these situations an interesting option is the *adaptive control* which is recommended for systems operating in variable environments and/or featuring variable parameters.

The object tracking and localization are two of the most important tasks in robots that allows to work necessarily with its environment, however an important problem is that although the object has been located in the image coordinated $x$ and $y$ it is not possible to recover the position from the object to the robot. Calculating the distance of various points in the scene relative to the position of the camera is an important task called stereo vision. A common *correspondence method* for extracting such depth information from intensity images is to acquire a pair of images using two cameras displaced from each other by a known distance.

In this work Neural networks and Fuzzy algorithms are used for the object localization. For tracking, searching most similar region approaches (Camshift and Particle filter) were utilized. A neuro-fuzzy prediction mechanism in tracking module made the tracking more stable. The stereo active vision system was controlled using adaptive and fuzzy algorithms which modify their behavior depending on the movements carried out by the tracked object. For the depth determination, we used a simple correspondence procedure based in a epipolar assumption. As a result of the combination of these modules and techniques, the system demonstrated real time tracking, velocity, and robust control.

## 1.2 Related work

Active vision systems have been studied extensively in the robot vision literature, both because of its intrinsic interest and because of the large number of applications. For example, autonomous robots may need to be able to follow objects in their environment ([169], [170], [171], [172] and [173]); one commonly studied special case of this concerns autonomous guided vehicles for driving on roads, which must track the features of the road ([174] and [175]) and also other moving vehicles [176]. Active vision may also be used in robot arm applications to capture multiple views of an object from a moving camera and thus compute trajectories for exploring freespace [177] or to select an optimal grasp to pick up the object [178]. There is increasing interest in using active vision to augment the robot-human interaction [179], including using lip-tracking to aid speech recognition ([180], [181] and [182]). Reliable hand-tracking is vital for this goal, and various systems have been proposed ([183], [184] and [185]) and gesture recognition ([186], [187], [188] and [189]). Hand gestures are a special case of the developing field of "perception of action" which attempts to use tracking information to infer knowledge about a scene. This has roots in the tracking of people ([190]; [191] and [192]) for surveillance applications, as well as creating artificial environments ([193] and [194]) which respond to human actions, for example creating an interactive playroom for children [195]. There is much current interest in learning to classify the output of such trackers into behaviours, for example ([189], [196], [186] and [187]). General techniques for tracking, not tied to any particular application, include the use of optic-flow information, for example [197], rigid three-dimensional models ([198] and [199]) and contour outlines [200].

Much research has been accomplished to detect and track objects of interest acquired by humanoid vision system. Cues such as color [201], disparities [202], optical flow and 2-D shape have been used to implement real-time active vision systems. Researchers have typically studied behaviors such as visual attention [203], vestibulo-ocular reflexes [204], saccadic movements, smooth pursuit and mimicking of human movements[205]. A modern tendency is the development of systems well-known as "biologically inspired" where are considered techniques and subsystems that try to reproduce the operation of active vision systems that are in the nature [206]. Although there are many works of active vision systems in humanoids few use the flexibility of the intelligent techniques for their construction, as exception we can mention [207] and [208], in both works were used neural networks and intelligent agents for the localization and tracking of objects.

## 1.3 Book Organization

The book tries as much as possible to be auto-contained, for this reason has chapters that seek to explain in detail some of the techniques considered as important in the development and control of the vision system.

Every chapter in this book contains the pertinent results needed to implement the stereo active vision system except for the chapters 2,6,7 and 10.

The organization of the remaining chapters is described below. In addition, the accompanying CD-ROM provides videos, MatLAB files and published papers to supplement this dissertation.

**Chapter 2** explains in detail the main algorithms used for the object tracking. The results produced by these algorithms in the object tracking under different conditions are also presented.

**Chapter 3** proposes the Fuzzy C-Means algorithm for the segmentation of not trivial-color objects. It is chosen as benchmark, by the characteristics that it represents the face segmentation.

**Chapter 4** explains the Learning Vector Quantization (LVQ) networks generalities and proposes their use in the segmentation of not trivial-color.

**Chapter 5** presents the control and tracking of the Monocular robotic Head. We describe in this chapter the implementation of fuzzy controllers based on the fuzzy condensed algorithm and also present a comparison of the obtained results with the PID controllers.

**Chapter 6** describes the use of ANFIS model to reduce the delays effects in the control for visual tracking and also explains how we resolved this problem by predicting the target movement using a neurofuzzy approach. This prediction mechanism in tracking module made the tracking more stable.

**Chapter 7** explains the stereo tracking problem and presents the control and tracking of the Stereo robotic Head. In this chapter the techniques and algorithms previously seen are used, for the construction of the stereo active vision system. Also algorithms of trajectory pursuit to increase the tracking robustness are explained.

In addition to the preceding material, the book contains three appendices.

**Appendix A** presents a summary of the Least Mean Square method used several times in the book for the parameters determination.

**Appendix B** presents the mechanical, electric and electronic description of the Monocular and Stereo robotic Heads.

**Appendix C** shows the functions and methodology, to use MatLAB and the image processing toolbox in computer vision applications.