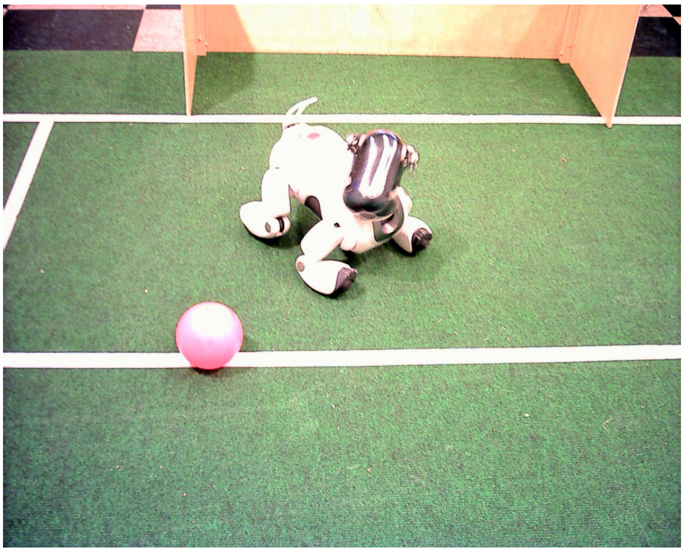
Attention Steering in Behavior-Based Vision

Bachelor Thesis

Supervisor: Arnoud Visser

Artificial Intelligence
University of Amsterdam / Informatics Institute
Kruislaan 403, Room F 105
1098 SJ Amsterdam
The Netherlands



Author: Bayu Slamet
0008540
baslamet@science.uva.nl

Publication: July 1, 2005

Summary

Mantz has shown that making vision behavior-based can yield significant performance gains for soccer playing Aibo's. Psychological research however outlines a severe limitation of selective attention. The phenomenon of 'change blindness' has been extensively studied in various experiments on human participants.

In this thesis a behavior-based model is presented to actively steer attention in order to overcome these limitations. Two approaches to steer attention are possible: based on time intervals and based on observation confidence. The timing-based approach releases attention after a certain interval threshold is reached. These thresholds can be defined by hand or learned from experience and it seems appropriate to define thresholds per behavior.

The confidence-based approach uses thresholds that are measures of confidence of the current observations. When current observations are sufficiently reliable, attention is relocated to broaden the robot's view. In this approach the confidence thresholds need to be calibrated by hand or learned from experience.

In our suggested model for attention steering we provide a synthesized view where confidence measures are combined with timing measures to provide an optimal solution to steering attention that avoids possible limitations that would result from using only one of the approaches.

Contents

Summary ... 2
Contents ... 3
Acknowledgements ... 4
Introduction ... 5
  The 4-Legged league ... 5
  The Dutch Aibo Team ... 5
Aibo Software Architecture ... 6
  Sense-Think-Act cycle ... 6
  Process Layout ... 7
  Modules and Solutions ... 7
  Algorithms ... 7
    Image Processing ... 7
    Self Localisation ... 9
    Behavior Control ... 9
    Motion Control ... 10
Image Processing Improvements ... 12
  German Team 2004 Image Processor ... 12
    Color Table Generalization ... 12
    Object-Dependent Image Processing ... 13
  Mantz: Behavior-Based Vision ... 13
    Object-Dependent Segmentation ... 13
    Object-Dependent Detection ... 14
    Behavior-Dependent Image Processing ... 14
  Integrating Best of Both for the Dutch Team ... 15
    Suggested Approach ... 15
    Integration Results ... 17
Selective Attention ... 18
  Modeling Attention ... 18
    Saliency ... 18
    Target-Driven Control of Attention ... 18
    Contextual Priors ... 19
  Knowledge Representation ... 19
  Change Blindness ... 19
An Improved Model to Attention Steering ... 20
  Attention Lingering ... 20
    Time-Based Lingering ... 20
    Confidence-Based Lingering ... 21
  Synthesis of Time and Confidence ... 21
Conclusion ... 22
Discussion ... 22
Future Work ... 22
Literature References ... 23

Acknowledgements

I would like to thank the Dutch Aibo Team for providing students with such a platform for doing research. Especially I would like to thank my supervisor Arnoud Visser for his confidence, support and insights and Jürgen Sturm for his help, patience and inspiration.

Introduction

In May 1997, IBM's Deep Blue supercomputer [7] defeated the reigning World Chess Champion Kasparov. Therefore, the year 1997 will be remembered as a turning point in the history of Artificial Intelligence.

In the same year the first RoboCup, the World Championships of robot soccer, was held which since then has been a major annual event for many intelligent autonomous agents researchers who participate in one of the leagues.

When compared to robot soccer, chess can be characterized as a static, turn-taking environment where a single agent has access to the complete world state. On the contrary, robot soccer can be characterized as a highly dynamic, real-time environment where knowledge is distributed on multiple agents that each have access to only a small part of the world state.

Each year, after the world championships, the participating teams have to publish their software. This way the teams can learn from each other and innovations can be based on the best implementation available. This serves the ultimate purpose of the RoboCup organization which is to improve on the various fields of science that correlate to intelligent autonomous systems so that in 2050 a team of robots can play a reasonable soccer match against a human team.

The 4-Legged league

The 4-legged league [23] is played on a field of approximately 4 by 6 meters with teams of 4 players where each player is one of the well known Sony Aibo robot dogs [27]. Each team should consist of a goalie and 3 field players [28], [29]. The Aibo has a programming interface and as participants are not allowed to make any modifications to the hardware in this league, the various teams that compete in this league win or lose by the quality of their software.

The Dutch Aibo Team

The Dutch Aibo Team [3] is a joint cooperation of various Dutch Universities. In 2004, the Dutch Aibo Team first participated in the 4-legged league of the international robot soccer competition with inspiring results. At the world championships of robot soccer the team managed to become 4th out of 6 in the soccer competition and to reach the 6th place out of 19 in the Challenges. This secured them a pre-qualification for the next RoboCup in July 2005 where they aim for a ranking in the top 3.

Aibo Software Architecture

When the Dutch Aibo Team first participated in the 4-legged league in 2004 they chose to use the German Team code of 2003 (GT2003 [24]) as a basis for their developments [13], [14], [15], [16]. In retrospect, this was a good choice, as the Germans won in 2004. Now the Dutch team wants to merge the German software of 2004 (GT2004 [25]) with some aspects of their own software of 2004 (DT2004 [14]) and some improvements that are issued by Mantz from the Technical University Delft [10].

In this chapter the key parts of the software architecture of the German Team code of 2003 will be described in order to build up a general understanding of the workings of the soccer playing agent. This gives the foundation to explain and compare the improvements implemented in 2004 by the Germans and by Mantz in the next chapter.

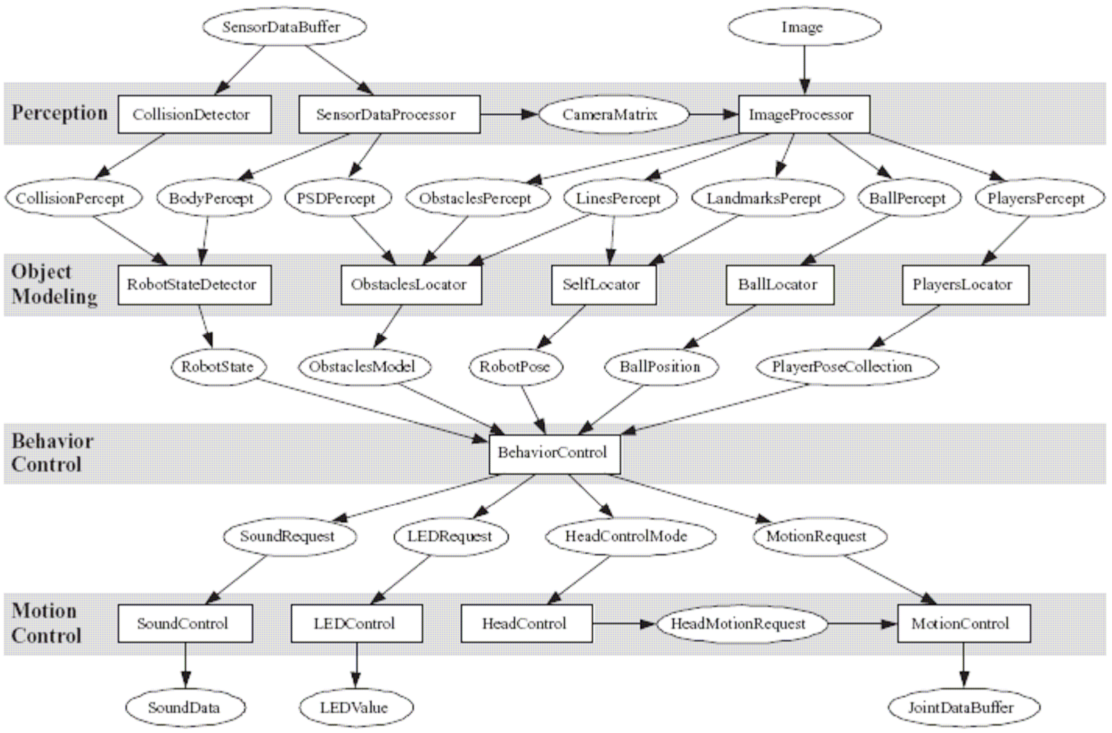


Fig. 1: Overview of German Team 2003 architecture

Sense-Think-Act cycle

A well known paradigm to characterize autonomous systems software is through the sense-think-act paradigm.

Sense: although the Aibo is equipped with several sensors, no sensor can provide the range and detail as the camera that is placed in the robot's nose. Most algorithms, like self localization, rely on the image data that is made available through the ImageProcessor which processes the raw images that are collected with the camera. All information extracted from sensor data is stored in a world model, internally this world model is called the *percepts* collection.

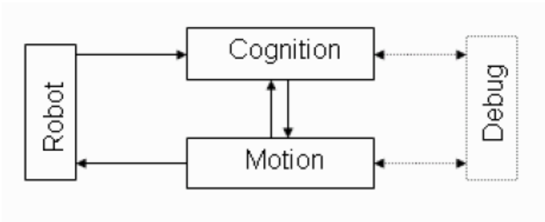
Think: the behavior control layer institutes the brain of the robot. Using symbolic reasoning it analyzes the world model that is stored in the percepts collection and chooses his actions. These actions are sent to the motion control layer.

Act: the action requests retrieved from behavior control are translated to specific motor commands and joint values. Using inverse kinematics and pre-programmed special actions the Aibo makes his moves.

Process Layout

In the software architecture there are several processes that run in parallel:

- Cognition
- Motion
- Robot
- Debug



The Cognition process is responsible for processing all sensor data and for executing the self localization and behavior control which make use of this sensor data. All these tasks are grouped in a single process as they all run in the same pace as sensor data is retrieved and processed. The Motion process is responsible for calculating the motor commands. This functionality is executed in a separate process as this always needs to run on full frame rate, without having to wait for Cognition. The Robot process is responsible for grabbing images and for controlling the motors.

During debugging a 4^{th} process can be activated: Debug. This process allows remote computers to connect with the Aibo and send and retrieve debug messages. This process never runs during actual RoboCup matches.

Modules and Solutions

All information processing is split into numerous *modules*. Every module has a specific task which is defined using interfaces. For most of the modules multiple exchangeable *solutions* exist. This means that different approaches to a particular task can be implemented as different solutions to the same module. The solution that is currently executed can be switched at runtime. This allows for easy testing and comparing different approaches to a particular problem (or task).

Algorithms

In this section the essential algorithms of the Aibo software will be covered.

Image Processing

The ImageProcessor takes the raw camera image as input and searches it for known objects like the field, field lines, goals, the ball and other players. All detected objects are stored in a percepts collection that is forwarded to subsequent modules.

The image processing is mainly based on color segmentation followed by shape detection. While doing so it tries to make smart use of only few of the available pixels to improve performance.

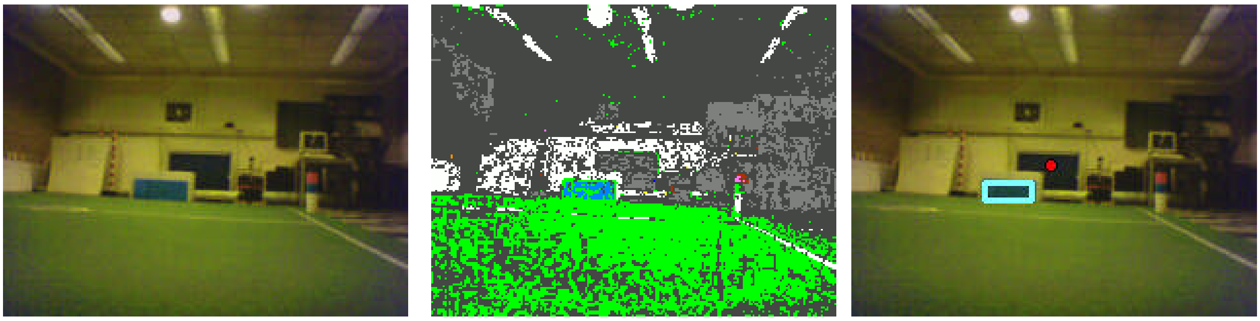


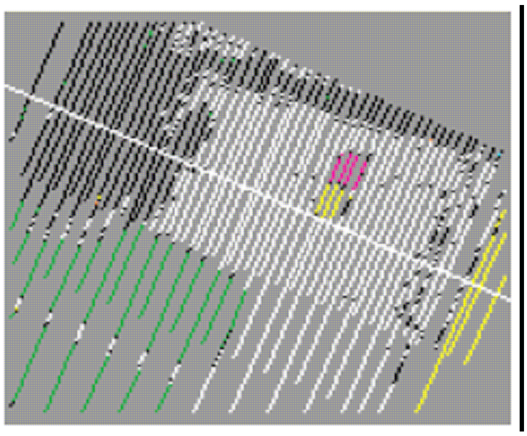
Fig. 2:Image Processing, on the left the raw image as grabbed by the camera, in the center the same image segmented using the color table and on the right the detected goal.

Color Table Segmentation

Every pixel in the image contains a three dimensional value p(y, u, v), where y represents the intensity and where u and v contain the color information. Each value takes an integer value between 0 and 254, which yields the potential for 255*255*255 different colors.

To simplify the object detection problem all these possible color values are mapped onto only a few number of color-classes. This mapping is done using the color table, which is in fact a huge 3-dimensional matrix that computes the color class for every possible pixel value. Calibration of this color table is a tedious task and several approaches to simplify this have been investigated [8], [26].

Object Detection



A priori knowledge about the possible objects to detect is used to guide the object detection algorithms. For example, it is known that the field is always below the horizon and that goals and beacons are at or above the horizon.

The first step in object detection is to compute the location of the horizon (yellow line in figure on the right). Based on the location of the horizon a grid of scanlines is constructed parallel to the horizon (the white lines). For every pixel that lies on one of the scanlines the color class is determined using the color table. Pixels that belong to the same color class and are sufficiently close to each other are clustered.

Each cluster represents a possible percept (detected object). The color information on the scanlines is compared with preprogrammed patterns to determine if the cluster should be considered a percept. For example: a series of green pixels followed by some white pixels below the horizon is likely to indicate a part of the field and a pink cluster above the horizon is likely to indicate a flag.

Self Localization

Self localization is primarily based on a Markov decision process that employs the Monte Carlo approach [2]. In this approach the current position of the robot on the field is modeled as the density of a set of particles. Each particle represents a possible position of the robot on the field. Therefore a particle mainly consists of a vector representing the hypothetical (x, y) coordinates of the robot in millimeters and its rotation in radians.

The localization technique first positions all particles based on the motion model of the (odometry) previous action of the robot. Then it computes the probability of each particle based on the current perceptions of the robot. The particles are resampled towards the locations with higher probabilities and then the average of the resampled probability distribution is taken as the estimate for current pose of the Aibo.

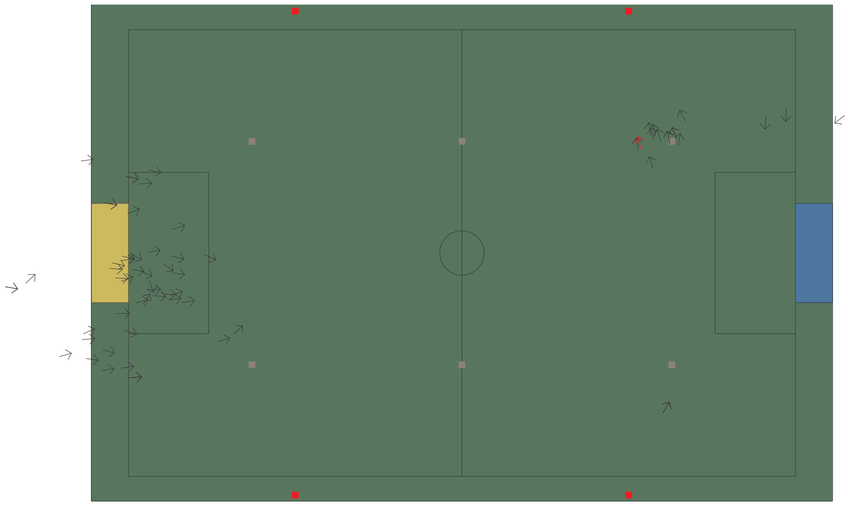


Fig. 3: The particles are shown in black, the estimated position would be near the yellow goal.

Behavior Control

The Behavior Control module is responsible for making decisions based on the world state, the current game state and the behavior that is executed currently by the Aibo. The Behavior Control outputs the following to the motion control layer:

- a motion request that specifies the next motion of the robot
- a head motion request that specifies the mode how the robot’s head is moved,
- a LED request that sets the states of the LED's,
- a sound request that selects a sound file to be played by the robot’s loudspeaker,
- a behavior team message that is sent to other players by wireless communication.

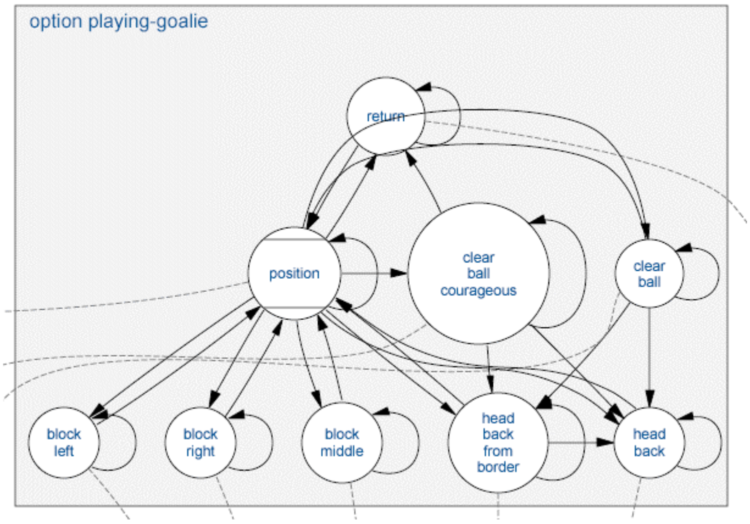


Fig. 4: The finite state machine for the playing-goalie behavior

Above you see a visualization of the finite state machine that specifies the behavior of the goalie. In every state a particular basic behavior is executed, specified by a motion requests. Using decision trees transitions can be made to subsequent states (indicated by arrows). These decision trees base their reasoning on symbols relevant to the soccer game like the position of the ball and the current pose (location and orientation) of the robot on the field.

Motion Control

This module computes the joint-values sent to the motors. The Aibo can make 3 types of motions:

- Walking Motions
- Special Actions
- Head Motions

The Walking Engine is used for all walking motions. Using a lot of parameters you can configure the walking engine which are then translated to the appropriate joint values using inverse kinematics.

The Special Actions are predefined sequences of joint values that result in a particular movement. All kick actions, get-up actions and cheering and artistry actions are defined in special '.mof' files.

The Head Control controls the head motions independently from leg motions. The head motions are mainly predefined loops of circular movements that facilitate searching and/or focusing.

All motions can be requested by the behavior control using motion requests.

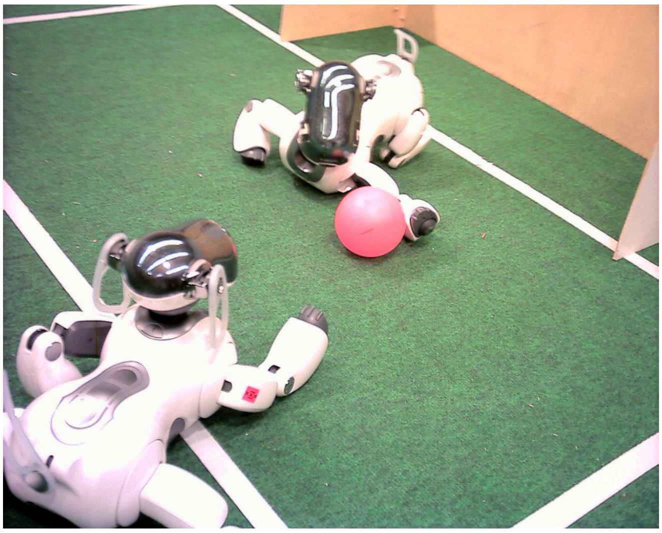


Fig.: Many motions, especially for the goalie and the striker, are predefined. The striker can execute a wide range of kicks and the goalie is equipped with several advanced defensive moves.

Image Processing Improvements

The German Team code from 2003 has served as the root for several development branches. In this chapter only subsequent work that involved improvements on the image processor are discussed. In the first place the German Team made many improvements in their software that they used for the RoboCup in 2004 (GT2004 [25]). While in 2003 they used a single approach to image processing that was used for the detection of all objects, in 2004 they modularized their approach to image processing into object-dependent algorithms

Mantz, from TU Delft, went a step further in his Master thesis work and not only improved the image processing by making it object dependent, he also added behavior dependency [10], [11].

German Team 2004 Image Processor

Before, almost all object detection algorithms were programmed in one big lump of code. They all shared the same approach to image processing, using a single color table for segmentation and a single grid of scanlines that was constructed around the horizon.

The improvements in 2004 are numerous. Most noteworthy are an attempt to color table generalization and the externalization of object-specific algorithms into so-called 'Specialist' classes (BallSpecialist, GoalSpecialst, etc).

Color Table Generalization

One of the key problems in the RoboCup domain is to reach a high degree of robustness of the vision system to changing lighting conditions. Despite several attempts to achieve automated self-calibration ([8], [26]), manual calibration remains the most efficient approach.

An attempt was made to improve the manually calibrated color table so that it would also cover lighting conditions not present in the images used for calibration. To achieve this, a generalization technique which uses an exponential influence model was implemented. In practice, the improvements that result from this generalization only have little positive effect.

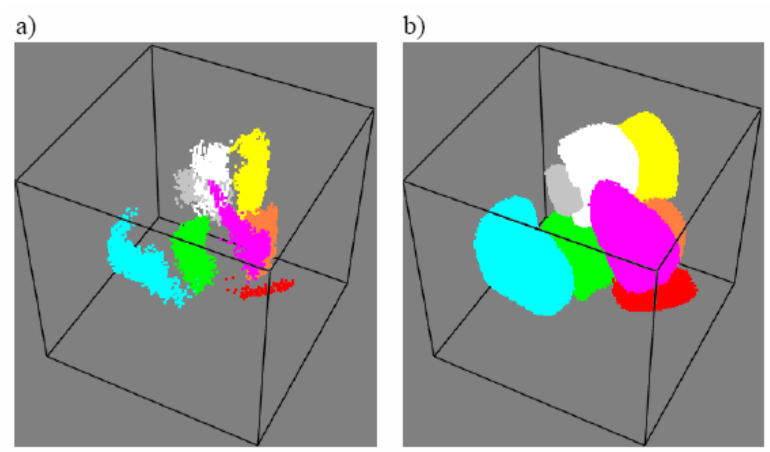


Fig. 5: plots of the color space covered by the color table a) original; b) generalized;

Object-Dependent Image Processing

In the central image processor still a single grid of scanlines is used to determine interesting parts of the input image. When interesting clusters of a certain color are found, then the appropriate Specialist is executed to determine if an object can be detected.

The scanlines grid used by the central image processor is coarser than the one used in the past, but the Specialists make use of fine-grained grids of scanlines that are positioned in a region that covers the cluster of interest and are optimized for the detection of the searched object.

So, the segmentation step of the image processing is still shared by all object detection algorithms, but the grids of scanlines are object dependent.

Mantz: Behavior-Based Vision

Mantz also addressed the key problem of the vision system's robustness to lighting variations. The first part of his approach is similar to the one taken by the German Team, he also implemented object dependent algorithms. However, in addition to object-dependent scanlines grids, Mantz also proposes the usage of object-dependent color segmentation.

Also, Mantz added another dimension of flexibility to the utilization of the image processor. He proposes to have a set of image processors that are each optimized to facilitate particular behavior. His idea is to have the behavior control control the image processor that is currently executed so that the behavior control can choose the image processor that is most likely to provide accurate perceptions needed for proper execution of this behavior.

Object-Dependent Segmentation

The original approach makes use of a single color table to do image segmentation. Several efforts have been made to improve the quality and robustness to changing lighting conditions of the mappings from pixel values to color classes. However, the color table remains a carefully balanced compromise between all color classes. that are needed to meaningfully detect all objects of interest. As is always the case: compromises lead to sub-optimality.

Mantz makes use of a set of color tables, where each color table is optimized for the detection of a particular type of object:

- Lines and ball, optimized for green, white and orange
- Yellow flag, optimized for pink, white and yellow
- Yellow goal, optimized for yellow and green
- Blue flag, optimized for pink, white and sky-blue
- Blue goal, optimized for sky-blue and green

As each color table only needs to differentiate among a few colors, the pixel values that correspond to a certain color class can be calibrated much coarser than in the single color table approach where one color table needed to be able to differentiate among all colors.

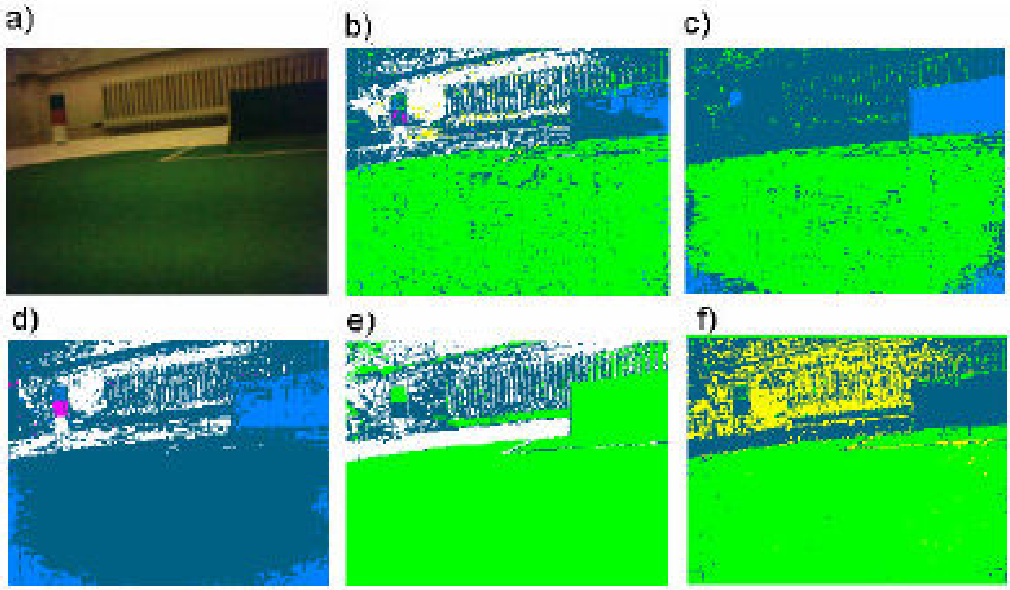


Fig. 6: a) camera image; b) segmented with the general color-table; c) segmented with the blue/green color-table for the detection of the blue goal; d) segmented with the blue/white/pink color-table for the detection of the blue flag; e) segmented with the green/white color-table for the detection of the field lines; f) segmented with the yellow/green color-table for the detection of the yellow goal. From [10].

Object-Dependent Detection

Mantz also externalized all object detection algorithms in specific classes. He also differentiates in what configuration of scanline grid he uses for the detection of particular objects. The main difference is that Mantz algorithms are supplied with color tables that are optimized for the particular task, where the German Team uses the same color table everywhere.

Tuning and testing would reveal for each type of object which solution works best. The one proposed by Mantz, the one proposed by the Germans or perhaps the best-of-breed solution proposed in this thesis.

Behavior-Dependent Image Processing

Another point where Mantz surpassed the Germans in flexibility is in the fact that in Mantz' approach the behavior has a means to specify which set of detection algorithms is preferred to facilitate the currently executed behavior.

For example, when a goalie wants to return to his goal he is only concerned with perceiving his own goal, possibly nearby flags and perhaps the field. This would result in several detection algorithms not being executed which saves processor time and resource consumption. Reasoning further along this line: as it is possible to specify only a subset of algorithms to run at the same time you could have these algorithms do the same amount of work in less time or perhaps do more in the same amount of time.

Another benefit of this is that you can let behavior control switch off algorithms that are known to be unreliable during the currently executed behavior. For example, the detection of far away objects is not so reliable during extensive movements, they are better suited for when a robot stands still.

Integrating Best of Both for the Dutch Team

The following diagram was taken from [30] and clearly outlines the differences and similarities between the approaches. At the top row you see the color tables used for segmentation, in the middle row the scanline grids used for object detection and at the bottom row the algorithms used for shape evaluation.

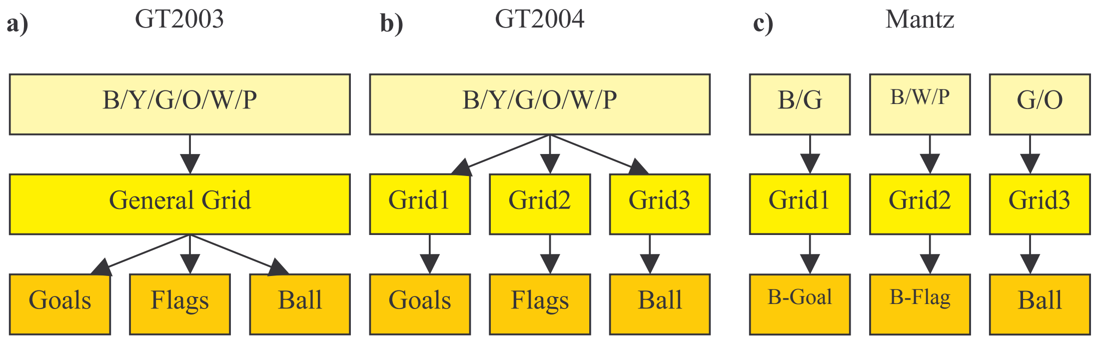


Fig 7: Generic versus object-specific steps in the object-recognition phase:
a) architecture of the German Team 2003, only the shape evaluation (bottom) is object-specific;
b) architecture of the German Team 2004, different scanline grids for different objects;
c) architecture of the Mantz; different color-tables for different objects;
B = Blue, Y = Yellow, G = Green, O = orange, W = White, P = Pink.

When compared to the German Team code of 2003, both mentioned approaches provide serious improvements for the vision system in terms of robustness, performance and especially in Mantz' approach also in terms of flexibility.

In 2004, the German Team was World Champion in the 4-legged league with their improved robot software and Mantz reports over 50% increased performance when his approach was applied to the goalie. Note that Mantz implemented his improvements on the German code of 2003, thus they are likely to yield less spectacular figures when applied on the current German code.

However, his changes are still expected to result in better image processing which will in turn have positive effects on the performance of all executed algorithms and behaviors that depend on it. The object-dependent algorithms may yield very similar results and may be a matter of fine-tuning, but the object-dependent segmentation of Mantz should certainly show improved perception results and it will allow for stricter algorithms in future.

In addition, the feature to be able to switch on or off the image processing algorithms from the behavior control is very beneficial in terms of processor usage and resource consumption on the Aibo.

Suggested Approach

The current Dutch team code is based the German team code of 2004, so the mentioned improvements made by the Germans are already included. This leaves the inclusion of Mantz principles.

As a first step the principle of object-dependent segmentation, by using a set of color tables instead of a single one, could be realized. This should already give an impression of the improvements that can be expected and give clues whether further investigation will be fruitful.

The second step would be to add the behavior dependency so that the behavior control can select which algorithms of the image processor are switched on to facilitate the currently executed behavior. This, together with object based segmentation, results in robots that employ selective attention as Mantz and Torralba describe.

Then one will also observe the limitation that occur due to change blindness. Implement the attention steering model presented in this thesis (see next chapter) to overcome these limitations.

As a final step one could make close comparisons between Mantz' solutions and those of the Germans. Fine-tuning and specific measurements would reveal which set of algorithms give best results.

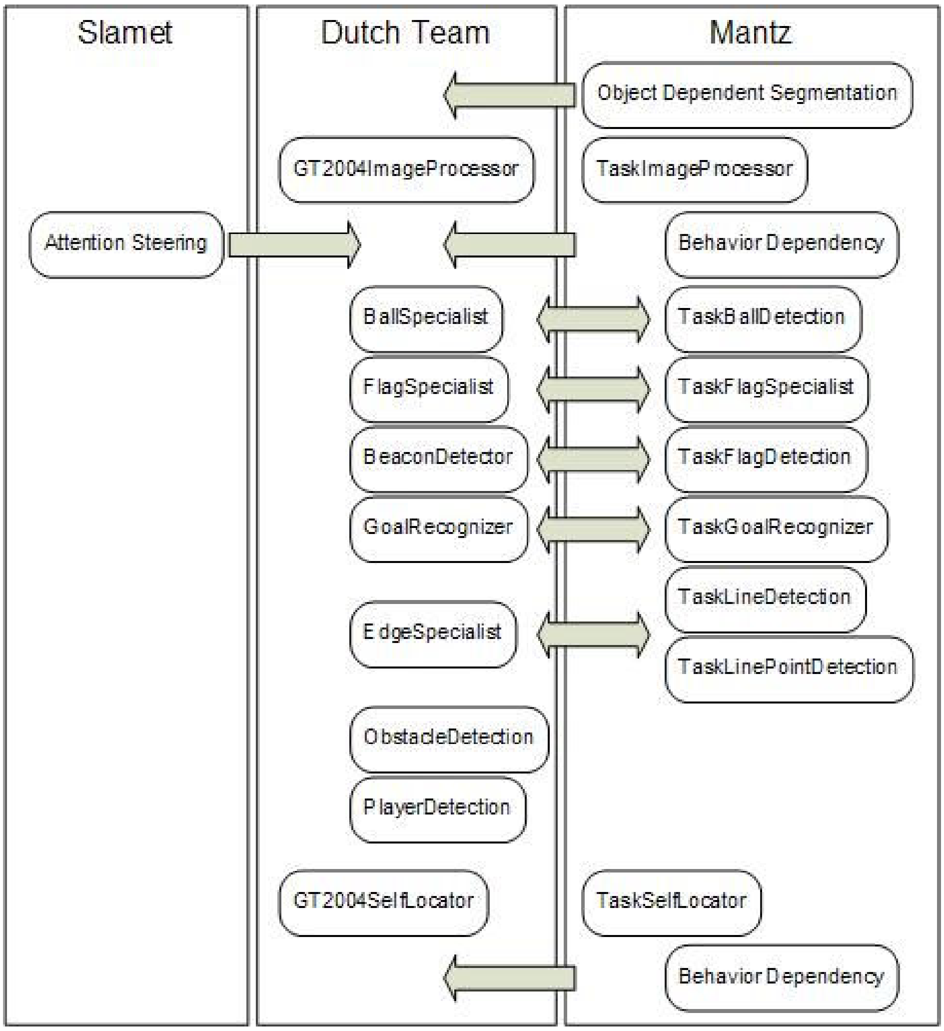


Fig.8: Suggested approach to integration

Note that in the above diagram the self locator modules are also included. Mantz adapted the self locator and made it behavior dependent like his image processor. In his thesis [10] he reports great improvements in self localization for the goalie when it is made behavior dependent, so investigating possible integration of this concept is recommended. In this thesis the self locator will not be further discussed.

Integration Results

When this project started, the modules as implemented by Mantz for the German Team code of 2003 were not yet ported for usage in the current Dutch Team code. As part of this project the modules were merged successfully. Although not of interest for this paper, the behavior dependent TaskSelfLocator was included as well.

The modules are sent to Paul van Rossum from TU Delft, the successor of Floris Mantz, to measure improvement against the original Dutch team code.

The insights gained and lessons learned during the merge were very valuable and are documented for future usage. More important, they enabled the development of the view on selective attention that is presented in the subsequent chapters.

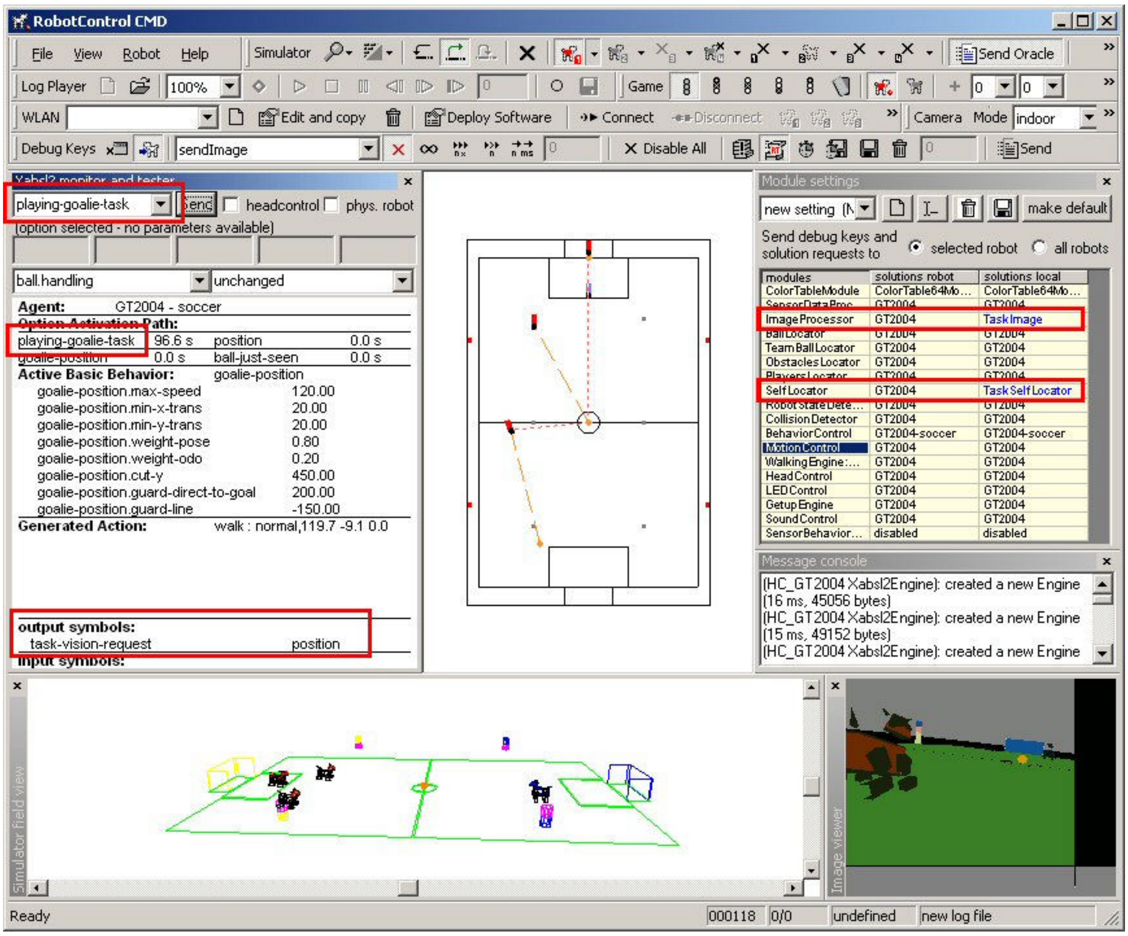


Fig. 9:The merge was successful as can be seen in the debugging interface:
- on the right you see the TaskImageProcessor and TaskSelfLocator modules currently active
- on the left you see the current behavior executed by the goalie which also controls the task-vision-request parameter which steers the selection of algorithms in the TaskImageProcessor. It is currently set to 'position' which means that the image processor will only run those algorithms that facilitate the positioning of the goalie in front of his goal.

Selective Attention

From a psychological point of view, the vision system implemented by Mantz has a closer resemblance to human vision than the German vision system has. How humans make use of their vision capabilities has been studied extensively in psychological research, see for examples Hayhoe [6], Wolfe [37] and numerous work by Rensink and O'Reagan and colleagues [17], [18], [19], [20], [21].

Modeling Attention

Torralba [32], [33] criticizes that models of visual attention predominantly focus on bottom-up approaches and ignore context and scene information. He proposes a change of paradigm where background information (context) is no longer seen as a hindrance or distracter, but is exploited to guide attention to the interesting regions of the image.

He enhances the existing bottom-up approach that uses saliency maps with a top-down approach that uses low-level features of the whole image to prime the presence or absence of objects in the scene. His model tries to predict location, scale and appearance before the actual exploring of the image starts. These predictions are then used to guide the subsequent exploration and are shown to provide an efficient shortcut for object detection.

Mantz approach [10] nicely fits in his computational framework and seems to share the positive effects of using steered attention. The single difference is that Torralba's approach is based on guiding attention towards regions of interest whereas Mantz focuses attention on colors of interest.

Torralba's computational framework is centered around the evaluation of the probability that an object is present given the set of local and context measures: $P(O|v_L, v_C)$

This probability density function is decomposed using Bayes in:

$$P(O|v_L, v_C) = \frac{1}{P(v_L|v_C)} P(v_L|O, v_C) P(O|v_C) \; .$$

Saliency

The first factor, $\frac{1}{P(v_L|v_C)}$, does not depend on the object searched or the task that is currently executed. Therefore this is a bottom-up factor. The general color table is the equivalent in Mantz' approach. The general color table (not to be confused with object specific color tables) is also not dependent on the currently executed task nor on the object searched.

Target-Driven Control of Attention

The second factor, $P(v_L|O, v_C)$, represents the top-down knowledge of the target appearance. Regions of the image with features unlikely to belong to the target object are vetoed and those similar to the searched object are enhanced. This is an exact match with what Mantz uses the object dependent color tables for.

Contextual Priors

The last factor, $P(O|v_C)$, provides context-based priors on object features. This factor does not depend on local measurements and therefore modulates the saliency of local image properties in the search for a particular object.

Given that an object is defined by the tuple $O = \{o, x, t\}$, where o denotes the object class, $x = (x, y)$ the object's location and t its appearance parameters like color and shape, Torralba chooses to further split the context priors by applying Bayes rule several times in order to acquire three factors that model different kinds of context priming on object search:

$$P(O|v_C) = P(t|x, v_C, o)P(x|v_C, o)P(o|v_C).$$

Here $P(t|x, v_C, o)$ gives the likely (prototypical) features of objects of a particular class. In Mantz approach this is realized in the object-dependent detection algorithms which make use of specialized scanline grids and specific shape evaluations.

The subsequent factor, $P(x|v_C, o)$, gives the most likely locations for the presence of an object given the context. This is a direct match with how Mantz positions scanlines in grids near the horizon in order to detect target objects.

The third factor, $P(o|v_C)$, provides the probability density function for objects of a particular type in the scene. Mantz uses the current robot pose (location and orientation on the field) to determine which objects are likely in view.

Knowledge Representation

Traditionally, knowledge representation has been seen as a prerequisite to informed action. Like is the case with the Aibo robots while playing soccer, representations are often assumed to comprise complete descriptions of the agent's environment. In the German Team software the robot tries to construct and maintain a world model that covers as much as can be perceived.

On the other side we have extreme approaches like that of Brooks [1], who denies the existence of representations underlying activity and a little less extreme approaches like that of Hayhoe [6] where representations are generated 'just-in-time' (JIT).

Hayhoe studied temporal dependencies of natural vision by measuring eye and hand movements while subjects made a sandwich. Her observations suggest that much natural vision can be accomplished with JIT representations, but also that JIT representations alone do not explain all. In addition to JIT representations, so she concludes, a coordinate frame representation is maintained that allows for coordination of hand movements independent of eye position.

This view of 'active vision' appears to be supported by various psychological evidence. The phenomenon of inattentional blindness [6], [20] or inattentional amnesia [37], [20] demonstrate the selective nature of natural vision and appear to add support to a selective approach in an agent's vision system like Mantz implemented for the soccer playing Aibo's.

Change Blindness

While research by Rensink, Hayhoe and Wolfe may suggest that human vision is also selective by nature and thus seem to add support to Mantz' approach to vision in the Aibo, this research also outlines potential problems we may encounter.

Rensink and colleagues ([18], [19], [21]) demonstrate several aspects of natural vision which result in failure to notice changes to entities in the visual scene when these take place during saccadic eye movements. Only when the object is fixated, changes can be detected.

Also Hayhoe demonstrates this phenomenon in [6], here the changed feature is even central to the task at hand. In her work she describes how participants fail to notice that a block changes color. It is striking that those participants were explicitly asked to pick up those blocks, which were either pink or blue, and to place these in a particular location according to color and that they just did not notice that the object changed color between pick up and final placement. It appears as if participants only pay attention to color as they pick up the block and during subsequent fixations participants are only concerned with location, in order to get the object in its target placement.

Equally striking is how Wolfe describes in [37] that 50% of the participants failed to notice that when a Stranger1 asks for directions to the participant and is subsequentially being replaced by another Stranger2 after being obstructed from view for a short moment by an enscened event.

It is not hard to see that in Mantz approach the robot dogs potentially suffer from similar limitations. When for example a goalie is executing behavior that is intended to position the robot back in front of his goal, the image processor may be configured for detection of the goal and flags which are typically good markers to guide this 'return to goal' behavior. If, during the execution of this behavior, an opponent striker comes into view with the ball in front of him, the goalie will ignore this new stimuli in his visual field as he is focused entirely on perceiving the goal and the flags. It is clear that different behavior is preferred under the presented circumstances.

An Improved Model to Attention Steering

Tsotsos [34] reports that on average human attention lasts around 250 ms and starts to linger after that. This contrasts sharply with Mantz' approach where fixation lasts as long as a particular behavior is executed, which can potentially be in the order of seconds. An improvement would be to include a similar notion of 'attention lingering' in the Aibo software.

Attention Lingering

Two approaches to attention dwelling are possible. One straightforward approach is to parameterize the time interval after which lingering starts. A more natural approach is to choose the appropriate moment for lingering attention based on the confidence of the current observations.

Time-Based Lingering

This would be straightforward to implement. Whenever the robot indulges in focused image processing a timer is started. When the timer reaches a predefined value a relaxation on this focusing is triggered and the vision system should start paying attention to a broader view.

An intuitive extension would be to have multiple behavior dependent time intervals or perhaps even allow behavior control to set the time interval. One can imagine that appropriate timings for a striker differ from that of a goalie.

The downside of this approach is that extensive calibration of the timings is required. If the time intervals are calibrated to large, the Aibo will start to broaden its view too late and the point of lingering is missed. On the other hand, if time intervals are set too short, then attention may be released before the current observations are reliable. In that case, releasing attention will interfere with accurate vision which is the initial purpose of attention.

One possible approach to overcome this limitation could be to make the timings dynamic and have the Aibo's learn appropriate values during playing. Several algorithms exist that are very well suited for highly dynamic environments [31]. Currently, the self localization algorithm also successfully employs a reinforcement learning algorithm [2], [25].

Having the Aibo learn the correct timings dynamically would also account for changing dynamics in the environment. Cognition algorithms have no constant processing time before reliable observations are acquired. For example, the time needed by the image processor to acquire reliable percepts highly depends on the current lighting conditions, which may change frequently during a single match of soccer.

Confidence-Based Lingering

Another approach would be to make choosing the appropriate moment for lingering directly dependent on the confidence in the current observations. Only when current observations are believed to be sufficiently reliable, lingering of attention is triggered.

The question then remains when perception is sufficiently reliable to allow for broadened attention. A learning approach similar to that mentioned with time-based lingering could be employed. When attention relaxation is triggered, the robot could start monitoring the confidence of the observations that are currently in the focus of attention. As the Aibo's view is broadened this confidence is likely to drop. The robot could learn the appropriate moments, dependent on behavior, where widening its view has less impact on confidence.

Synthesis of Time and Confidence

The view presented in this paper is that should at least employ confidence measures to steer attention. A model of confidence based attention steering could then be fine-tuned using time-based measures.

Using time-based measures alone is in fact an indirect approach to confidence based lingering. Learning the appropriate timings is in fact compensating for the effects of changing attention to soon or to late on the reliability of perceptions. Therefore, one should base attention lingering directly on confidence measures as this is likely to yield more accurate results than in the indirect approach.

Reliability requirements for observations and attention lingering for avoiding change blindness are two opposing forces. The appropriate balancing factor between these forces can not be covered in a single fraction. A behavior based prioritizing scheme is recommended so that attention can be balanced in accordance with the priorities for the currently executed behavior.

So far, this synthesis leaves one hypothetical situation unaddressed: the situation where the vision system fails to acquire sufficiently reliable perceptions. In the currently presented model this would lead to locked focusing and attention would never be released during the execution of a particular behavior.

On could argue that this is quite desirable, as under those circumstances selective attention is the robot's best available utility to still acquire perceptions with some reliability. Our view is exactly the contrary. Especially in these situations, broadening the view may give some key new information that allows the Aibo to climb out of the local lock-up. So, these situations should be detected as early as possible and if encountered attention should be broadened rapidly.

Time-based measures are suggested in a monitoring process at the background that detects these kinds of local lock-ups and then overrules the confidence based steering.

Conclusion

Mantz has shown that soccer playing robots benefit greatly from behavior-based vision [10], [11]. Related AI research in attention modeling by Torralba and colleagues ([12], [32], [33]) and psychological research to natural vision ([6], [17], [20]) add support to his view of selective attention.

However, Mantz does not adequately address the limitations of selective attention, like the phenomenon of change blindness (Rensink [18], [19], [21]) or inattentive amnesia (Wolfe [37]). In this thesis a model has been presented that allows for further extension of Mantz' approach. By using time-based and confidence-based measures the attention can be actively steered to avoid change blindness.

Discussion

In [35] Wood develops a view on selective attention based on experiments conducted by Rensink [21], Hayhoe [6]and Wolfe [37]. She concludes that experience is a strong denominator for guiding attention. In the robot soccer domain however, this experience is built up in the minds of researchers as they participate in the robot soccer competitions. And not, as in Woods scenario, by the agents themselves.

So one should take care when one lets the Aibo learn certain threshold values through own experience. In the past [13], [25] it has already been shown that despite existing automated solutions [8], [26], manual calibration of color tables is still preferred for best performance during soccer matches.

For improved predictability and to allow for a fixed definition that is carefully elaborated during calibration processes, one should always offer the option to manually configure important threshold values that guide part of the information processing..

Future Work

Mantz has shown that making the vision system behavior-based yields significantly better results. In this thesis a behavior-based model was presented that will further improve his results by actively steering attention.

Future work would also lie in this line of making particular information processing based on behavior. Mantz already experimented with behavior-based self localization. Short mention of this was made in this thesis and it is recommended to further investigate the potential benefits.

In addition to the camera sensor the Aibo is also equipped with for example infrared distance sensors, which have been used only scarcely yet. The information provided by these sensors proved to be not valuable enough for extensive use in perception. However, when the acquired information is put into context by including information on the currently executed behavior the provided sensor data may become much more meaningful. Future work would be to investigate possibilities to exploit other sensor data than vision data by making the processing behavior-based.

Literature References

[1] R. A. Brooks. A robust layered control system for a mobile robot. IEEE Journal of Robotics and Automation, RA-2(1), March 1986.

[2] F. Dallaert et al. "Monte Carlo Localization for Mobile Robots", IEEE Conference on Robotics and Automation (ICRA99), May 1999.

[3] Dutch Aibo Team, http://www.dutchaiboteam.nl/

[4] German Team, http://www.germanteam.org/

[5] W. Hammoumi, V. Nedovic, B. Slamet, R. Valenti, "Improving self-localization and behavior for Aibo's soccer playing robots (implementation according to the new RoboCup rules of 2005)", University of Amsterdam, February 2005.

[6] M. M. Hayhoe, A. Shrivastava, R. Mruczek, J.B. Pelz, "Visual memory and motor planning in a natural task", Journal of Vision 3 (2003), pages 49-63.

[7] IBM Deep Blue, http://www.research.ibm.com/deepblue/

[8] M. Jüngel, 'Using Layered Color Precision for a Self-Calibrating Vision System', in the Proceedings of the 8th International Works on RoboCup 2004, Lecture Notes on Artificial Intelligence, 2005. Springer Verlag.

[9] M. Lötzsch, J. Bach, H.-D. Burkhard, M. Jüngel. "Designing Agent Behavior with the Extensible Agent Behavior Specification Language XABSL". In 7th International Workshop on RoboCup 2003 (Robot World Cup Soccer Games and Conferences), Lecture Notes in Artificial Intelligence, Padova, Italy, 2004. Springer (to appear).

[10] F.A. Mantz, 'A Behavior-Based Vision System on a Legged Robot', MSc. Thesis, Delft University of Technology, February 2005.

[11] F.A. Mantz, P.P. Jonker, W. Caarls, 'Thinking in Behaviors, not in Tasks; a Behavior-Based Vision System on a Legged Robot', to appear in the Proceedings of the RoboCup Workshop, Osaka, Japan, July 2005.

[12] A. Oliva, A. Torralba, M. S. Castelhano and J. M. Henderson, "Top-down control of visual attention in object detection". Proceedings of the IEEE International Conference on Image Processing. Vol. I, pages 253-256. September 14-17 2003, in Barcelona, Spain.

[13] S. Oomes, P. Jonker, M. Poel, A. Visser, M. Wiering, "The Dutch Aibo Team 2004", University of Amsterdam, January 2004..

[14] S. Oomes, P. Jonker, M. Poel, A. Visser, M. Wiering, W. Caarls, S. Leijnen, S. van Weers, N. Wijngaards, F. Dignum, "The Dutch Aibo Team Report on RoboCup 2004", January 2005..

[15] B. Ottens, A.Abbo, P.J. van der Meer, M. Stienstra, "Aibo Project 2004 – German Team Report", University of Amsterdam, January 2004.

[16] P. de Oude, T. van Erven, J. Liem, T. van Kasteren, "Evaluation of CMPack 2003", University of Amsterdam, January 2004.

[17] R. A. Rensink, "The dynamic representation of scenes", Visual Cognition 7 (2002), pages 17-42

[18] R. A. Rensink, "When good observers go bad: change blindness, inattentional blindness and visual experience", Pshyche 6(09), August 2000

[19] R. A. Rensink, J. K. O'Reagan, J.J. Clark, "On the failure to detect changes in scenes across brief interruptions", Visual Cognition 7 (2002), pages 127-145

[20] R. A. Rensink, "Seeing, sensing and scrutinizing", Visual Res. 40 (2000), pages1469-1487

[21] R. A. Rensink, "Change Detection", Annual Rev. Psychology 53 (2002), pages 245-277

[22] RoboCup Organization, http://www.robocup.org/

[23] RoboCup Sony Four-Legged Robot League, http://www.tzi.de/4legged

[24] T. Röfer, H-D. Burkhard, U. Düffert, J. Hoffmann, D. Göhring, M. Jüngel, M. Lötzsch, O. von Stryk, R. Brunn, M. Kalluik, M. Kunz, S. Petters, M. Risler, M. Stelzer, I. Dahm, M. Wachter, K. Engel, A. Osterhues, C. Schumann, J. Ziegler, 'GermanTeam RoboCup 2003', 2003. http://www.germanteam.org/GT2003.pdf
[25] T. Röfer, T. Laue, H-D. Burkhard, J. Hoffmann, M. Jüngel, D. Göhring, M. Lötzsch, U. Düffert, M. Spranger, B. Altmeyer, V. Goetzke, O. von Stryk, R. Brunn, M. Dassler, M. Kunz, M. Risler, M. Stelzer, D. Thomas, S. Uhrig, U. Schwiegelshohn, I. Dahm, M. Hebbel, W. Nistico, C. Schumann, M. Wachter, 'GermanTeam RoboCup 2004', 2004. http://www.germanteam.org/GT2004.pdf
[26] D. Schulz and D. Foz, 'Bayesian color estimation for adaptive vision-based robot localization', in the Proceedings of IROS, 2004.
[27] Sony Aibo, http://www.aibo.com/
[28] Sony Four-Legged Robot Football League Rulebook 2004, available at http://www.tzi.de/4legged/pub/Website/History/Rules2004.pdf
[29] Sony Four-Legged Robot Football League Rulebook 2005, available at http://www.tzi.de/4legged/pub/Website/History/Rules2005.pdf
[30] B. Slamet, A. Visser, "Purposeful perception by attention-steered robots", University of Amsterdam, submitted to BNAIC 17, june 2005.
[31] R.S. Sutton, A.G. Barto, "Reinforcement Learning – An Introduction", MIT Press, Cambridge, 1998.
[32] A. Torralba, "Modeling global scene factors in attention", J. Opt. Soc. Am. 20 (7) 1407–1418. 2003.
[33] A. Torralba, "Contextual priming for object detection". International Journal of Computer Vision. Vol. 53(2), 169-191. 2003
[34] J.K. Tsotsos, S. M. Culhane, Winky Yan Kei Wai, Yuzhong Lai, N. Davis, F. Nuflo, "Modeling visual attention via selective tuning", Artificial Intelligence 78 (2005) pages 507-545
[35] S..Wood, "When being reactive just won't do", in: Proceedings of the AAAI Spring Symposium on Integrated Planning Applications, Stanford University, Palo Alto, CA, pages 102–106, March 27–29, 1995
[36] S. Wood, "Representation and purposeful autonomous agents", Robotics and Autonomous Systems 49 (1-2), 2004.
[37] J.M. Wolfe, "Inattentional Amnesia", in: V. Coltheart(Ed.), Fleeting Memories, MIT Press, Cambridge, 1999
[38] XABSL: The Extensible Agent Behavior Specification Language, http://www.ki.informatik.hu-berlin.de/XABSL/