Reinforcement Learning for a Self-Improving Walking Pattern with Force Sensors in RoboCup

Fabian Fiedler

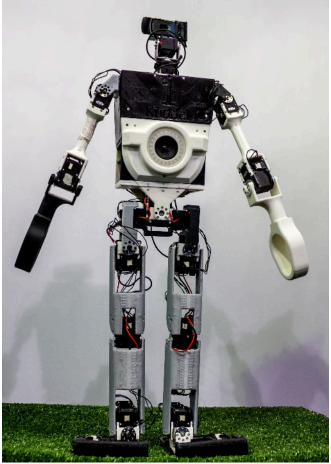
Universität Hamburg
Fakultät für Mathematik, Informatik und Naturwissenschaften
Fachbereich Informatik
Technische Aspekte Multimodaler Systeme

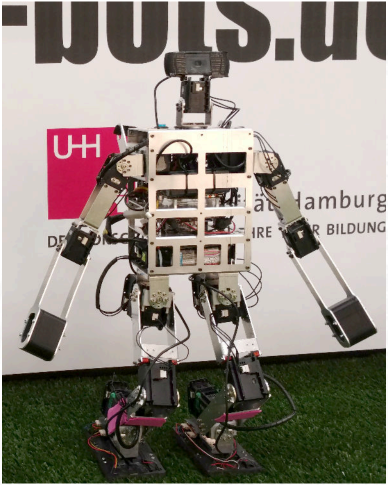
13 December 2016

Motivation

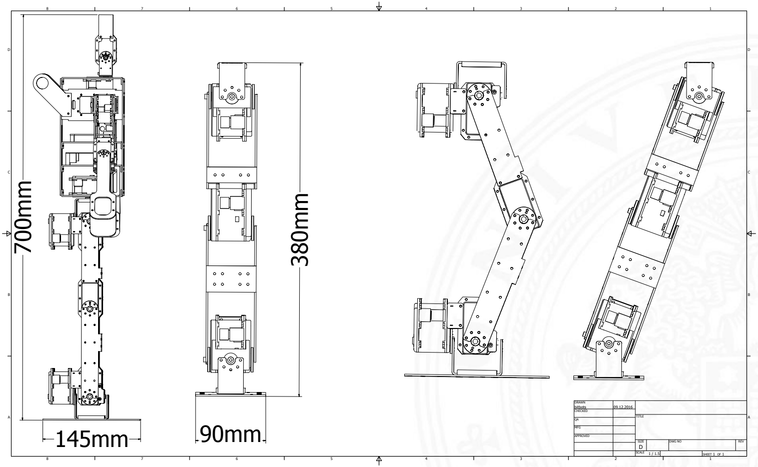
- The Hamburg Bit-Bots compete in the Robocup humanoid Kid-Size league
- 2050 Goal: to defeat the human world champion soccer team in a fair game according to the FIFA rules
- Biped Walking is still one of the most challenging topics in RoboCup

New Robots





Minibot



Motors

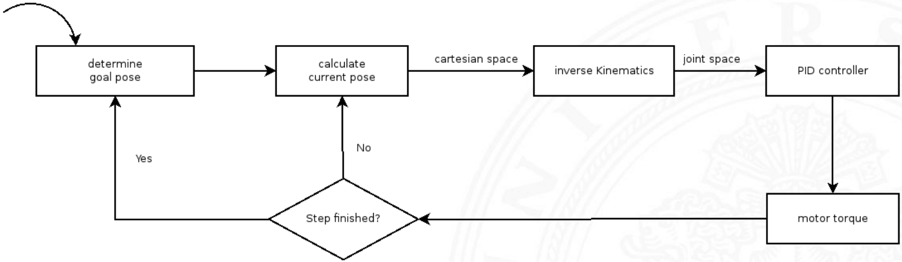
MX-106T/ MX-106R/ EX-106+

	MX-106T/ MX-106R			EX-106+
Weight	153 g (5.39 oz)			154 g (5.43 oz)
Dimension(mm) / (inch)	40.2×65.1×46(mm) 1.58×2.56×1.81(inch)			40.2×65.1×46(mm) 1.58×2.56×1.81(inch)
Gear Ratio (material)	225 : 1 (metal)			184 : 1 (metal)
Network Interface	TTL / RS-485			RS-485
Position Sensor (Resolution)	Contactless Absolute Encoder (360°/ 4096)			Magnetic Encoder (251°/ 4096)
Motor	Maxon Motor			Maxon Motor
Operation Voltage (V)	11.1	12.0	14.8	12~18.5
Stall Torque (N.m)	8.0	8.4	10.0	8.0 at 14.8V
Stall Current (A)	4.8	5.2	6.3	6.1
No Load Speed (RPM)	41	45	55	69

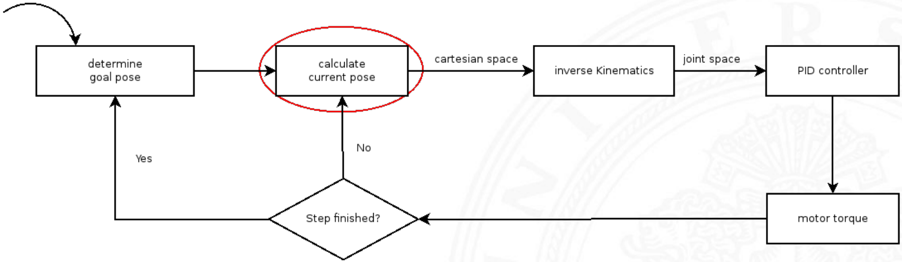
- High level PID-position-control possible
- However oddly implemented by Robotis, mostly only a P controller used
- The Stall Torque is hardly achieved by the servos, realistic is half of it

[Robotis]

Walking Flowchart



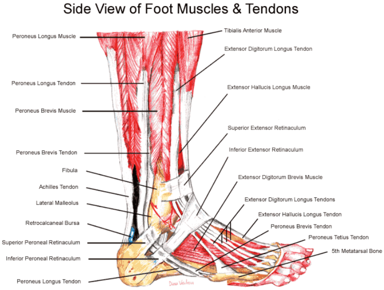
Walking Flowchart



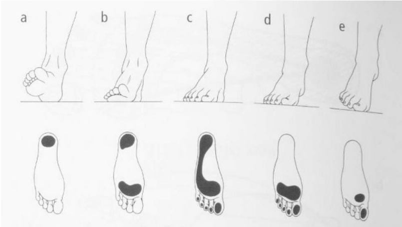
Inverse Kinematics (IK)

- Calculation of the motor angles of the robot for a given goal position of its limbs and torso
- For DARwIn-OP-like robots there is a well working analytic IK-Solver by the team NUbots from Newcastle
- It is enough for a walking pattern to generate the position of the legs and torso of the robot

Human Gait

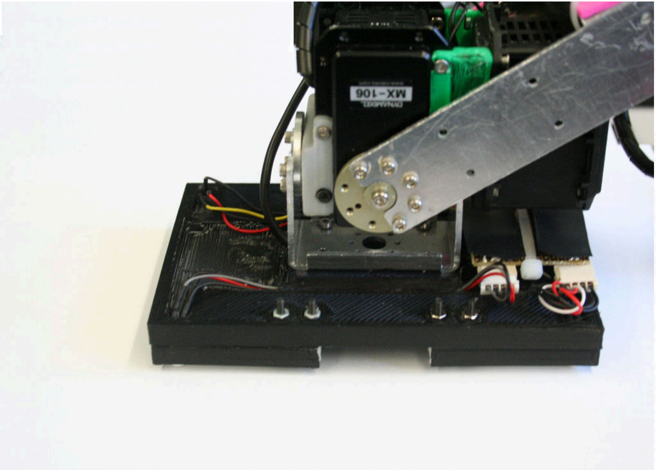


[Swierzewski]

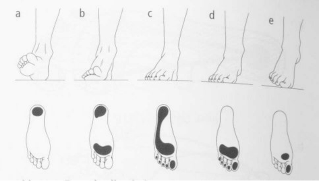


[Debrunner]

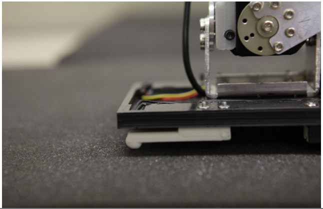
Foot of a robot



Robot Gait



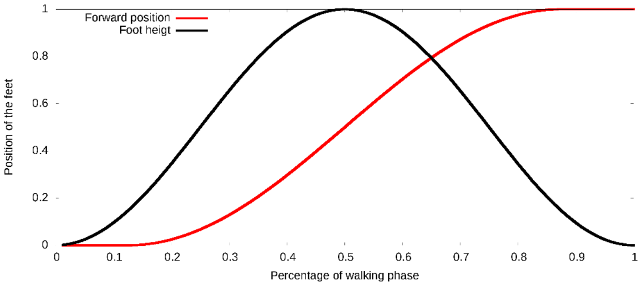
[Debrunner]



- No human-like rolling motion possible
- Foot can be parallel to the ground at all time

Foot Movement

- Motors are torque controlled, so the movement of the foot should be two-times derivable
- On artificial turf the robot should first lift his foot before moving it forward
- $cos(x + \pi) + 1$, from 0 to π for the forward/sideward movement, from 0 to 2π for the height



Stabilizing the gait

- Without control of the torso the robot will fall over
- Usually the movements of the torso are highly manually tuned and the tuning is very tedious
  - Team Darwin-Walking: 20 configuration values
  - Nimbro-OP-Walking: 60 (140) configuration values
- An optimal or self-improving control would be more suitable

A stable upright posture

- A still standing robot with no external force is stable if the Center of Mass (COM) is inside the convex hull of its feet.

$$COM = \frac{1}{\sum_{i=1}^{n} m_i} \sum_{i=1}^{n} m_i r_i$$

- where m_i is the mass of the i part and r_i its position
- In case of high friction between surface and foot the Zero-Moment-Point (ZMP) can be referred as a generalization of stability

Zero-Moment-Point

- "The Zero-Moment-Point is that point on the ground at which the net moment of the inertial forces and the gravity forces has no component along the horizontal axes." [Dasgupta]

$$\boldsymbol{OD} = \frac{mgz \times \boldsymbol{OG} \times z + z \times \dot{\boldsymbol{H}}_G}{mg + m\boldsymbol{a}_g \cdot z} \tag{1}$$

Where $\boldsymbol{OD}$ represents the vector from an origin O to the ZMP D, m is the total mass, $\boldsymbol{OG}$ reaches from the origin to the Center of Mass. z denotes the z-axis, i.e. the normal vector of the ground. $\dot{\boldsymbol{H}}_G$ is the angular momentum around the Center of Mass G and $\boldsymbol{a}_g$ is the acceleration of of the Center of Mass G. [Sardain]

Optimal Walking

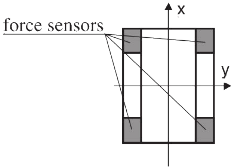
$$\min \int ||\boldsymbol{OD} - \boldsymbol{OZ}||^2 dt \tag{2}$$

- Where $\boldsymbol{OD}$ denotes the vector to the current ZMP and $\boldsymbol{OZ}$ denotes the vector to the optimal ZMP [Dasgupta99]
- Generally $\boldsymbol{OD}$ should be in the center of the foot
- Suitable as reward/objective function

Calculation of the ZMP

- $\boldsymbol{OD} = \frac{mgz \times \boldsymbol{OG} \times z + z \times \dot{\boldsymbol{H}}_G}{mg + m\boldsymbol{a}_g \cdot z}$
- Especially the angular momentum $\dot{\boldsymbol{H}}_G$ is not computable without deep knowledge of the current torque of each motor
- The torque of a motor is depending on the temperature, the quality of the gears, the stability of the screws et cetera
- It is therefore neither possible to calculate the ZMP correctly nor simulate it

Center of Pressure



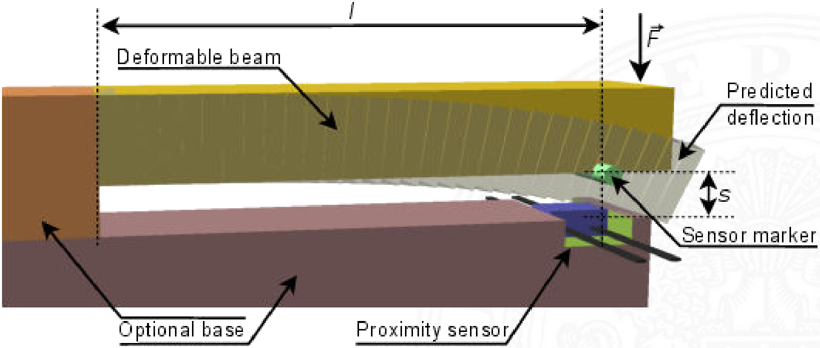
[Vukubratovic]

- The pressure between the foot and the ground can be replaced by a force acting at the Center of Pressure.
- In case of a stable gait COP and the ZMP match [Sardain]
- With a flat surface and an inflexible foot it is sufficient to add a sensor in each corner of the feet to calculate the COP [Vukubratovic]

$$\boldsymbol{OD} = \frac{\sum_{i=1}^{4} f_i r_i}{\sum_{i=1}^{4} f_i}$$

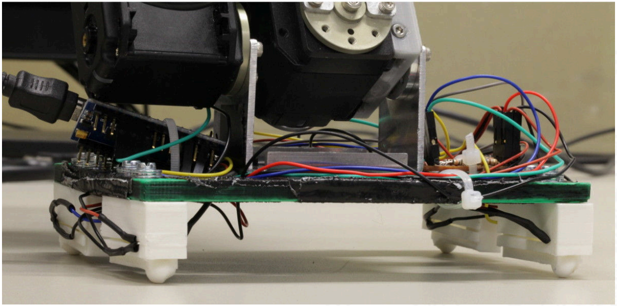
- Where f_i is the pressure on the sensor and r_i is the position.

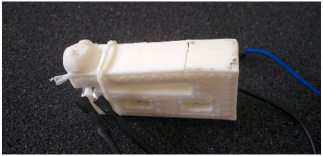
Design of the Force Sensors



[Wasserfall]

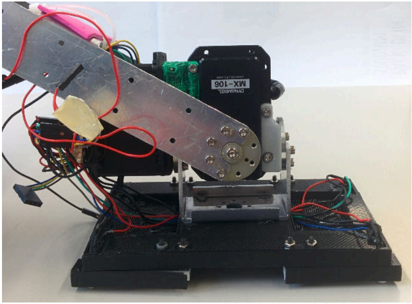
First Prototype

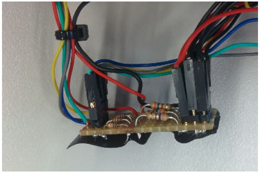




- Cables easily ripped off the sensors
- The deformable beam tends to break

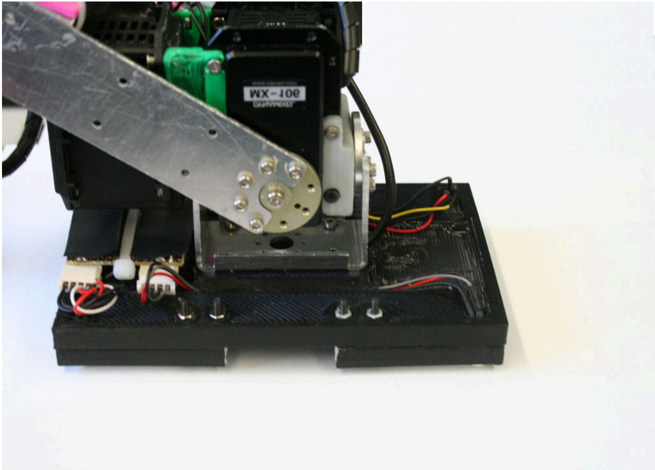
Second Prototype



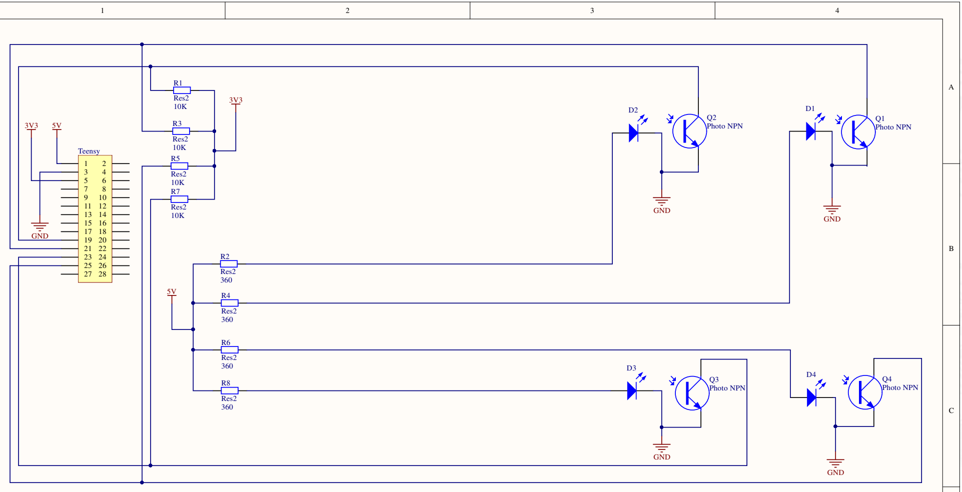


- The cable routing is a mess
- Connectors tend to be too loose, thus to slip out

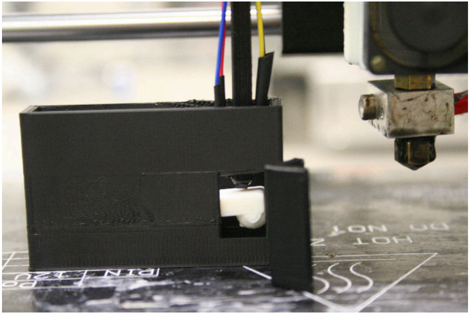
Third Prototype

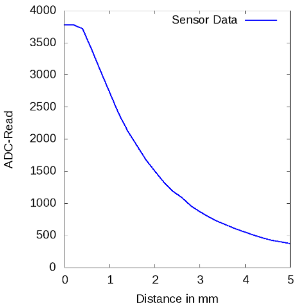


Schematics

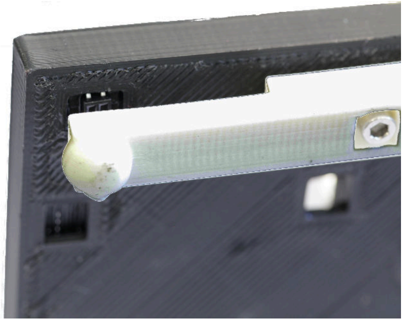

1
2
3
4
A
B
C
R1
Res2
10K
R3
Res2
10K
R5
Res2
10K
R7
Res2
10K
3V3
3V3
5V
Teensy
GND
R2
Res2
360
R4
Res2
360
5V
R6
Res2
360
R8
Res2
360
D2
Q2
Photo NPN
GND
D1
Q1
Photo NPN
GND
D3
Q3
Photo NPN
GND
D4
Q4
Photo NPN
GND

Sensor Calibration



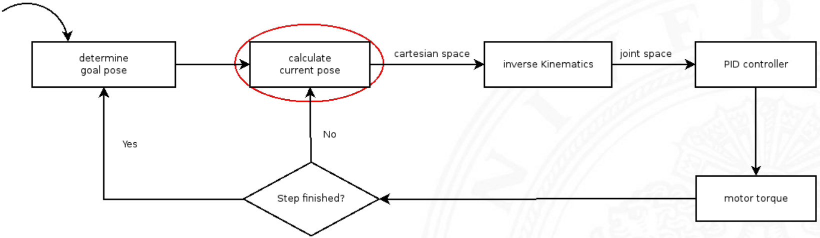


Hooke’s Law

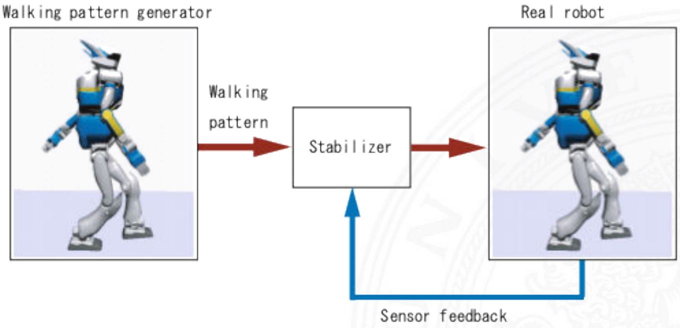


- Hooke’s Law
  - $F = k \cdot x$
  - F is the force, k the Hookes constant, x the deformation

Control system



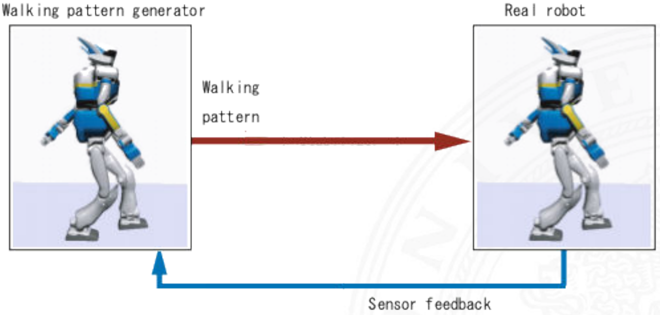
Control system



[Kajita]

- The history of feedback is not saved over multiple runs

Control system



[Kajita, modified]

- Improvement of Walking Pattern

Reinforcement Learning

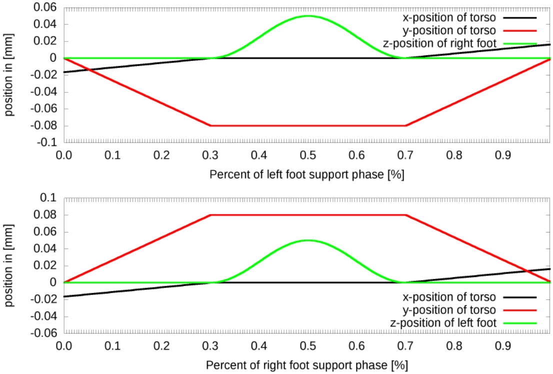
- "Reinforcement Learning is learning what to do - how to map situations to actions - so as to maximize a numerical reward signal. [...] The learner [...] must discover which actions yield the most reward by trying them."[Sutton98]
- State s, action a, policy $\pi : s \rightarrow a$, reward function $r : s, a \rightarrow \mathbb{R}$ value function $V : s, a \rightarrow \mathbb{R}$

Reinforcement Learning

- Learning on real hardware needs a very fast convergence – a small learning space is better
- Trajectory of the feet of the robot is given
- Trajectory of the torso should be optimal with respect to the ZMP
- State s: percentage of walking phase
- Action a: goal position of the torso
- Morimoto et al. argues that one can consider the x-axis (sagittal) and y-axis (lateral) disjunct from each other [Morimoto07]

1D-tile coding

- "Tile coding is a piecewise constant function approximation" [Whiteson05]



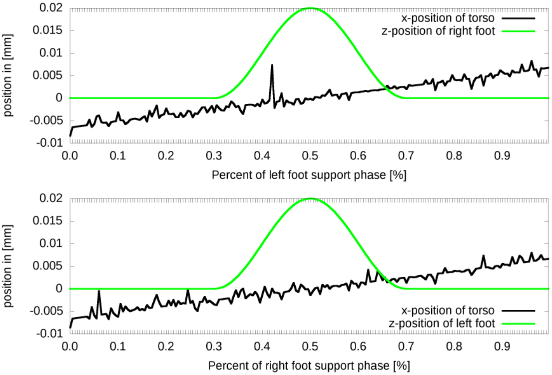
Reinforcement Learning

- State s: percent of the walking phase
- Action a: goal position of the torso
- Reward r : difference from actual ZMP to desired position
- Calculation of a value function is hardly feasible
- Policy Search Reinforcement Learning seems better
- No random exploration leads to reduced chance of falling and damage

Policy Gradient Methods

- Policy Gradient Methods are widely used for Policy Search Reinforcement Learning
- Easiest black-box method is the finite difference policy gradient [Deisenroth]
  - $\nabla E(R(\pi)) = (\delta\Theta^T\ \delta\Theta)^{-1}\delta\Theta^T\ \delta R$
  - where $\delta\Theta$ is an list of added perturbations, in this case out of the Gaussian distribution
  - δR is the list of observed differences in the reward

Finite Difference Policy Gradient

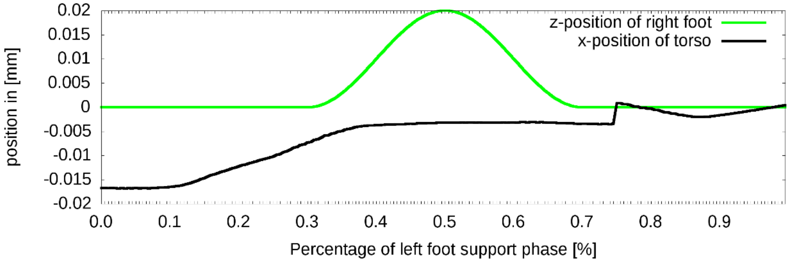


- Signal is to noisy for Finite Differences Methods

Gradient Estimation

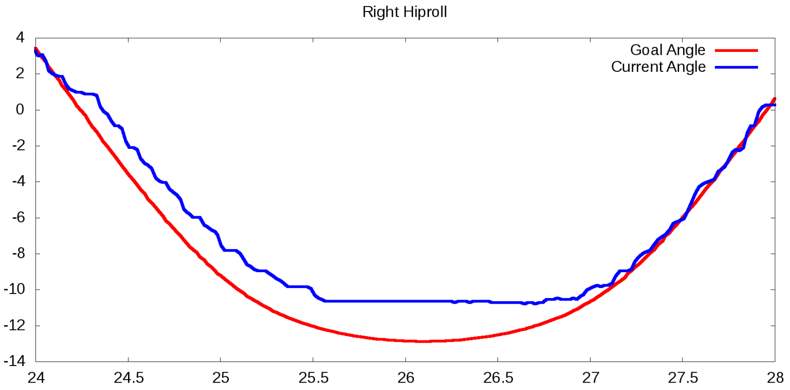
- There are better black-box gradient estimators like REINFORCE [Williams90] but they are likely to have the same problem
- However the position of the ZMP itself is a gradient estimator for the expected reward
- Goal-position of the torso influences not only the next step but beyond $\rightarrow$ Temporal Difference Learning $TD(\lambda)$(cf. [Sutton98])

$TD(\lambda)$

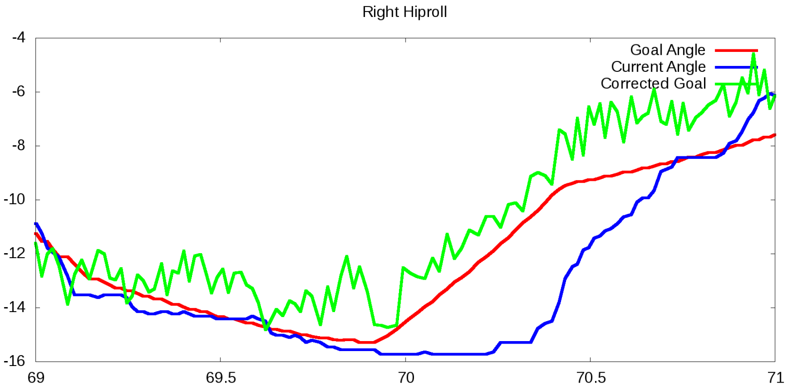


- Glitch occurred due to a broken sensor
- However the history of actions should be similar to a two times derivable function
- Gaussian smoothing of the $TD(\lambda)$-learning over the actions prevents jumps in the actions

Current Challenge

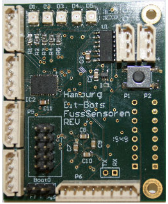


Current Challenge



Future Work

- ROS → Robot specification as URDF for an Inverse-Kinematic-Solver, e.g. *KDL* or *IKFast*, should lead to a stable walking for any given humanoid



- Each foot needs an own USB cable
- Would support the Dynamxiel-TTL-Bus
- Would support strain gauge based load cells

Live Demo

Live Demo

Bibliography

Robotis	`http://en.robotis.com/index/product.php?cate_code=101010` [Retrieved 27.12.16
Swierzewski	`www.healthcommunities.com/foot-anatomy/muscles-tendons-ligaments.shtml` [Retrieved 27.12.16]
Debrunner	Debrunner Alfred M.: Orthopädie. Orthopädische Chirurgie. Patientenorientierte Diagnostik und Therapie des Bewegungsapparates. 2002.
Vukubratovic	Vukobratović, Miomir and Borovac, Branislav: Zero-Moment Point - Thirty Five Years of its Life in International Journal of Humanoid Robotics 2005
Sardain	P. Sardain and G. Bessonnet: Forces acting on a biped robot. Center of pressure-zero moment point in IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans 2004

Bibliography

Wasserfall	Florens Wasserfall, Norman Hendrich, Fabian Fiedler and Jianwei Zhang: 3D-Printed Low-Cost Modular Force Sensors
Kajita	Kajita et al. Introduction to Humanoid Robotics. 2014
Sutton98	Richard S. Sutton and Andrew G. Barto: Reinforcement Learning an Introduction. 1998
Morimoto07	J. Morimoto and C. G. Atkeson: Learning Biped Locomotion in IEEE Robotics Automation Magazine 2007
Whiteson	Shimon Whiteson et al. Adaptive Tile Coding for Value Function Approximation in Proceedings of the 6th International Conference on Abstraction, Reformulation and Approximation. 2005
Deisenroth	Marc Peter Deisenroth, Gerhard Neumann, Jan Peters: A Survey on Policy Search for Robotics in Foundations and Trends in Robotics. 2013