Principled Construction of Perception and Action Skills for Bold Hearts 3D 2010

Sander van Dijk and Daniel Polani

School of Computer Science
University of Hertfordshire, UK

Abstract. We aim at systematically developing a battery of principled methods to generate behaviours useful to achieve a viable RoboCup 3D gameplay.
behaviours. This is particularly interesting and challenging in humanoids since any possible tactical, strategical or cooperative aspects can only be successfully tackled once the basic skills are in place.
The construction of basic skills in humanoids is usually an intricate business that requires a large amount of hand-tuning. We aim to develop a systematic path towards reducing this amount of handtuning both on the perception as well as the actuation side.
We approach this by combining principled methods, many grounded in information-theory, some in well-known Kalman- and particle-filtering, as well as hand-coded components. The long-term goal is to ultimately replace the hand-crafted structuring of the code by learnt frameworks. This Team Description Paper discusses the aspects the Bold Hearts team currently concentrates upon.

1 Base and Locomotion

Team Bold Hearts has competed in the RoboCup Soccer Simulation league since 2003. The first two years the team participated in the 2D competitions, in 2005 it joined the 3D community. At the beginning of 2009 a full restart of the team was initiated, after attracting Sander van Dijk to the team, former member of the successful team Little Green BATS[1]. To get the Bold Hearts up to steam quickly, the new code is based on the `libbats` library released by the Little Green BATS[2].

A simple open loop oscillator model, based on [13], and similar to that released by the Little Green BATS, is momentarily used as a gait generator. This already results in a fast walking behaviour, however the lack of using any sensory feedback leaves open several possibilities for enhancing stability.

Firstly, we attempt to handle sideways stability. In [13] a method is presented to couple the phase of the oscillator to the natural phase of the robot's dynamics. The underlying idea is that this natural phase ϕ_r can be inferred from the current

[1] See http://www.littlegreenbats.nl/
[2] See http://www.launchpad.net/littlegreenbats/

location of the Centre Of Pressure (COP), x_{COP}, and its velocity, $\dot{x}_{COP}$:

$$\phi_r = -\arctan(\frac{\dot{x}_{COP}}{x_{COP}}). \tag{1}$$

This is then combined with the oscillator's static rate of change ω_c to determine the rate of change of the phase of the walking gait:

$$\dot{\phi}_c = \omega_c + K_c \sin(\phi_r - \phi_c), \tag{2}$$

where K_c is a coupling constant, which determines the strength of the coupling. When set to 0, the model reduces to a fully open loop controller.

Initial experiments have shown that the Force Resistance Preceptor used in the simulation has poor resolution, affecting the usability of the COP measure and thus the effectiveness of the phase coupling. To overcome this, we research the effect of different stability measures to replace the COP in (1), most notably the Zero-Rate of Angular Momentum (ZRAM) point, introduced by [4]. This measure takes into account additional postural information, resulting in a more detailed stability indication. Experiments are currently being done to test the usefulness of the ZRAM-point for phase coupling quantitatively.

Besides lateral stability, another important issue to handle is anterior/posterior stability. Currently, a basic method is used to lean the torso in the direction of movement to counteract angular momentum caused by the walking gait:

$$f = min(\frac{v(t)}{v_{max}} \frac{a_{forward}(t)}{a_{side}(t)}, 1) \tag{3}$$

$$\theta(t) = f \cdot \theta_{max}, \tag{4}$$

where $v(t)$ is the agent's velocity at time t, v_{max} its maximum velocity, a its acceleration, $\theta(t)$ the added torso pitch and θ_{max} the maximum torso pitch. Efforts are under way to make this method more adaptive and effective, using reinforcement learning methods and stability measures like the ZRAM-point.

2 General Approach

Learning, specifically *reinforcement learning*, has been part of the RoboCup endeavour for a long time [18, 16, 1, 12]. Reinforcement learning methods are of interest because of their generality and mathematical grounding. They are also quite successful in nontrivial problems [17]; in conjunction with kernel methods, they can address even larger problems in a highly efficient way [2, 8, 7, 6].

Still, the problems to address are quite large (and large-dimensional). However, realistic embodied agents offer a selection of possible partial decompositions [5]. There is significant indication that Shannon information can be a powerful indicator of where "interesting" properties of the environment lie. The use of information-theoretic (or information-theoretically motivated) decompositions is a natural while computationally expensive approach. [3] It has been shown that

it can lead to self-organized feature extraction [10], sensoritopic map formation [14], or identify interesting states in state space [11].

Here, we have several tasks for which we will use informational approaches. The current server dynamics contain limited vision and noise. Combined with a doubling of the amount of players, intelligent vision is an important issue. Part of it will be covered by conventional Kalman-filter approaches. However, we intend to use novel informational principles to address the active-vision task imposed through the limited vision. For this, we will use the novel *Infotaxis* principle [19] to guide actions to identify objects of importance such as ball, goal and other players. Section 3 will give a formal description of our use of this principle in localization, the first use of it in robot control, and the research we will follow after this first step.

3 Active Vision Through Infotaxis

One of the new challenges introduced in 2009 for the Robocup 3D Simulation teams are the restrictions placed on the vision sensor. The previous two years the simulated robots were equipped with perfect 360-degrees omni vision cameras, making the environment fully accessible. From this year, however, a restricted vision sensor is introduced, similar to that used in the spheres version of the simulator until 2006. This sensor has a range of 120 degrees on both the horizontal as the vertical axis and supplies noisy data about the objects within its field of sight. The next sections describe ways we use and directions of research to handle this new challenge.

3.1 Localization

We supply the agents with a localization mechanism that maintains their global location in world coordinates. Many tasks can be achieved with only relative position information, for instance to kick the ball into a goal the relative position of the agent to the ball and to the goal is sufficient. Global coordinates however make it easier for the agent to deduce more about the world, like the trajectory of the ball and whether the team is in an attack or defence situation. In this section we will describe the Kalman filter localization method, a traditional method used to solve prediction problems, as described in [9] and [21].

With this method, the agent's estimated location is described by a multivariate normal distribution $N(\mathbf{x}, \Sigma)$ with means $\mathbf{x}$, here a 3-dimensional vector depicting the agent's x, y and z coordinates in the field, and covariance matrix Σ. After each time step this estimate is refined in two steps: first a *prediction* is made based on the dynamics of the environment and the agent's actions, secondly this prediction is *updated* by integrating observations.

Predict In the prediction step at timestep k the mean $\mathbf{x}_{k|k-1}$ and covariance matrix $\Sigma_{k|k-1}$, where $(\cdot)_{k|l}$ means 'at timestep k, given all observations up to

and including timestep l', are determined as follows:

$$\mathbf{x}_{k|k-1} = \mathbf{A}\mathbf{x}_{k-1|k-1} + \mathbf{B}\mathbf{u}_{k-1} \tag{5}$$

$$\Sigma_{k|k-1} = \mathbf{A}\Sigma_{k-1|k-1}\mathbf{A}^T + \mathbf{Q} \tag{6}$$

where $\mathbf{A}$ is the state transition model relating the state of the previous timestep to that of the current timestep, $\mathbf{u}_k$ is the control vector at timestep k reflecting the agent's actions and $\mathbf{Q}$ is the process noise.

For now we assume $\mathbf{A} = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix, indicating that there is no effect on the agent's location besides its actions. Later on this can be extended by appending the agent's velocity to $\mathbf{x}$. Also, the input control is defined in world coordinates, so $\mathbf{B} = \mathbf{I}$. This results in the simplified equations:

$$\mathbf{x}_{k|k-1} = \mathbf{x}_{k-1|k-1} + \mathbf{u}_{k-1} \tag{7}$$

$$\Sigma_{k|k-1} = \Sigma_{k-1|k-1} + \mathbf{Q} \tag{8}$$

Update The update step uses observations of landmarks at the current timestep, $\mathbf{z}_k$, to refine the estimate:

$$\mathbf{K}_k = \Sigma_{k|k-1}\mathbf{H}^T(\mathbf{H}\Sigma_{k|k-1}\mathbf{H}^T + \mathbf{R}_k)^{-1} \tag{9}$$

$$\mathbf{x}_{k|k} = \mathbf{x}_{k|k-1} + \mathbf{K}_k(\mathbf{z}_k - \mathbf{H}\mathbf{x}_{k|k-1}) \tag{10}$$

$$\Sigma_{k|k} = (\mathbf{I} - \mathbf{K}_k\mathbf{H})\Sigma_{k|k-1} \tag{11}$$

where $\mathbf{H}$ is the observation model relating an observation to a location world coordinates and $\mathbf{R}_k$ is the observation noise covariance matrix. $\mathbf{K}$ is the *gain* or *blending factor* that minimizes the a posteriori error covariance. Note that the observation noise model depends on the current timestep, since the noise when observing far away landmarks is larger than with nearer objects.

The observations are defined in the global coordinate system, so $\mathbf{H} = \mathbf{I}$, resulting in the simplifications:

$$\mathbf{K}_k = \Sigma_{k|k-1}(\Sigma_{k|k-1} + \mathbf{R}_k)^{-1} \tag{12}$$

$$\mathbf{x}_{k|k} = \mathbf{x}_{k|k-1} + \mathbf{K}_k(\mathbf{z}_k - \mathbf{x}_{k|k-1}) \tag{13}$$

$$\Sigma_{k|k} = (\mathbf{I} - \mathbf{K}_k)\Sigma_{k|k-1} \tag{14}$$

3.2 Single-Landmark Observations

As mentioned in the previous section, an observation consists of agent coordinates in the global coordinate system. These can be obtained through triangulation or trilateration of the observed locations of several landmarks. However, the gyroscopic sensor of the agent gives sufficient information to achieve an observation by using only a single landmark.

To achieve this, we maintain the current rotation matrix $\mathbf{T}_k$ of the agent relative to the field. After beaming we set $\mathbf{T}_k = \mathbf{I}$. In all subsequent timesteps we update the matrix using the angular velocity measurement $\dot{\theta}_k$, given by the

gyroscopic sensor. Firstly, the previous rotation estimate is used to transform this measurement from the local to the global coordinate frame:

$$\dot{\theta}'_k = \mathbf{T}_{k-1}\dot{\theta}_k \tag{15}$$

Based on this a rotation matrix Θ_k is constructed, describing the rotation since the last time step in the global coordinate frame. Finally, this matrix is used to obtain the new global rotation matrix estimate:

$$\mathbf{T}_k = \Theta_k \cdot \mathbf{T}_{k-1} \tag{16}$$

With this new estimate a local observation can be transformed into a global location observation, enabling accurate localization based on a single landmark.

3.3 Information Gathering

During a match an agent that focusses solely on the ball will receive sufficient observations of landmarks to be able to localize effectively using the method described in the previous section. However, there are more interesting objects in the field. Especially with the current increase in team size, coordination and keeping track of the opponent's players becomes an important issue, hindered by the fact that an agent can only pay attention to a small part of the field at the same time. Therefore we search for active vision strategies which optimize the amount of useful information gathered by the agent.

To do this we will use the infotaxis strategy which 'locally maximizes the expected rate of information gain'[20]. The information gain resulting from an observation can be measured by the decrease of the entropy $H(f)$ of the distribution $f(\mathbf{x})$. In our case of multivariate normal distribution we have:

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{N/2}\left|\Sigma\right|^{1/2}} e^{-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)} \tag{17}$$

$$H(f) = -\int_{-\infty}^{\infty} f(\mathbf{x}) log(f(\mathbf{x}))d\mathbf{x} \tag{18}$$

$$= \log\left(\sqrt{(2\,\pi\,e)^N\left|\Sigma\right|}\right), \tag{19}$$

where N is the number of dimensions, in our case $N = 3$.

The problem we need to solve is which action $a \in \mathcal{A}$ of the possible actions $\mathcal{A}$ to take to maximize the decrease in entropy:

$$a_k = \arg\max_a -\Delta_a H(X) \tag{20}$$

$$= \arg\min_a H(X)_{k+1|a} - H(X)_k \tag{21}$$

$$= \arg\min_a H(X)_{k+1|a} \tag{22}$$

$$= \arg\min_a \log\left(\sqrt{(2\,\pi\,e)^N\left|\Sigma_{k+1|k+1,a}\right|}\right) \tag{23}$$

$$= \arg\min_a \left|\Sigma_{k+1|k+1,a}\right| \tag{24}$$

$$= \arg\min_a \left|\mathbf{I} - (\Sigma_{k|k} + \mathbf{Q})(\Sigma_{k|k} + \mathbf{Q} + \mathbf{R}_{k+1|a})^{-1}\right| \tag{25}$$

3.4 Future Directions

There are several ways to continue from here and multiple problems we are or will be researching. Firstly, the choice of set of actions $\mathcal{A}$ is important to get the best results. If for instance it consists of 'turn head n degrees left/right' the agent may focus on a single set of landmarks, unwilling to sweep over empty areas, even though that may lead to observing better landmarks.

Secondly, the current model of infotaxis assumes that all information is equally valueble. However, in football this is not the case; information about the location of the ball may be worth more than where your keeper is. Moreover, the relative value of different types of information could change during a game. A tradeoff has to be made to decide on which target to focus, e.g. by limiting $\mathcal{A}$ to actions that will not loose sight of the ball or by alternating between the targets based on the current value of the information about each to the agent. To optimize the latter case we will use another information theoretical principle, *relevant information*, which measures the amount of information present in a random variable that is relevant for an agent's optimal strategy [15]. This amount gives an indication which variable should get more attention.

References

1. S. Buck and M. Riedmiller. Learning situation dependent sucess rates of actions in a robocup scenario. In *Proceedings of PRICAI '00, Melbourne, Australia, 28.8.-3.9.2000*, page 809, 2000.
2. Y. Engel, S. Mannor, and R. Meir. Bayes meets bellman: The gaussian process approach to temporal difference learning. In *Proc. of ICML 20*, pages 154–161, 2003.
3. Santiago Franco and Daniel Polani. Skill learning via information-theoretical decomposition of behaviour features. In Daniel Polani, Brett Browning, Andrea Bonarini, and Kazuo Yoshida, editors, *RoboCup 2003: Robot Soccer World Cup VII*, volume 3020 of *LNCS*. Springer, 2004. Team Description (CD supplement).
4. Ambarish Goswami and Vinutha Kallem. Rate of change of angular momentum and balance maintenance of biped robots. *International Conference on Robotics and Automation*, 2004.
5. David Jacob, Daniel Polani, and Chrystopher L. Nehaniv. Legs that can walk: Embodiment-based modular reinforcement learning applied. In *IEEE Computational Intelligence in Robotics & Automata (IEEE CIRA 2005)*, pages 365–372. IEEE, 2005.
6. Tobias Jung and Daniel Polani. Sequential learning with ls-svm for large-scale data sets. In *Proc. 16th International Conference on Artificial Neural Networks, 10-14. September 2006, Athens, Greece*, volume 2, pages 381–390, 2006.

7. Tobias Jung and Daniel Polani. Kernelizing lspe(λ). In *Proc. 2007 IEEE International Symposium on Approximate Dynamic Programming and Reinforcement Learning, April 1-5, Hawaii*, pages 338–345, 2007.
8. Tobias Jung and Daniel Polani. Learning robocup-keepaway with kernels. In Neil Lawrence, Anton Schwaighofer, and Joaquin Quionero Candela, editors, *Gaussian Processes in Practice*, volume 1 of *JMLR Workshop and Conference Proceedings*, pages 33–57, 2007.
9. Kalman, Rudolph, and Emil. A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, 82(Series D):35–45, 1960.
10. Alexander Klyubin, Daniel Polani, and Chrystopher Nehaniv. Representations of space and time in the maximization of information flow in the perception-action loop. *Neural Computation*, 19(9):2387–2432, 2007.
11. Alexander S. Klyubin, Daniel Polani, and Chrystopher L. Nehaniv. Keep your options open: An information-based driving principle for sensorimotor systems. *PLoS ONE*, 3(12):e4018, Dec 2008.
12. Martin Lauer and Martin Riedmiller. An algorithm for distributed reinforcement learning in cooperative multi-agent systems. In *Proc. 17th International Conf. on Machine Learning*, pages 535–542. Morgan Kaufmann, San Francisco, CA, 2000.
13. J. Morimoto, G. Endo, J. Nakanishi, S. Hyon, G. Cheng, and D. Bentivegna. Modulation of simple sinusoidal patterns by a coupled oscillator model for biped walking. In *Proceedings of the 2006 IEEE international conference on robotics and automation*, pages 1579–1584, 2006.
14. Lars Olsson, Chrystopher L. Nehaniv, and Daniel Polani. From unknown sensors and actuators to actions grounded in sensorimotor perceptions. *Connection Science*, 18(2):121–144, 2006. Special Issue on Developmental Robotics, Douglas Blank and Lisa Meeden, editors.
15. Daniel Polani, Thomas Martinetz, and Jan T. Kim. An information-theoretic approach for the quantification of relevance. In *ECAL '01: Proceedings of the 6th European Conference on Advances in Artificial Life*, pages 704–713, London, UK, 2001. Springer-Verlag.
16. Peter Stone. *Layered Learning in Multiagent Systems: A Winning Approach to Robotic Soccer*. MIT Press, 2000.
17. Peter Stone, Richard S. Sutton, and Gregory Kuhlmann. Reinforcement learning for robocup-soccer keepaway. *Adaptive Behavior*, 13(3):165–188, 2005.
18. Peter Stone and Manuela Veloso. A layered approach to learning client behaviors in the robocup soccer server. *Applied Artificial Intelligence*, 12, 1998.
19. Massimo Vergassola, Emmanuel Villermaux, and Boris I. Shraiman. 'infotaxis' as a strategy for searching without gradients. *Nature*, 445:406–409, 2007.
20. Massimo Vergassola, Emmanuel Villermaux, and Boris I. Shraiman. 'infotaxis' as a strategy for searching without gradients. *Nature*, 455:406–409, 2007.
21. Greg Welch and Gary Bishop. An introduction to the kalman filter. Technical report, Chapel Hill, NC, USA, 1995.