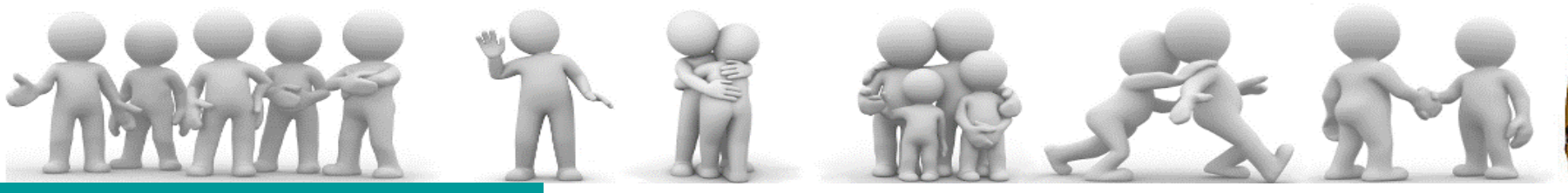
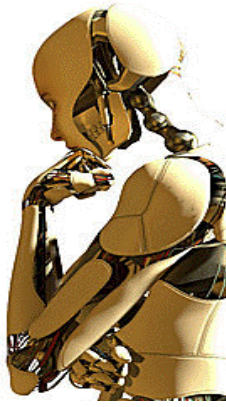
3rd International Workshop on Human Behavior Understanding





7 October 2012 – Algarve in conjunction with IROS'12

Co-Chairs:

Albert Ali Salah
Boğaziçi University, Turkey
salah@boun.edu.tr

Javier Ruiz-del-Solar
Universidad de Chile, Chile
jruizd@ing.uchile.cl

Çetin Meriçli
Carnegie Mellon University, USA
cetin@cmu.edu

Pierre-Yves Oudeyer
INRIA, France
pierre-yves.oudeyer@inria.fr

Program Committee:

Levent Akın, Turkey
Brenna Argall, USA
Kai Arras, Germany
Sven Behnke, Germany
Tony Belpaeme, UK
François Brémond, France
Ginevra Castellano, UK
Kerstin Dautenhahn, UK
Özlem Durmaz İncel, Turkey
Vanessa Evers, The Netherlands
Ian Fasel, USA
Michael Goodrich, USA
Verena Hafner, Germany
Luca Iocchi, Italy
Hatice Köse, Turkey
Ben Kröse, The Netherlands
Manuel Lopes, France
Tekin Meriçli, Turkey
Yukie Nagai, Japan
Mark Neerincx, The Netherlands
Catherine Pelachaud, France
Dennis Reidsma, The Netherlands
Erol Şahin, Turkey
Komei Sugiura, Japan
Leila Takayama, USA
Adriana Tapus, France
Tijn van der Zant, The Netherlands
Manuela Veloso, USA
Rodrigo Verschae, Chile
Juan Wachs, USA
Zeynep Yücel, Japan

Welcome

[8:30-8:45] **Human Behavior Understanding for Robotics**, *Albert Ali Salah, Javier Ruiz-del-Solar, Çetin Meriçli, Pierre-Yves Oudeyer*

Sensing Human Behavior

[8:45- 9:30] **Keynote: Scene Understanding and Assisted Living, *François Brémond***

[9:30- 9:50] **Real-Time Exact Graph Matching with Application in Human Action Recognition,** *Oya Çeliktutan, Christian Wolf, Bülent Sankur and Eric Lombardi*

[9:50-10:10] **An Efficient Approach for Multi-view Human Action Recognition based on Bag-of-Key-Poses,** *Alexandros Andre Chaaraoui, Pau Climent-Pérez and Francisco Flórez-Revuelta*

[10:10-10:30] **Bayesian Fusion of Ceiling Mounted Camera and Laser Range Finder on a Mobile Robot for People Detection and Localization,** *Ninghang Hu, Gwenn Englebienne and Ben J.A. Kröse*

[10:30-11:00] **Coffee break**

Social and Affective Signals

[11:00-11:20] **Using speech data to recognize emotion in human gait,** *Angelica Lim and Hiroshi G. Okuno*

[11:20-11:40] **Gender differences in the Perception of Affective Movements,** *Ali-Akbar Samadani, Rob Gorbet and Dana Kulic*

[11:40-12:00] **Vagueness and dreams. Analysis of body signals in vague dream telling,** *Laura Vincze, Isabella Poggi and Francesca D'Errico*

[12:00-12:20] **Computing and evaluating the Body Laughter Index,** *Maurizio Mancini, Giovanna Varni, Donald Glowinski and Gualtiero Volpe*

[12:20-14:00] **Lunch break**

Human-Robot Interaction

[14:00-14:45] **Keynote: Affordances and Concepts, *Erol Sahin***

[14:45-15:05] **Recognizing the Visual Focus of Attention for Human Robot Interaction,** *Samira Sheikhi and Jean-Marc Odobez*

[15:05-15:25] **Contextual analysis of human non-verbal guide behaviors to inform the development of FROG, the Fun Robotic Outdoor Guide,** *Daphne E. Karreman, Elisabeth M.A.G. van Dijk, and Vanessa Evers*

[15:25-15:45] **Between Initial Expectations and Acquaintance: Interacting with a Developing Robot,** *Kerstin Fischer and Joe Saunders*

[15:45-16:15] **Coffee break**

Imitation and Learning from Demonstration

[16:15-17:00] **Keynote: Robots and the Human, *Oussama Khatib***

[17:00-17:20] **Learning the combinatorial structure of demonstrated behaviors with inverse feedback control,** *Olivier Mangin and Pierre-Yves Oudeyer*

[17:20-17:40] **Internal Simulations for Behaviour Selection and Recognition,** *Guido Schillaci, Bruno Lara and Verena Hafner*

[17:40-18:00] **Automatic Imitation Assessment in Interaction,** *Stéphane Michelet, Koby Karp, Emilie Delaherche, Catherine Achard and Mohamed Chetouani*

http://www.cmpe.boun.edu.tr/hbu/2012/

Preface

Domains where human behavior understanding is a crucial need (e.g., robotics, human-computer interaction, affective computing, and social signal processing) rely on advanced pattern recognition techniques to automatically interpret complex behavioral patterns generated when humans interact with machines or with others. This is a challenging problem where many issues are still open, including the joint modeling of behavioral cues taking place at different time scales, the inherent uncertainty of machine detectable evidences of human behavior, the mutual influence of people involved in interactions, the presence of long term dependencies in observations extracted from human behavior, and the important role of dynamics in human behavior understanding. Implementing these methods on robotic platforms introduces further constraints on processing resources, tracking over time, model building, and generalization.

The Third Workshop on Human Behavior Understanding (HBU), organized as a satellite to IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'2012), gathers researchers dealing with the problem of modeling human behavior under its multiple facets (expression of emotions, display of relational attitudes, performance of individual or joint actions, imitation, etc.), with particular attention to implications in robotics, including additional resource and robustness constraints of robotic platforms, social aspects of human-robot interaction, and developmental approaches to robotics.

The Workshop features three invited talks by François Brémond (INRIA, France), Erol Şahin (METU, Turkey), and Oussama Khatib (Stanford University, USA). We received 31 submissions in total, and each paper was peer-reviewed by at least two members of the technical program committee.

The proceedings volume, published by Springer Verlag in the LNCS series as LNCS 7559, contains the papers presented at the workshop and a summarizing paper. In this short version, we include the abstracts of all the papers, as well as abstracts of the keynotes, and speaker bios. We thank Springer Verlag for giving us permission to publish these abstracts here.

We would like to take the opportunity to thank our program committee members and reviewers for their rigorous feedback, our authors and our keynote speakers for their contributions. We thank the PAL project of INRIA, BAP 6531 project of Boğaziçi University, and the EUCogIII network for their financial support.

Albert Ali Salah
Javier Ruiz-del-Solar
Çetin Meriçli
Pierre-Yves Oudeyer

Organization

Conference Co-chairs

Albert Ali Salah	Boğaziçi University, Turkey
Javier Ruiz-del-Solar	Universidad de Chile, Chile
Çetin Meriçli	Carnegie Mellon University, USA
Pierre-Yves Oudeyer	INRIA, France

Technical Program Committee

Levent Akın	Boğaziçi University, Turkey
Brenna Argall	Northwestern University, USA
Kai Arras	Albert-Ludwigs-Universität Freiburg, Germany
Sven Behnke	University Bonn, Germany
Tony Belpaeme	University of Plymouth, UK
François Brémond	INRIA, France
Ginevra Castellano	University Birmingham, UK
Kerstin Dautenhahn	University Hertfordshire, UK
Özlem Durmaz İncel	Boğaziçi University, Turkey
Vanessa Evers	University of Twente, The Netherlands
Ian Fasel	University of Arizona, USA
Michael Goodrich	Brigham Young University, USA
Verena Hafner	Humboldt-University Berlin, Germany
Luca Iocchi	Roma University, Italy
Hatice Köse	Istanbul Technical University, Turkey
Ben Kröse	Univ. of Amsterdam & Hogeschool van Amsterdam, The Netherlands
Manuel Lopes	INRIA, France
Tekin Meriçli	Boğaziçi University, Turkey
Yukie Nagai	Osaka University, Japan
Mark Neerincx	Delft Univ. of Technology, The Netherlands
Catherine Pelachaud	TELECOM-ParisTech, France
Dennis Reidsma	University of Twente, The Netherlands
Erol Şahin	Middle East Technical University, Turkey
Komei Sugiura	NICT, Japan
Leila Takayama	Willow Garage, USA
Adriana Tapus	ENSTA-ParisTech, France
Tijn van der Zant	University of Gröningen, The Netherlands
Manuela Veloso	Carnegie Mellon University, USA
Rodrigo Verschae	AMTC-Chile, Chile
Juan Wachs	Purdue University, USA
Zeynep Yücel	ATR, Japan

Additional Reviewers

Thomas Cederborg
Barış Evrim Demiröz
Jonathan Grizou
Olivier Mangin
Sao Mai Nguyen
Ronan Sicre

Table of Contents

Imitation and Learning from Demonstration

Human Behavior Understanding for Robotics

Albert Ali Salah[1], Javier Ruiz-del-Solar[2], Çetin Meriçli[3], and Pierre-Yves Oudeyer[4]

[1] Boğaziçi University, Department of Computer Engineering, Istanbul, Turkey
salah@boun.edu.tr
[2] Universidad de Chile, Department of Electric Engineering Av. Tupper 2007, Santiago, Chile
jruizd@ing.uchile.cl
[3] Carnegie Mellon University, Computer Science Department Pittsburgh, PA 15213, USA
cetin@cmu.edu
[4] FLOWERS research team, INRIA Bordeaux Sud-Ouest 33405 Talence, Cedex, France
pierre-yves.oudeyer@inria.fr

Abstract. Human behavior is complex, but structured along individual and social lines. Robotic systems interacting with people in uncontrolled environments need capabilities to correctly interpret, predict and respond to human behaviors. This paper discusses the scientific, technological and application challenges that arise from the mutual interaction of robotics and computational human behavior understanding. We supply a short survey of the area to provide a contextual framework and describe the most recent research in this area.

Invited talk: Scene Understanding and Assisted Living

François Brémond

INRIA, 06902 Sophia Antipolis, France
francois.bremond@inria.fr

Scene understanding is the process, often real time, of perceiving, analyzing and elaborating an interpretation of a 3D dynamic scene observed through a network of sensors (e.g. video cameras). This process consists mainly in matching signal information coming from sensors observing the scene with models which humans are using to understand the scene. Based on that, scene understanding is both adding and extracting semantic from the sensor data characterizing a scene. This scene can contain a number of physical objects of various types (e.g. people, vehicle) interacting with each others or with their environment (e.g. equipment) more or less structured. The scene can last few instants (e.g. the fall of a person) or few months (e.g. the depression of a person), can be limited to a laboratory slide observed through a microscope or go beyond the size of a city. Sensors include usually cameras (e.g. omni-directional, infrared), but also may include microphones and other sensors (e.g. optical cells, contact sensors, physiological sensors, radars, smoke detectors). Scene understanding is influenced by cognitive vision and it requires at least the melding of three areas: computer vision, cognition and software engineering. Scene understanding can achieve five levels of generic computer vision functionality of detection, localization, tracking, recognition and understanding. But scene understanding systems go beyond the detection of visual features such as corners, edges and moving regions to extract information related to the physical world which is meaningful for human operators. Its requirement is also to achieve more robust, resilient, adaptable computer vision functionalities by endowing them with a cognitive faculty: the ability to learn, adapt, weigh alternative solutions, and develop new strategies for analysis and interpretation. In this work, we discuss how scene understanding can be applied to Home Care Monitoring.

François Brémond is leading the STARS team at INRIA Sophia Antipolis. He designs and develops generic systems for dynamic scene interpretation that detect and track mobile objects, which can be either humans or vehicles, and recognize their behaviours. He is particularly interested in filling the gap between sensor information (pixel level) and recognized activities (semantic level). In 1997 he obtained his PhD degree at INRIA in video understanding and worked as a post doctorate at USC on the interpretation of videos taken from UAV (Unmanned Airborne Vehicle). He participated in many European and industrial research projects in activity monitoring. François Brémond is author of more than 100 scientific papers published in international journals or conferences in video understanding. In 2005 he co-founded Keeneo, a company in intelligent video surveillance.

Real-Time Exact Graph Matching with Application in Human Action Recognition

Oya Çeliktutan[1,2], Christian Wolf[2], Bülent Sankur[1], and Eric Lombardi[2]

[1]Electrical and Electronics Engineering,
Boğaziçi University, Istanbul, Turkey
[2]Université de Lyon, CNRS,
INSA-Lyon, LIRIS, UMR CNRS 5205, F-69621, France

Abstract. Graph matching is one of the principal methods to formulate the correspondence between two set of points in computer vision and pattern recognition. However, most formulations are based on the minimization of a difficult energy function which is known to be NP-hard. Traditional methods solve the minimization problem approximately. In this paper, we show that an efficient solution can be obtained by exactly solving an approximated problem instead of approximately solving the original problem. We derive an exact minimization algorithm and successfully apply it to action recognition in videos. In this context, we take advantage of special properties of the time domain, in particular causality and the linear order of time, and propose a novel spatio-temporal graphical structure.

Keywords: Spatio-temporal graph, Hyper-graph matching, Action recognition

An Efficient Approach for Multi-view Human Action Recognition Based on Bag-of-Key-Poses

Alexandros Andre Chaaraoui[1], Pau Climent-Pérez[1], and Francisco Flórez-Revuelta[2]

[1] Department of Computing Technology, University of Alicante, Alicante, Spain
{alexandros,pcliment}@dtic.ua.es
[2] Faculty of Science, Engineering and Computing, Kingston University, Kingston upon Thames, United Kingdom
F.Florez@kingston.ac.uk

Abstract. This paper presents a novel multi-view human action recognition approach based on a bag-of-key-poses. In the case of multi-view scenarios, it is especially difficult to perform accurate action recognition that still runs at an admissible recognition speed. The presented method aims to fill this gap by combining a silhouette-based pose representation with a simple, yet effective multi-view learning approach based on *Model Fusion*. Action classification is performed through efficient sequence matching and by the comparison of successive key poses which are evaluated on both feature similarity and match relevance. Experimentation on the MuHAVi dataset shows that the method outperforms currently available recognition rates and is exceptionally robust to actor-variance. Temporal evaluation confirms the method's suitability for real-time recognition.

Keywords: human action recognition, multi-view action recognition, key pose, bag-of-key-poses, MuHAVi dataset

Bayesian Fusion of Ceiling Mounted Camera and Laser Range Finder on a Mobile Robot for People Detection and Localization

Ninghang Hu[1], Gwenn Englebienne[1], Ben Kröse[1,2]

[1] University of Amsterdam
[2] Amsterdam University of Applied Science

Abstract. Robust people detection and localization is a prerequisite for many applications where service robots interact with humans. Future robots will not be stand-alone any more but will operate in smart environments that are equipped with sensor systems for context awareness and activity recognition. This paper describes a probabilistic framework for the fusion of data from a laser range finder on a mobile robot and an overhead camera fixed in a domestic environment. The contribution of the framework is that it enables seamless integration with other sensors. For tracking multiple people it is possible to use a probabilistic particle filter tracker. We show that the fusion improves the results of the individual subsystems.

Using speech data to recognize emotion in human gait

Angelica Lim and Hiroshi G. Okuno

Graduate School of Informatics, Kyoto University
Sakyo-ku, Kyoto, Japan 606-8501
{angelica, okuno}@kuis.kyoto-u.ac.jp

Abstract. Robots that can recognize emotions can improve humans' mental health by providing empathy and social communication. Emotion recognition by robots is challenging because unlike in human-computer environments, facial information is not always available. Instead, our method proposes using speech and gait analysis to recognize human emotion. Previous research suggests that the dynamics of emotional human speech also underlie emotional gait (walking). We investigate the possibility of combining these two modalities via perceptually common parameters: Speed, Intensity, irRegularity, and Extent (SIRE). We map low-level features to this 4D cross-modal emotion space and train a Gaussian Mixture Model using independent samples from both voice and gait. Our results show that a single, modality-mixed trained model can perform emotion recognition for both modalities. Most interestingly, recognition of emotion in gait using a model trained uniquely on speech data gives comparable results to a model trained on gait data alone, providing evidence for a common underlying model for emotion across modalities.

Keywords: robot emotions, emotional gait, emotional voice, affect recognition

Gender differences in the perception of affective movements

Ali-Akbar Samadani, Rob Gorbet, and Dana Kulić

University of Waterloo, Electrical and Computer Engineering Department
Waterloo, Ontario N2L 3G1, Canada

Abstract. Identifying human capabilities in perceiving affective expressions is essential for developing interactive machines that can engage with their human users. In order to ensure that the behaviour of the interactive machine is perceived as intended, any gender–specific differences in the perception of affective expressions are an important design consideration. This paper presents a preliminary study investigating the role of gender in the perception of affective hand movements displayed on both anthropomorphic and non-anthropomorphic structures. The results show that gender significantly influences the participants' perception and that the impact of the display structure and intended-emotion on the perception of the affective movements differs between male and female observers.

Keywords: Affective movements, Gender differences, Display structure, Perception, User study.

Vagueness and dreams. Analysis of body signals in vague dream telling

Laura Vincze, Isabella Poggi and Francesca D'Errico

Educational Department, Roma Tre University
Rome, Italy
laura.vincze@gmail.com, poggi@uniroma3.it, fderrico@uniroma3.it

Abstract. The paper provides a conceptual definition of the notions of vagueness, approximation, hesitation and hastiness and an analysis of multimodal cases of vague, approximate, hesitant and hasty communication during dream-telling.

Keywords: Vagueness, approximation, hesitation, hastiness, social signals, multimodal communication

Computing and evaluating the Body Laughter Index

Maurizio Mancini, Giovanna Varni, Donald Glowinski, and Gualtiero Volpe

InfoMus - University of Genoa
maurizio.mancini@dist.unige.it
giovanna@infomus.org
donald.glowinski@dist.unige.it
gualtiero@infomus.org

Abstract. The EU-ICT FET Project ILHAIRE is aimed at endowing machines with automated detection, analysis, and synthesis of laughter. This paper describes the Body Laughter Index (BLI) for automated detection of laughter starting from the analysis of body movement captured by a video source. The BLI algorithm is described, and the index is computed on a corpus of videos. The assessment of the algorithm by means of subject's rating is also presented. Results show that BLI can successfully distinguish between different videos of laughter, even if improvements are needed with respect to perception of subjects, multimodal fusion, cultural aspects, and generalization to a broad range of social contexts.

Invited talk: Affordances and Concepts

Erol Şahin

KOVAN Research Lab, Middle East Technical University
Ankara, Turkey
erol@metu.edu.tr

Do we perceive all the qualities of the environment to accomplish a simple task like grasping? Do we perceive all properties of an object, and only then infer whether it is graspable or not? Do we think "this green spherical object towards my right is a tennis ball, and I know that balls that are smaller than my hand can be grasped, therefore I can successfully grasp it"? J.J. Gibson, one of the most influential figures in the field of psychology, objected such a view. Instead, he conceived the notion of affordances to explain how inherent "values" and "meanings" of things in the environment can be directly perceived and how this information can be linked to the action possibilities offered to the organism by the environment as: "The affordances of the environment are what it offers the animal, what it provides or furnishes, either for good or ill. The verb to afford is found in the dictionary, but the noun affordance is not. I have made it up. I mean by it something that refers to both the environment and the animal in a way that no existing term does. It implies the complementarity of the animal and the environment." Although introduced in psychology, the notion shifted the focus from the agent or the environment alone, to their interaction. In this paper, I first review this popular notion through its different, sometimes contradictory, interpretations in other fields ranging from human-computer interaction to autonomous robotics to develop a formalization of affordances for its use at different levels of autonomous robot control. Using this formalization as a framework, I show methods how robots can automatically learn to perceive affordances in their environments, use learned affordance relations to ground symbolic planning mechanisms in the continuous sensory-motor experiences of the robot and link them with concepts represented by verbs and nouns in language on robots to communicate with humans. I conclude by pointing out the future directions on this line of research, briefly discussing social affordances observed in human-robot interactions.

Erol Şahin is an Assist. Prof. at the Computer Engineering Department of METU, heading the KOVAN Research Laboratory. He has a Ph.D. in Cognitive & Neural Systems from Boston University, USA, where he worked with Prof. Paolo Gaudiano at the now-defunct Neurobotics Lab. His research on robotics is being funded by the ROSSI project (FP7-ICT-21625). Recently the group was awarded an iCub humanoid robot platform, being developed by the RobotCub project. The group is developing a neurocomputational model based on the notion of affordances in psychology and mirror and canonical neuron studies in neurophysiology. Dr. Şahin has edited a number of books and special issues on Swarm Robotics and is an associate editor of the Adaptive Behavior journal.

Recognizing the Visual Focus of Attention for Human Robot Interaction

Samira Sheikhi[1,2] and Jean-Marc Odobez[1,2]

[1] Idiap Research Institute, Switzerland
[2] École Polytechnique Fédérale de Lausanne, Switzerland

Abstract. We address the recognition of people's visual focus of attention (VFOA), the discrete version of gaze that indicates who is looking at whom or what. As a good indicator of addressee-hood (who speaks to whom, and in particular is a person speaking to the robot) and of people's interest, VFOA is an important cue for supporting dialog modelling in Human-Robot interactions involving multiple persons. In absence of high definition images, we rely on people's head pose to recognize the VFOA. Rather than assuming a fixed mapping between head pose directions and gaze target directions, we investigate models that perform a dynamic (temporal) mapping implicitly accounting for varying body/shoulder orientations of a person over time, as well as unsupervised adaptation. Evaluated on a public dataset and on data recorded with the humanoid robot Nao, the method exhibit better adaptivity and versatility producing equal or better performance than a state-of-the-art approach, while the proposed unsupervised adaptation does not improve results.

Keywords: Human robot interaction, visual focus of attention, gaze, head pose.

Contextual analysis of human non-verbal guide behaviors to inform the development of FROG, the Fun Robotic Outdoor Guide

Daphne E. Karreman, Elisabeth M.A.G. van Dijk and Vanessa Evers

Human Media Interaction, University of Twente
Enschede, the Netherlands
{d.e.karreman, e.m.a.g.vandijk, v.evers} @utwente.nl

Abstract. This paper reports the first step in a series of studies to design the interaction behaviors of an outdoor robotic guide. We describe and report the use case development carried out to identify effective human tour guide behaviors. In this paper we focus on non-verbal communication cues in gaze, gestures and movements. The work reported involves the observation of human tour guide behaviors and visitor responses as well as interviews with guides. An affinity diagram is used to identify effective communication cues of human guides and the relations between them. The opportunities for a robotic guide are discussed. We argue that human guide behaviors and strategies cannot be one-on-one ap-plied to robot tour guides. Instead, we aim to develop abstractions of the human behaviors, appropriate for robot tour guides and effective in realizing visitor engagement. The results of this study will be used to create a first Fun Robotic Outdoor Guide prototype with the abstracted interactive robot guide behaviors implemented to assess the effects on visitor experience in 'the wild.'

Keywords: Human Tour Guide Behavior, Non-Verbal Robot Behavior, Contextual Analysis

Between Initial Expectations and Acquaintance: Interacting with a Developing Robot

Kerstin Fischer[1] and Joe Saunders[2]

[1] IFKI, University of South Denmark,
Slagelse, Denmark
kerstin@sitkom.sdu.dk
[2] University of Hertfordshire,
Hertfordshire, UK
j.1.saunders@herts.ac.uk

Abstract. Two factors that been suggested to influence the ways in which people interact with robots, namely users' initial expectations on the one hand and their increasing acquaintance with their robotic partner due to repeated interaction over time on the other. In the current study, eight participants interacted with a humanoid robot in five different sessions. Between the sessions, the robot was trained on the linguistic material presented to it by its human tutor in the preceding session, and thus the robot exhibits increasingly more knowledge of the domain. The results uncover the interaction between users' preconceptions and feedback-driven interactional effects that shape human-robot interactions. While considerable differences between users can be observed, all users respond to the robot's feedback and increasing linguistic capabilities in comparable ways.

Invited talk: Robots and the Human

Oussama Khatib

Artificial Intelligence Laboratory, Stanford University, CA, USA
khatib@robotics.stanford.edu

In the field of robotics, the motivation to emulate human movement has been driven by the desired to endow robots, humanoids in particular, with human-like movement properties. Understanding the fundamental characteristics of human motion is a challenging multifaceted problem that requires, in particular, the development of accurate models of the kinematics, dynamics, and actuation of human musculoskeletal systems. These models are essential for building full human motion simulation and for performing motion reconstruction from captured data, as well as for the analysis and characterization of human movement. Another major element in the synthesis of human motion is the control architecture needed for dealing with the highly redundant and tightly constrained nature of musculoskeletal systems. These issues have much in common with the problems encountered in redundant articulated body systems profoundly studied in robotics. Given these shared underlying problems and given the progress and advances made in computational robotics, which has been strongly motivated by real-time needs, algorithms, methodologies, and techniques developed in robotics are being increasingly used in studies of human motion. Analytical models originated in robotics are providing much needed tools for human motion synthesis. The discussion focuses on our ongoing studies into the connection between humans and robots and on the new insights and results this exploration has produced. These developments, which are proving extremely valuable in human biomechanics, are providing new avenues for exploring human motion – with exciting prospects for novel clinical therapies, athletic training, character animation, and human performance improvement.

Oussama Khatib received his PhD in Electrical Engineering from SupAero, Toulouse, in 1980. He is Professor of Computer Science at Stanford University. His work on advanced robotics focuses on methodologies and technologies on human-centered robotics including humanoid control architectures, human motion synthesis, interactive dynamic simulation, haptics, and human-friendly robot design. He is Co-Editor of the Springer Tracts in Advanced Robotics series, and has served on the Editorial Boards of several journals as well as the Chair of numerous conferences. He coedited the Springer Handbook of Robotics, which received the PROSE Award for Excellence in Physical Sciences & Mathematics and was also the winner in the category Engineering & Technology. He is a Fellow of IEEE and has served RAS as a Distinguished Lecturer, as a member of the Administrative Committee, and as the Program Chair of ICRA 2000. He is the President of the International Foundation of Robotics Research (IFRR) and a recipient of the Japan Robot Association (JARA) Award in Research and Development. Professor Khatib received the 2010 IEEE RAS Pioneer Award in Robotics and Automation.

Learning the combinatorial structure of demonstrated behaviors with inverse feedback control

Olivier Mangin[1,2] and Pierre-Yves Oudeyer[1]

[1] Flowers Team, INRIA, France
[2] Université Bordeaux 1, France

Abstract. In many applications, such as virtual agents or humanoid robots, it is difficult to represent complex human behaviors and the full range of skills necessary to achieve them. Real life human behaviors are often the combination of several parts and never reproduced in the exact same way. In this work we introduce a new algorithm that is able to learn behaviors by assuming that the observed complex motions can be represented in a smaller dictionary of concurrent tasks. We present an optimization formalism and show how we can learn simultaneously the dictionary and the mixture coefficients that represent each demonstration. We present results on a idealized model where a set of potential functions represents human objectives or preferences for achieving a task.

Internal Simulations for Behaviour Selection and Recognition

Guido Schillaci[1], Bruno Lara[2] and Verena V. Hafner[1]

[1] Cognitive Robotics Group,
Department of Computer Science,
Humboldt-Universität zu Berlin.
Germany.
{guido.schillaci, hafner}@informatik.hu-berlin.de,
[2] Cognitive Robotics Group,
Faculty of Science,
Universidad Autonoma del Estado de Morelos.
Cuernavaca, Mexico.
bruno.lara@uaem.mx

Abstract. In this paper, we present internal simulations as a methodology for human behaviour recognition and understanding. The internal simulations consist of pairs of inverse forward models representing sensorimotor actions. The main advantage of this method is that it both serves for action selection and prediction as well as recognition. We present several human-robot interaction experiments where the robot can recognize the behaviour of the human reaching for objects.

Keywords: behaviour recognition, internal simulation, human-robot interaction, internal models

Automatic Imitation Assessment in Interaction

Stéphane Michelet, Koby Karp, Emilie Delaherche, Catherine Achard and Mohamed Chetouani

Institute of Intelligent Systems and Robotics,
University Pierre and Marie Curie, 75005 Paris, France.
{michelet, karp, delaherche}@isir.upmc.fr,
{catherine.achard, mohamed.chetouani}@upmc.fr

Abstract. Detecting social events such as imitation is identified as key step for the development of socially aware robots. In this paper, we present an unsupervised approach to measure immediate synchronous and asynchronous imitations between two partners. The proposed model is based on two steps: detection of interest points in images and evaluation of similarity between actions. Firstly, spatio-temporal points are detected for an accurate selection of the important information contained in videos. Then bag-of-words models are constructed, describing the visual content of videos. Finally similarity between bag-of-words models is measured with dynamic-time-warping, giving an accurate measure of imitation between partners. Experimental results obtained show that the model is able to discriminate between imitation and non-imitation phases of interactions.

Keywords: Imitation, DTW, unsupervised learning

Author Index