

An Intuitive Interface for designing Social behaviors based on Human motions

Praveenkumar Vasudevan¹ and Gentiane Venture²

¹ Graduate Student, École Centrale de Nantes, Nantes, France,
praveenv4k@gmail.com

² Associate Professor, Tokyo University of Agriculture and Technology, Japan
venture@cc.tuat.ac.jp

Abstract. Humans interacting with intelligent robots has been seen as a potential game changer of the future. In scenarios where robots coexist with humans in a social environment, understanding not only verbal communication, but also non-verbal communication is extremely inevitable. The non-verbal communication carries information such as intention, emotion and health of a human, that adds value to the way robots participate in an interaction. Additionally, the people who design interaction scenarios are from diverse fields who do not essentially have the required robot programming skills. In this paper we propose an easy to use and intuitive programming paradigm which gives the power to design robot behaviors taking into account human motions. We propose a distributed architecture which gives the capability to plug and play multi-modal motion recognition systems and diverse class of robots. We present results of Nao humanoid robot performing actions understanding human motions using kinect motion recognition system.

Keywords: human robot interaction, motion recognition, robot behaviors, kinect, nao

1 Introduction

Human Robot Interaction (HRI) is a challenging research field at the intersection of psychology, cognitive science, social sciences, artificial intelligence, computer science, robotics, engineering and human-computer interaction [13]. Goodrich [15] in his extensive survey proposed two main types of HRI namely Remote interaction or Teleoperation and Proximate interaction. The latter is particularly important where the humans and the robots are co-located (for example, service robots may be in the same room as humans). Proximate interaction has gained importance due to the successful encounters of putting robots to work with human beings. It has led to the development of a new class of robots called Social Robots. Fong et al. [14] define that social robots are able to recognize each other and engage in social interactions; Breazeal et al. [9] explain that a social robot is a robot which is able to communicate with humans in a personal way; Bartneck and Forlizzi [7] describe that a social robot is an autonomous

or semi-autonomous robot that interacts with humans by following some social behaviors; Hegel et al. [17] define that a social robot is a combination of a robot and a social interface. Summarizing all these Yan et al. [35] defines *A social robot is a robot which can execute designated tasks and the necessary condition turning a robot into a social robot is the ability to interact with humans by adhering to certain social cues and rules.*

The Social robots already entered the human spaces as entertainers[3], educators[3], caring agents[1] and personal assistants[4]. Given that social robotics has emerged as a promising field, designing and developing interaction systems need to be approached in a systematic manner wherein the robots should be able to understand the human motions and intentions in order to interact in a better way. To make it possible it is necessary to develop robotic systems with essential cognitive skills for efficient and natural interaction. Most often the on-board sensors on the robots fail to satisfy this demanding requirement due to various constraints like space, power and computational requirements. Hence consideration of augmenting exteroceptive sensors that are commonly available in the smart home/public environments to this purpose is important.

Another important aspect in HRI is the fact that the users of such systems are from diverse backgrounds. So the tools needed to design behaviors of a social robot should be intuitive and user friendly. With increased availability of social robots and cost effective motion recognition sensors, we could observe a huge void which inhibits the exploitation of available technology for designing human motion driven robot behaviors. More efficient tools are needed which could tackle this issue.

The main contribution of this work will be to develop an application independent experimental platform wherein a social robot will be augmented with essential perceptual ability to understand human motions. The behavior design of such a social robot will be made possible by an easy to use behavior design interface. The experimental platform will be used to design and evaluate the interaction between the social robot and the human driven by his/her motion.

The key problem statements of this work are

- Human Pose estimation and motion recognition
- Localization of the robot
- Behavior design interface

2 Related work

2.1 Human Pose Detection

Vision based motion capture and analysis has been studied widely and a summary of all the approaches developed during the past three decades have been presented in the surveys [25] [26][27]. All these studies have investigated vision based human motion capture and analysis in general, however our particular focus is to use low cost RGB-D sensors like Microsoft Kinect[19] to this purpose.

Vision based human pose estimation has traditionally suffered from the requirement to adopt an initialization pose and losing track after a few frames. These problem has been addressed by the approaches by Xbox[19] team which are capable of accurately predicting the 3D positions of body joints using single depth images without using any temporal information[31][32]. Unlike the approaches used in the Kinect SDK, the approach presented in [10] uses both the depth and color(RGB-D) data for human body detection and pose estimation using a customizable human kinematic model.

Understanding of human motion is not complete if the gesture of the human could not be understood. Hidden Markov models (HMM) which had been widely used for speech recognition[29] also inspired to be used for the gesture recognition applications. In the survey by Microsoft research[16], a background study on various algorithms used for human activity analysis is presented. Recently data-driven machine learning approaches like neural networks, Support vector machines, clustering, decision trees and bayesian networks are being used to this purpose[20].

The localization of humanoid robots is a challenging issue, due to rough odometry estimation, noisy onboard sensing, and the swaying motion caused by walking[11]. The Point cloud library[30] which is one of the most widely used 3D perception software library, implements ready to use probabilistic tracking algorithms[34]. This could be readily used to track an object of known geometry and this information has to be fed to the PCL through a point cloud mesh of the object. Studies on robot localization, obstacle mapping, and path planning in multilevel 3D environments by equipping Nao with a consumer-level depth camera has been reported in [24]. Localization and motion planning in smart home environment using an external Kinect sensor have been proposed in [11]. In [12] a robust particle filter parallelized on a GPU that can track a known 3D object model over a sequence of RGB-D images is proposed. However all these methods are computationally demanding and it could cause overall performance degradation particularly when we want to share the same sensor for both motion recognition and localization of the robot. Tracking rectangular fiducial markers using augmented reality toolkits like ARToolKit[23] can be interesting if we could embed those markers on the humanoid robot. This is one of the simplest and cheapest solution in terms of the computational power as it can provide position and orientation relative to physical markers in real time.

The users of social robots do not have necessary backgrounds in programming and design of robot behaviors. The main challenge in the behavior design is the ability of defining the behavior which can abstract complex data flows from the end user. There exists several visual programming languages which allow non-programmers to create robot applications using a set of pre-built behavioral blocks and connecting them to one another to get the desired flow of action [21],[2]. These programs are very intuitive and allow the users to realize complex sequence of movements and sequential behaviors. However when it comes to designing reactive behaviors taking into account human in-the-loop, the existing visual programming methods increase the cognitive load on the

end users. Specialized robot programming techniques like Task description language[33], scripting techniques like Universal Robotic Body Interface(URBI)[5] and distributed asynchronous architecture like ROS[28] have been proposed in the literature. Hierarchical organization of behaviors and modularity are also being investigated [22],[6],[18]. Though these programs provide modularity, distributed nature ,support multiple sensors and robots, all these tools require high level of skill in robotics and programming skill to use them. Recently a new non-domain-specific solution called *Target Drives Means (TDM)* is proposed in [8] however it lacks an intuitive interface.

2.2 Behavior Design Frameworks

3 System Architecture

4 Implementation

5 Experimentation

6 Conclusions

For the moment we have evaluated our system only for the Kinect motion capture system working seamlessly with the Nao humanoid robot for a set of predefined gestures and robot behaviors. We are planning to develop an extensive database containing commonly encountered gestures and also an extensive set of primitive robot actions. These will be made available to the end user through the intuitive programming interface which could be then used for defining complex motion driven behaviors. Additionally we are also planning to integrate our system to work with other modes of motion recognition like IMU, Accelerometers and Gyroscopes that are available in smart-phones and wearable devices. Similarly we would like to integrate other system with other robots like Pepper which we expect to receive soon.

7 Acknowledgements

We would like to thank the members of GVLab of Tokyo University of Agriculture and Technology for helping us arranging the resources to realize this work and participate in the experimentation.

References

1. Aldebaran. Ask nao: Autism solution for kids. <https://asknao.aldebaran.com>. Accessed: 2014-12-12.
2. Aldebaran. Choregraphe software. <http://www.aldebaranrobotics.com/en/Discover-NAO/Software/choregraphe.html>. Accessed: 2014-11-20.

3. Aldebaran. Nao humanoid robot. <https://www.aldebaran.com/en/humanoid-robot/nao-robot>. Accessed: 2014-11-15.
4. Aldebaran. Project romeo. <http://projetromeo.com/en/scenario>. Accessed: 2015-04-20.
5. J. C. Baillie. Urbi 2: Introduction to concurrent real-time programming. In *Proceedings of the Third International Workshop on Software Development and Integration in Robotics*. IEEE ICRA, May 2008.
6. Gianluca Baldassarre and Marco Mirolli. *Computational and Robotic Models of the Hierarchical Organization of Behavior*. Springer Publishing Company, Incorporated, 2013.
7. C. Bartneck and J. Forlizzi. A design-centred framework for social human-robot interaction. In *RO-MAN 2004. 13th IEEE International Workshop on Robot and Human Interactive Communication (IEEE Catalog No.04TH8759)*, pages 591–594. IEEE, 2004.
8. Vincent Berenz and Kenji Suzuki. Targets-drives-means: A declarative approach to dynamic behavior specification with higher usability. *Robotics and Autonomous Systems*, 62(4):545–555, 2014.
9. Cynthia Breazeal. *Designing Sociable Robots*. MIT Press, Cambridge, MA, USA, 2002.
10. Koen Buys, Cedric Cagniard, Anatoly Baksheev, Tinne De Laet, Joris De Schutter, and Caroline Pantofaru. An adaptable system for rgb-d based human body detection and pose estimation. *Journal of Visual Communication and Image Representation*, 25(1):39 – 52, 2014. Visual Understanding and Applications with RGB-D Cameras.
11. Enric Cervera, Amine Abou Moughlbay, and Philippe Martinet. Localization and navigation of an assistive humanoid robot in a smart environment. In *IEEE International Workshop on Assistance and Service Robotics in a Human Environment*, 2012.
12. Changhyun Choi and Henrik I. Christensen. RGB-D object tracking: A particle filter approach on GPU. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems, Tokyo, Japan, November 3-7, 2013*, pages 1084–1091, 2013.
13. Kerstin Dautenhahn. Methodology and themes of human-robot interaction: A growing research field. *International Journal of Advanced Robotic Systems*, Vedran Kordic, Aleksandar Lazinica and Munir Merdan (Ed.), 2007.
14. Terrence Fong, Charles E. Thorpe, and Charles Baur. Collaboration, dialogue, human-robot interaction. In *Robotics Research, The Tenth International Symposium, ISRR 2001, Lorne, Victoria, Australia, November 9-12 2001*, pages 255–266, 2001.
15. Michael A. Goodrich and Alan C. Schultz. Human-robot interaction: A survey. *Foundations and Trends in Human-Computer Interaction*, 1(3):203–275, 2007.
16. J. Han, L. Shao, D. Xu, and J. Shotton. Enhanced computer vision with microsoft kinect sensor: A review. *IEEE Trans. Cybernetics*, 43(5), October 2013.
17. Frank Hegel, Claudia Muhl, Britta Wrede, Martina Hielscher-Fastabend, and Gerhard Sagerer. Understanding social robots. In *Second International Conference on Advances in Computer-Human Interaction, ACHI 2009, February 1-7, 2009, Cancun, Mexico*, pages 169–174, 2009.
18. J.G. Hurdus and D.W. Hong. Behavioral programming with hierarchy and parallelism in the darpa urban challenge and robocup. In *IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems, 2008. MFI 2008.*, pages 503–509, Aug 2008.

19. Microsoft Inc. Kinect for windows. <http://www.microsoft.com/en-us/kinectforwindows/>. Accessed: 2014-11-20.
20. Microsoft Inc. Kinect for windows sdk. <http://www.microsoft.com/en-us/download/details.aspx?id=44561>. Accessed: 2014-11-20.
21. Microsoft Inc. Microsoft robotics developer studio 4. <http://msdn.microsoft.com/library/bb483024>. Accessed: 2014-11-28.
22. H. Jaeger and T. Christaller. Dual dynamics: Designing behavior systems for autonomous robots. *Artificial Life and Robotics*, 2(3):108–112, 1998.
23. Hirokazu Kato and Mark Billinghurst. Marker tracking and HMD calibration for a video-based augmented reality conferencing system. In *2nd IEEE and ACM International Workshop on Augmented Reality, IWAR '99, San Francisco, CA, USA, October 20-21, 1999*, pages 85–94, 1999.
24. Daniel Maier, Armin Hornung, and Maren Bennewitz. Real-time navigation in 3d environments based on depth camera data. In *12th IEEE-RAS International Conference on Humanoid Robots (Humanoids 2012), Osaka, Japan, November 29 - Dec. 1, 2012*, pages 692–697, 2012.
25. Thomas B. Moeslund and Erik Granum. A survey of computer vision-based human motion capture. *Computer Vision and Image Understanding*, 81(3):231 – 268, 2001.
26. Thomas B. Moeslund, Adrian Hilton, and Volker Krger. A survey of advances in vision-based human motion capture and analysis. *Computer Vision and Image Understanding*, 104(23):90 – 126, 2006. Special Issue on Modeling People: Vision-based understanding of a persons shape, appearance, movement and behaviour.
27. Ronald Poppe. Vision-based human motion analysis: An overview. *Computer Vision and Image Understanding*, 108(12):4 – 18, 2007. Special Issue on Vision for Human-Computer Interaction.
28. Morgan Quigley, Ken Conley, Brian Gerkey, Josh Faust, Tully Foote, Jeremy Leibs, Rob Wheeler, and Andrew Y Ng. Ros: an open-source robot operating system. *ICRA workshop on open source software*, 3(3.2):5, 2009.
29. Lawrence R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. In Alex Waibel and Kai-Fu Lee, editors, *Readings in Speech Recognition*, pages 267–296. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1990.
30. Radu Bogdan Rusu and Steve Cousins. 3d is here: Point cloud library (PCL). In *IEEE International Conference on Robotics and Automation, ICRA 2011, Shanghai, China, 9-13 May 2011*, 2011.
31. Jamie Shotton, Andrew Fitzgibbon, Mat Cook, Toby Sharp, Mark Finocchio, Richard Moore, Alex Kipman, and Andrew Blake. Real-time human pose recognition in parts from single depth images. In *In CVPR, 2011. 3*, 2011.
32. Jamie Shotton, Ross Girshick, Andrew Fitzgibbon, Toby Sharp, Mat Cook, Mark Finocchio, Richard Moore, Pushmeet Kohli, Antonio Criminisi, Alex Kipman, and Andrew Blake. Efficient human pose estimation from single depth images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(12):2821–2840, 2013.
33. Reid G. Simmons and David Apfelbaum. A task description language for robot control. In *Proceedings 1998 IEEE/RSJ International Conference on Intelligent Robots and Systems. Innovations in Theory, Practice and Applications, October 13-17, 1998, Victoria, BC, Canada*, pages 1931–1937, 1998.
34. Ryohei Ueda. Tracking 3d objects with point cloud library. <http://pointclouds.org/news/2012/01/17/tracking-3d-objects-with-point-cloud-library/>. Accessed: 2015-01-10.

35. Haibin Yan, Jr. Ang, Marcelo H., and Aun Neow Poo. A survey on perception methods for humanrobot interaction in social robots. *International Journal of Social Robotics*, 6(1):85–119, 2014.