Toward High-Performance 24/7 Cognitive Humanoids

Tamim Asfour

November 30, 2011

Abstract

Recently, considerable progress has been made towards the realization of humanoid robot systems which are able to move in a human-like way and perform tasks in human-centered environment. However, current systems are still limited in their actuation, sensing, prediction, interaction and learning capabilities. We define High-Performance Humanoids as integrated complete humanoid robot systems able to act, interact, predict and learn in 24/7 manner in the real world and to perform a wide variety of tasks.

1 State of the Art

In recent years there are renewed efforts to develop robot systems that can perceive, move and perform actions. An encouraging spectrum of many isolated elements in the area of cognitive systems has been realized with a focus on performance in well-defined, narrow domains. The development of cognitive robots relies on artificial embodiments having complex and rich perceptual and motor capabilities. This leads to robots with rich sensorial inputs and complex actions necessary to develop higher cognitive processes. These aspects are, thus, particularly supported by humanoid robots ([7, 6, 3, 2, 5, 14, 11, 16, 13, 1, 12, 4, 15, 10, 8]), i.e. (embodied) robots that perceive, move, and perform diverse actions, which are often acquired by learning techniques.

Although current systems are technologically advanced, they are not able to learn in an open-ended way and their behaviours and lifelong learning capabilities are limited. Successful attempts in building complete systems are still limited to systems designed for "sunshine" environments with limited scope and simple tasks in a given scenario. The transferability of the developed skills and abilities to varying contexts and tasks without a costly redesign of specific solutions is still impossible. Complete robot systems integrating perception, action, planning and lifelong learning capabilities, which are necessary to interact with the environment, as well as a variety of functionalities which are needed to carry out diverse tasks in real environments are still missing.

Recent examples of complete systems which are able to perform a variety of tasks are the Willow Garage Personal Robot PR2, a wheel-driven robot with two arms which is able to perceive its environment detect wall outlets and plug itself in for recharging. The Twendy-one robot (see [8]) developed at the Waseda University in Tokyo posses a wide range of capabilities in human environments such as carrying a tray, fetching objects from the refrigerator, etc. Similar tasks have been presented using the HRP-2 humanoid robot series in [12]. The humanoid robot ARMAR-III (see [1]) is endowed with a variety of capabilities to perform tasks in a kitchen environment such as grasping daily objects, loading the dishwasher and fetching objects for the refrigerator as well as learning various behaviours from human observation. On the humanoid robot DB at ATR, Japan, various behaviours have been demonstrated such as paddling a single ball on a racket, learning a folk dance by observing a human perform it, drumming synchronized to sounds the robot hears (karaoke drumming), juggling three balls, performing a Tai Chi exercise in contact with a human, and various oculomotor behaviours [2]. However, although these robots can perform different tasks, they do not possess the ability of autonomous, lifelong learning, which is crucial for robots with the ambition to operate in human living spaces. Although learning has been employed to acquire single tasks, the applied learning techniques were specialized and did not consider the problem of lifelong learning from sensorimotor experience.

2 Challenges and Research Roadmap

Today, humanoid robots can be considered as highly advanced mechatronics systems with complex and rich sensorimotor capabilities. Thus, such systems are as the most suitable experimental platform for studying human behaviors and cognitive information processing. In the following some of the challenges towards the realization of high-performance humanoids are briefly discussed.

2.1 New Bodyware for Humanoids

Research efforts related to the development of high-performance humanoid robots must address the question of how the body morphology must support processes and representations for emergence of cognitive capabilities.

- The design of humanoid robots with human-like capabilities requires a new thinking regarding design mechanisms, materials and control. Novel technologies and methodologies are needed for the development of compliant, high-performance and energy efficient actuators, sensor technologies (in particular skin), soft materials, as well as dynamically reconfigurable software and hardware architectures and high density lightweight power sources.
- Investigation of design principles and quantitative models for the development of systems that 1) explore their own sensorimotor primitives and body morphology 2) explore the environments and the effective interaction with it and 3) predict the body dynamics and the physics of the world.

- How body morphology allows to cope with morphological change arising through the interaction with the environment and tolerance to uncertain variability in performance of single robot components.
- How reconfigurability and self-reconfigurablility, redundancy, robustness and flexibility in technical systems can be implemented.

2.2 Objects, Actions and Prediction

To deal with problems on perception and action researchers in the late 80s introduced two new frameworks with parallel efforts, in the field of Computer Vision and the field of AI/Robotics, under the headings of Active Vision (Animate, Purposive, Behavioural) and Behaviour-based robotics respectively. In both formalisms, the old idea of conceiving an intelligent system as a set of modules (perception, action, reasoning) passing results to each other, was replaced by the new idea of thinking of the system as a set of behaviours. Behaviours are sequences of perceptual events and actions. These efforts still go on, but have only led to limited success. One reason for that is that such perceptual events involve recognition which is such a hard problem that it prevented the new formalisms from making a breakthrough. A further reason for failure is that the behaviours were never meant to involve objects into the action (e.g. for recognition). A third reason was that no one managed to formulate any theory for behaviour-based robotics. Hence, it was impossible to predict how the systems developed would scale up and deal with new situations.

Human understanding of objects is essentially multi-sensorial. It develops during an intensive exploration making use of visual and haptic information. Therefore, the cross-connection between haptic and vision must be analysed. The gained knowledge would play a key role in modelling multi-sensorial processes in artificial cognitive systems, which then can develop a more holistic understanding of the perception-action coupling and thus objects and actions. In traditional information theory, the environment only plays the part of a passive, undirected disturbance (for example also in closed loop control theory) negatively affecting the input-to-output transfer characteristics of a system. Here we propose, instead that the environment should be treated as an active component. It is active through "my own actions" (the actions of ego) and those of "the others" (the actions of alter), which feed back to ego. Thus, traditional information theory is not sufficient to describe the interaction of an agent with its world correctly. Instead this problem needs to be addressed in a closed loop paradigm where ego acts in its environment and observes the consequences of its actions (in interrelation also with alter). This notion has entered modern robotics theories by the qualitative term "rootedness", which refers to the necessity to embed an artificial acting agent in an environment. Thus, instead of using the conventional I/O paradigm, new approaches should introduce so called encased closed loop situations defined by the mutual interactions of an organism (ego) with its environment. An encased closed loop describes a conventional sensor-motor feedback control loop but with an active environment monitored from the perspective of ego. This represents a central shift of paradigm and follows a constructivist's viewpoint where the environment becomes an integral part of the system's description. This notion goes clearly beyond the conventional concept of a perception-action loop. It embeds the agent into its environment and into its social group by the same formalism. On the side of theory this will lead to intrinsically consistent and technologically applicable measures of "autonomy", "contingency", and "complexity" of agent-world- as well as agent-agent interactions, resulting in the first steps towards an information theory of encased closed loops.

Research into cognitive robots should combine the study of perceptual representations that facilitate motor control, motor representations that support perception, and learning based on actively exploring the environment and interacting with people that provides the constraints between perception and action. This will then allow, e.g., to learn the actions that can be carried out on and with objects when making use of the interplay of different sensorial modalities, such as vision, haptics and acoustics. Action-centred cognition presupposes that artificial cognitive systems will be equipped with eyes, sophisticated haptic sensors for its end-effectors and microphone-ears. This allows for efficient interaction with the world making use of the full potential of multi-sensorial representations.

2.2.1 Object-Action Complexes

The European project PACO-PLUS (Perception, Action and Cognition through Learning of Object-Action Complexes, www.paco-plus.org) has introduced the concept of Object-Action Complexes (OACs) to emphasize the notion that for a cognitive agent objects and actions are inseparably intertwined and that categories are therefore determined (and also limited) by the action an agent can perform and by the attributes of the world it can perceive (see [9]). The resulting OACs (pronounsed "oaks" are the entities on which cognition develops (action-centered cognition). Entities "things" in the world of a robot (or human) will only become semantically useful objects through the action that the agent can/will perform on them.

The OAC concept is based on, but extends the Gibsonian concept of "affordance". In contrast to constructivist approaches, Gibson claimed that objects and events in our environment provide an actor all the information they need about the actions they "afford". This claim was motivated by the idea that perception is not a static process, but rather is a temporally extended act of information acquisition. In other words, the affordances our environment provides are revealed by actively exploring it.

PACO-PLUS has made significant use of the Gibsonian approach in two respects. First, PACO-PLUS has made the robot an information-seeker that is actively experimenting with the objects it is facing in order to find out more about its perceptual features and the action opportunities they provide. In other words, the PACO-PLUS agent is no longer a passive knowledge receiver but an *active explorer*. Second, the idea that active information acquisition reveals the objective structure of our environment makes it possible to ground cognitive representations. Once objective environmental information about sensorimotor opportunities is encoded in the lowest level OACs, these OAC level provide a reliable basis for forming higher-level, more abstract representations suitable for reasoning and action planning, while still being grounded in the robot's sensorimotor experience.

2.3 Representations

Building humanoid robots able to learn to operate in the real world and to interact and communicate with humans, must model and reflectively reason about their perceptions and actions in order to learn, act, predict and react appropriately. Such capabilities can only be attained through physical interaction with and exploration of the real world and requires the simultaneous consideration of perception and action. Representations built from such interactions are much better adapted to guiding behaviour than human crafted rules and allow situated and embodied systems, such as humanoid robots in human-centered environments, to gradually extend their cognitive horizon. Such representations should allow for learning and extending representation in ways that transform intractable problems into tractable ones and support generalization and knowledge transfer between different cognitive systems. These representations should take into account space and motion, objects (things that move) and actions, properties and affordances, goals, plans, beliefs and desires, communication, and models of other minds. In this context several research questions must be addressed.

- How to extend and improve exploration-based and stimulus-driven knowledge acquisition?
- How to define the actual algorithmic mechanisms by which an agent can generalize knowledge across domains leading to a generative extension of its experience?
- How to embed these two mechanisms in a dynamically stable process to drive the extension of knowledge in a generative way while interacting with its environment and other agents (humans)?
- How to allow the agent to predict its own perception-action loops, but also - importantly - the actions of other agents, leading to advanced abilities to cooperate, interact and communicate?
- How to integrate exploration-based and generative inside-out processes into an advanced, complete embodied cognitive system?

3 Examples for Research Challenges

3.1 High-Performance 24/7 Humanoid for daily life

The challenges of creating high-performance 24/7 humanoid robots can be summarized as follows

- Understanding and interpretation of scenes, contexts and situations
- Categorization of daily objects
- Grasping and manipulating any object (Pin, book, . . . , beer box)
- Navigation in every environment (Home, street, super market, etc.)
- Human-Robot interaction
  - Multimodal interaction
  - Physical interaction
  - Natural communication
  - Action and activity and intention recognition
  - Human tracking, gesture detection, face detection and identification, emotion recognition
- Social interaction (Humor, trust, privacy)
- Personalization: Adapt to humans needs and habits.

What to measure?

Criterion	
Energy consumption	Similar to other household appliances (oven, fridge, dishwasher, etc.)
Program complexity	FLOPs, Memory requirements
Performance	2015: set/clean the table, load the dish washer or the washing machine, prepare food. 2030: Clean the apartment, go shopping (in super market, shopping center, Italian shop, etc.) 2049: Similar to human caregiver in performance and social interaction
Price	Cheap car

3.2 High-performance humanoid robot that can play tennis

It is not about tennis but about the following scientfic and technological challenge:

- Understanding the body dynamics, body balancing and motor coordination
- Safe falling and recovery
- Real-time prediction: reaction based on vision would be too late. *Sense-Plan-Act* would not work. Instead **Predict-Act-Sense**.

- Learning of others behavior and adaptation of own behavior based on past experience and learnign to predict and adapt from little experience and few examples.
- Multisensory integration (vision, vestibular, haptics, ..)
- High speed perception and high speed control.

What to measure?

Criterion	
Energy consumption	Humanoid robot should be able to play a game with the energy equivalent of a ”Maultaschen” dish.
Program complexity	FLOPs, Memory requirements
Performance	2020: Perform basic tennis playing 2030: Steadily win against number 500 of the ATP ranking 2049: Steadily win against number one of the ATP ranking
Price	Cheap car

References

[1] T. Asfour, K. Regenstein, P. Azad, J. Schröder, N. Vahrenkamp, and R. Dillmann. ARMAR-III: An Integrated Humanoid Platform for Sensory-Motor Control. In *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, pages 169–175, Genova, Italy, December 2006.

[2] C. G. Atkeson, J. Halle, F. Pollick, M. Riley, S. Kotosaka, S. Schaal, T. Shibata, G. Tevatia, A. Ude, S. Vijayakumar, and M. Kawato. Using Humanoid Robots to Study Human Behavior. *IEEE Journal on Intelligent Systems*, 15(4):46–55, July/August 2000.

[3] R.A. Brooks and et al. The Cog Project: Building a Humanoid Robot. In *The First International Conference on Humanoid Robots and Human friendly Robots*, Tsukuba, Japan, Oktober 26 - 27 1998.

[4] G. Cheng, S.-H. Hyon, J. Morimoto, A. Ude, G. Colvin, W. Scroggin, and S.C. Jacobsen. Cb: A humanoid research platform for exploring neuroscience. In *6th IEEE-RAS International Conference on Humanoid Robots*, pages 182 –187, dec. 2006.

[5] G. Cheng, A. Nagakubo, and Y. Kuniyoshi. Continuous Humanoid Interaction: An Integrated Perspective - Gaining Adaptivity, Redundancy, Flexibility - In One. *IEEE Journal on Robotics and Autonomous Systems*, 37(2-3):161–183, November 2001.

[6] K. Hirai, M. Hirose, and Y. Haikawa nad T. Takenaka. The Development of Honda Humanoid Robot. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pages 1321–1326, Leuven, Belgium, May 1998.

[7] K. Hirai, M. Masato, Y. Haikawa, and T. Takenaka. The Development of Honda Humanoid Robot. In *IEEE International Confernce on Robotics and Automation*, pages 1321–1326, Leuven, Belgium, May 1998.

[8] H. Iwata and S. Sugano. Design of human symbiotic robot TWENDY-ONE. In *IEEE International Conference on Robotics and Automation*, pages 580 –586, may 2009.

[9] N. Krüger, Ch. Geib, J. Piater, R. Petrick, M. Steedman, F. Wörgötter, A. Ude, T. Asfour, D. Kraft, D. Omrcen, A. Agostini, and R. Dillmann. Object-Action Complexes: Grounded abstractions of sensory-motor processes. *Robotics and Autonomous Systems*, 59(10):740–757, 2011.

[10] G. Metta, G. Sandini, D. Vernon, L. Natale, and F. Nori. The iCub humanoid robot: an open platform for research in embodied cognition. In *PerMIS: Performance Metrics for Intelligent Systems Workshop*, aug. 2008.

[11] T. Miyashita and H. Ishiguro. Natural behavior generation for humanoid robots. *I. J. Humanoid Robotics*, 1(4):637–649, 2004.

[12] K. Okada, M. Kojima, Y. Sagawa, T. Ichino, K. Sato, and M. Inaba. Vision based behavior verification system of humanoid robot for daily environment tasks. In *Humanoid Robots, 2006 6th IEEE-RAS International Conference on*, pages 7 –12, dec. 2006.

[13] I.W. Park, J.Y. Kim, J. Lee, and J. Oh. Mechanical design of humanoid robot platform khr-3 (kaist humanoid robot 3: Hubo). In *Humanoid Robots, 2005 5th IEEE-RAS International Conference on*, pages 321 –326, dec. 2005.

[14] Y. Sakagami, R. Watanabe, C. Aoyama, S. Matsunaga, N. Higaki, and K. Fujimura. The intelligent ASIMO: system overview and integration. In *IEEE/RSJ International Conference on Intelligent Robots and System*, volume 3, pages 2478–2483, 2002.

[15] G. Sandini, G. Metta, and D. Vernon. 50 years of artificial intelligence. chapter The iCub cognitive humanoid robot: an open-system research platform for enactive cognition, pages 358–369. Springer-Verlag, Berlin, Heidelberg, 2007.

[16] K. Yokoi, F. Kanehiro, K. Kaneko, S. Kajita, K. Fujiwara, and H. Hirukawa. Experimental Study of Humanoid Robot HRP-1S. *I. J. Robotic Res.*, 23(4-5):351–362, 2004.