



A methodological approach relating the classification of gesture to identification of human intent in the context of human-robot interaction

Chrystopher L Nehaniv, Kerstin Dautenhahn, Jens Kubacki, Martin Haegele, Christopher Parlitz, Rachid Alami

► To cite this version:

Chrystopher L Nehaniv, Kerstin Dautenhahn, Jens Kubacki, Martin Haegele, Christopher Parlitz, et al.. A methodological approach relating the classification of gesture to identification of human intent in the context of human-robot interaction. ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, Aug 2005, Nashville, United States. <hal-01976823>

HAL Id: hal-01976823

<https://laas.hal.science/hal-01976823v1>

Submitted on 10 Jan 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

A Methodological Approach relating the Classification of Gesture to Identification of Human Intent in the Context of Human-Robot Interaction*

Chrystopher L. Nehaniv, Kerstin Dautenhahn
Adaptive Systems Research Group
School of Computer Science, University of Hertfordshire
College Lane, Hatfield Herts AL10 9AB, United Kingdom
C.L.Nehaniv@herts.ac.uk

Jens Kubacki, Martin Haegele, Christopher Parlitz
Fraunhofer Institute for Manufacturing
Engineering & Automation (IPA)
Nobelstraße 12
70569 Stuttgart, Germany

Rachid Alami
LAAS/CNRS, Robotics & Artificial Intelligence Group
7, Avenue du Colonel Roche
31077 Toulouse Cedex 4, France

Abstract—In order to infer intent from gesture, a broad classification of types of gestures into five main classes is introduced. The classification is intended as a generally applicable basis for incorporating the understanding of gesture into human-robot interaction (HRI). Examples from human-robot interaction show the need to take into account not only the kinematics of gesture, but also the interactional context. Requirements for the operational classification of gesture by a robot interacting with humans are suggested and initial steps in its deployment are discussed.

Index Terms—interaction context, classification of gestures, human-robot activity and interaction.

I. INTRODUCTION: THE NEED FOR CLASSIFYING GESTURE

The word *gesture* is used for many different phenomena involving human movement, especially of the hands and arms. Only some of these are interactive or communicative. The pragmatics of gesture and meaningful interaction are quite complex (cf. [9], [11], [12]), and an international journal [6] now exists entirely devoted to the study of gesture. Applications of service or ‘companion’ robots that interact with humans, including naive ones, will increasingly require human-robot interaction (HRI) in which the robot can recognize *what* humans are doing and to a limited extent *why* they are doing it, so that the robot may act appropriately, e.g. either by assisting, or staying out of the way. Due to the situated embodied nature of such interactions and the non-human nature of robots, it is not possible to directly carry over methods from human-computer interaction (HCI) or rely entirely on insights from the psychology of human-human interaction. Insights from proxemics and kinesics, which study spatial and

temporal aspects of human-human interaction [7], [4], [9] and some insights of HCI, e.g. recognizing the diversity of users and providing feedback acknowledgment with suitable response timing (e.g. [16]), may also prove to be extremely valuable to HRI. Notwithstanding, the nascent field of HRI must develop its own methods particular to the challenges of embodied interaction between humans and robots. New design, validation, evaluation methods and principles particular to HRI must be developed to meet new challenges such as *legibility*, making the robot’s actions and behaviour understandable and predictable to a human, and ‘*robotiquette*’, respecting human activities and situations (e.g. not interrupting a conversation between humans or disturbing a human who is concentrating or working intensely — without sufficient cause), as well as respecting social spaces, and maintaining appropriate proximity and levels of attention in interaction. Part of meeting these challenges necessarily involves some understanding of human activity at an appropriate level. This requires the capabilities of recognizing human gesture and movement, and inferring intent. The term “*intent*” is used in this paper in a *limited* way that refers to *particular motivation(s) of a human being that result in a gestural motion directly or indirectly relevant for human-robot interaction*.

In inferring the intent from a human’s gesture it is helpful to have a classification of which type of gesture is being observed. Without a sufficiently broad classification, understanding of gesture will be too narrow to characterize what is happening and appropriate responses will not be possible in many cases.

Knowing how to recognize and classify gesture may also serve to inform the design of robot behaviour, including gestures made by the robot to achieve legibility and convey aspects of the robot’s state and plans to humans. This in turn will contribute to robot interaction with humans that is legible, natural, safe, and comfortable for the humans interacting with the robot. To begin to approach the complexity of gesture in the context of situated human-

*The work described in this paper was conducted within the EU Integrated Project COGNIRON (“The Cognitive Robot Companion”) and was funded by the European Commission Division FP6-IST Future and Emerging Technologies under Contract FP6-002020. This paper extends and supercedes C. L. Nehaniv, “Classifying Gesture and Inferring Intent” *Proc. AISB’05 Symposium on Robot Companions: Hard Problems and Open Challenges in Robot-Human Interaction*, The Society for the Study of Artificial Intelligence and Simulation of Behaviour, 2005.

robot interaction, the rough classes of gesture described below are developed in order to provide a broad level of description and the first steps toward a pragmatic, operational definition that could be used by an autonomous system such as a robot to help it (1) to infer the intent of human interaction partners, and (2), as an eventual goal, to help the robot use gestures itself (if possible) to increase the legibility of its behaviour.

II. SOME RELATED WORK ON RECOGNIZING GESTURE AND INTENT

The questions of how gestures are acquired and come to be recognized as meaningful by particular individuals in the course of their development (ontogeny of gesture and its recognition), and conventionalized, elaborated, or lost within particular cultures (evolution of gesture) are large and deep issues, but will not be addressed within the scope in this paper. Psychological/linguistic studies of human gesture use and understanding, related classifications relevant for interaction, language evolution, and language acquisition, e.g. by hearing or deaf children, have all been undertaken (cf. [17]). Such understanding of the development of gesture and its functions may help shed light on gesture in human-robot interaction.

While this paper does not attempt a comprehensive survey of the role and recognition of gesture in human-robot interaction, it does suggest inherent limitations of approaches working with a too narrow notion of gesture, excluding entire classes of human gesture that should eventually be accessible to interactive robots able to function well in a human social environment. Much work with data gloves, typically at a low level for hand gesture recognition for virtual reality or of manipulative grasping has been carried out since the 1990's (e.g. [5], [1]). The important role of gesture for intent communication in human-robot interaction is increasingly being acknowledged, although some approaches still focus only on static hand poses rather than dynamic use of more general types of gesture in context; a survey of hand gesture understanding in robotics appears in [13].

Multimodal and voice analysis can also help to infer intent via prosodic patterns, even when ignoring the content of speech. Robotic recognition of a small number of distinct prosodic patterns used by adults that communicate praise, prohibition, attention, and comfort to preverbal infants has been employed as feedback to the robot's 'affective' state and behavioural expression, allowing for the emergence of interesting social interaction with humans [3]. Hidden Markov Models (HMMs) have been used to classifying limited numbers of gestural patterns (such as grasps or letter shapes) and also to generate trajectories by a humanoid robot matching those demonstrated by a human [2]. Multimodal speech and gesture recognition using HMMs has been implemented for giving commands via pointing, one-, and two-handed gestural commands together with voice for intention extraction into a structured symbolic data stream for use in controlling and programming a vacuuming cleaning robot [8]. Many more examples in

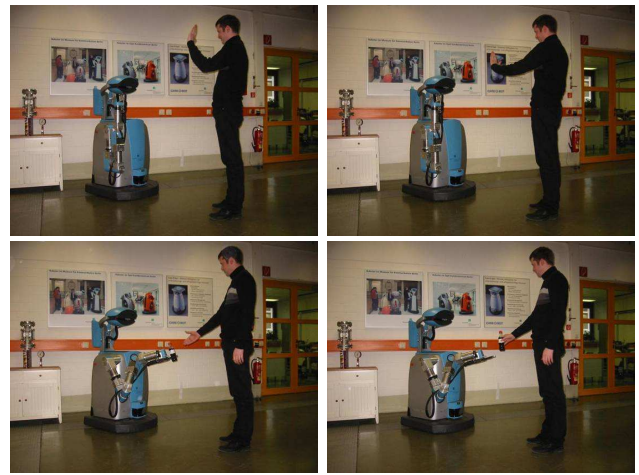


Fig. 1. **Gestures with similar kinematics but different functions.** Top row: HELLO (left) an interactional gesture (class 4) is similar to STOP (right) a conventional symbol (class 3). Bottom row: PASS OBJECT (left) is similar and TAKE OBJECT are both multifunctional interactional (class 4) and manipulative (class 1) gestures. Activity and situational context – e.g. stage of interaction and current activity (top row), or location of manipulandum, here a bottle (bottom row) – are used to disambiguate between such kinematically similar gestures.

robotics exist. Nevertheless, approaches pursued so far in robotics thus tend to use very limited, constrained, and specific task-related gestural repertoires of primitives, and do not attempt to identify general gestural classes. They have tended to focus on a fixed symbolic set of gestures (possibly an extensible one, in which new gestures can be learned), or focus on only a few representatives from one or two of the gestural classes identified here (e.g. grasps (a subclass of manipulative gestures), or on symbolic and pointing gestures).

III. INSUFFICIENCY OF BODY MODEL FITTING ANALYSES: RELATING CONTEXT TO KINEMATICS.

It should be stressed that a single specific instance of a particular the kind of physical gestural motion could, depending on context and interaction history, reflect very different kinds of human intent. It will not always be possible to infer intent based solely on based the mechanical aspects of human movements (such as changes in joint angles) without taking context into account.

Gestures with identical or near identical human kinematics can be different classes. In general, the kinematic picture alone is not enough to determine the class of a gesture or the human's intent. Examples in shown in figure 1 require contextual information in order to be disambiguated. Relating the context and history of interaction to the kinematics is a key point for recognizing human gestures in HRI. Figure 1 using our classification (see below) illustrates this ambiguity.

IV. CLASSIFICATION OF GESTURES

To approach this problem, a classification of gesture for inferring intent and assisting in the understanding of human activity should closely relate gesture with limited categories

of intent in situated human activity. The categories of the broad classification presented here thus correspond to and allow the attribution of limited kinds of intent to humans. This classification is developed as an aid for helping robots to achieve limited recognition of situated human gestural motion, so as to be able to respond appropriately if required, while these robots are working in an environment of ambient human activity (such as a home or office), in which, at times, the robots are also assisting or cooperating with the humans. *Applications of this classification will require the mapping of physical aspects of gestural motion in interactional contexts to the five gestural classes (and their subtypes) suggested here.*

The following is a rough, tentative classification. Gestures are classed into five major types with some subtypes.

A. Five Classes (with Subtypes)

- 1) **'Irrelevant'/Manipulative Gestures.** These include *irrelevant gestures, body / manipulator motion, side-effects of motor behaviour, and actions on objects.* Broadly characterized, manipulation by a human is here understood as *doing something to influence the non-animate environment or the human's relationship to it (such as position).* Gestural motions in this class are manipulative actions (in this sense) and their side effects on body movement. These 'gestures' are neither communicative nor socially interactive, but instances and effects of human motion. They may be salient, but are not movements that are primarily employed to communicate or engage a partner in interaction. Cases include, e.g. motion of the arms and hands when walking; tapping of the fingers; playing with a paper clip; brushing hair away from the face with the hand; scratching; grasping a cup in order to drink its contents. (Note: it may be very important to distinguish among the subtypes listed above for robot understanding of human behaviour.)
- 2) **Side Effect of Expressive Behaviour.** In communicating with others, motion of hands, arms and face (changes in their states) occur as part of the overall communicative behaviour, but without any specific interactive, communicative, symbolic, or referential roles (cf. classes 3-5)
Example: persons talk excitedly raising and moving their hands in correlation with changes in voice prosody, rhythm, or emphasis of speech.
- 3) **Symbolic Gestures.** Gestural motion in symbol gesture is a *conventionalized signal in a communicative interaction.* It is generally a member of a limited, circumscribed set of gestural motions that have specific, prescribed interpretations. A symbolic gesture is used to trigger certain actions by a targeted perceiver, or to refer to something or substitute as for another signal according to a code or convention. Single symbolic gestures are analogous to discrete actions on an interface, such as clicking a button.
Examples: waving down a taxi for it to stop; use of a conventional hand signals (a command to halt

indicated open flat hand; a military salute); nodding 'yes'; waving a greeting 'hello' or 'goodbye'.

Note that the *degree of arbitrariness* in such gestures may vary: The form of the gesture may be an arbitrary conventional sign (such as a holding up two fingers with palm forwards to mean 'peace', or the use of semaphores for alphabetic letters). On the other hand, a symbolic gesture may resemble to a lesser or greater extent iconically or, in ritualized form, a referent or activity.

Further examples: holding up two fingers to indicate 'two'; opening both (empty) hands by turning palms down to indicate a lack of something. Nearly all symbolic gestures are used to convey *content* in communicative interactions.

- 4) **Interactional Gestures.** These are gesture used to *regulate interaction with a partner*, i.e. used to initiate, maintain, invite, synchronize, organize or terminate a particular interactive, cooperative behaviour: raising a empty hand toward the partner to invite the partner to give an object; raising the hand containing an object toward the partner inviting them to take it; nodding the head indicating that one is listening. The emphasis of this category is neither reference nor communication but on gestures as mediators for cooperative action.¹ Interactional gestures thus concern regulating the form of interactions, including the possible regulation of communicative interactions but do not generally convey any of the content in communication. Interactional gestures are similar to class 1 manipulative gestures in the sense that they influence the environment, but in contrast to class 1, they influence the "animated environment" – *doing something to influence human agents (or other agents) in the environment*, but not by conveying symbolic or referential content.²
- 5) **Referential/Pointing Gestures.** These are *used to refer to or to indicate objects (or loci) of interest* – either physically present objects, persons, directions or locations the environment – by pointing (*deixis* – showing), or indication of locations in space being used as proxies to represent absent referents in discourse. Deictic gesture can involve a hand, finger,

¹Note that we are using the word "cooperative" in a sense that treats regulating communication or interaction as an instance of cooperation.

²Some more subtle examples include putting one's hand on another person's arm to comfort them. Such actions, and others involving physical contact, may be quite complex to interpret as understanding them may require understanding and modeling the intent of one person to influence that state of mind of another. At this point, we class simply them with interactional gestures recognizing that future analysis may reveal deep issues of human-human interaction and levels of complexity beyond the rudimentary types of human intent considered here. A special case worthy of note is human contact with the robot, unless this is directly a manipulation of the robot's state via an interface - e.g. via button presses — which would fall into class 3 (symbolic gesture), non-accidental human contact with the robot is likely to be indicative of an intent to initiate or regulate interaction with the robot (class 4). Physical contact between humans might also involve expression of affection (kissing), or aggression (slapping, hitting) – which generally indicate types of human-human interaction it would be better for a robot to steer clear of!

other directed motion, and/or eye gaze. Checking the eye gaze target of an interaction partner is commonly used to regulate reference and interaction.³

Table I summarizes the five classes.

Data on the interaction history and context may help in determining the class of a gesture. If the class is known, then the set of possible gestures can remain large, or be narrowed significantly. Symbolic gestures (class 3) correspond to discrete symbols in a finite set, of which there may be only be a small number according to context or size of the given repertoire of the given symbolic gestural code. Interactional gestures (class 4) are likely to comprise a small, constrained class. Class 1 gestures are either “irrelevant”, or to be understood by seeking the intent of the associated motor action or object manipulation (e.g. grasping or throwing an object, arms moving as a side effect of walking). Class 5 (referential and pointing gestures) comprise a very limited class, although pointing can also at times carry affective force (e.g. hostility).

Knowledge of specific conventional codes and signs can help the identification of particular signs within class 3, and also in determining that the gesture in fact belongs to class 3, i.e. is a symbolic communicative signal. Machine learning methods such as Hidden Markov Models may be used successfully to learn and classify gestures for a limited finite set of fixed gestures (e.g. [18]). It seems likely that HMM methods would be most successful with class 3 (symbolic gestures) or within narrow domains within other classes (manipulative grasps with class 1), but how successful they would be at differentiating between classes or for whole classes remains uninvestigated at present.

V. IMPORTANT ISSUES

A. Target and Recipient of a Gesture

If a gesture is used interactively or communicatively (classes 2-5), it is important to recognize whether the gesture is directed toward the current interaction partner (if any) — which may be the robot, another person (or animal) present in the context, or possibly neither (*target*). If pointing, what is the person pointing to? Who is the pointing designed to be seen by? (*recipient*). If speaking, to whom is the person speaking? If the gesture is targeted at or involves a contact with an object, this suggests it may belong to class 1 (or possibly 5, even without contact). A gesture of bringing an object conspicuously and not overly quickly toward an interaction partner is manipulative (in the sense explained in the discussion of class 1, since an object is being manipulated), but it may well at the same time also be a solicitation for the partner to take the object (class 4). Similarly if the partner has an object, an open hand conspicuously directed toward the partner or object may be a solicitation for the partner to give the object (class 4).

³Eye gaze following develops and supports joint attention already in preverbal infants. Language, including deictic vocabulary (e.g. demonstratives such as the words “these” and “that”), and other interactional skills, typically develop on this scaffolding (see [10]).

B. Multipurpose Gestures

It is possible for a single instance of a particular gesture to have aspects of more than one class or to lie intermediate between classes. As mentioned above, handing over an object is both class 1 and 4. And, for example, holding up a yellow card in football has aspects of classes 1 and 3, object manipulation and conventional symbolic signal. Many ritualized symbolic gestures (class 3) also can be used to initiate or regulate interaction (class 4), e.g. the ‘come here’ gesture: with palm away from the recipient, moving the fingers together part way toward the palm; waving forearm and open hand with palm facing recipient to get attention. More complex combinations are possible, e.g. a gesture of grasping designed by the human to be seen by a recipient interaction partner and directed toward a heavy or awkwardly-sharped target object as a solicitation of the partner to cooperatively carry the object with the gesturer (classes 1, 4, 5).

C. Ritualization: Movement into Classes 3 and 4

Gestures that originate in class 1 as manipulations of the non-animate environment and the person’s relationship to it may become *ritualized* to invite interactions of certain types, e.g., cupping the hand next to the ear can indicate that person doing it cannot hear, so that the interaction partner should speak up. Originally cupping the hand near the ear served to improve a person’s ability to hear sounds in the environment from a particular direction (class 1), but it may be intended to be seen by a conversational partner who then speaks up (class 4). The hand cupped at the ear can even be used as a conventionalized symbol meaning ‘speak up’ (class 3). Other examples of ritualization toward regulation of interaction and also symbolic gesture include mimicking with two hands the motions of writing on a pad as a signal to a waiter to ask for the bill; miming a zipping action across the mouth to indicate that someone should be ‘shut up’; or placing a raised index finger over lips which have been pre-formed as if to pronounce /sh/.

D. Cultural and Individual Differences

Different cultures may differ in their use of the various types of gesture. Some symbolic gestures such as finger signs (e.g. the “OK” gesture with thumb and index finger forming a circle) can have radically different interpretations in other cultures, or no set interpretation depending on the culture of the recipient (e.g. crossing fingers as a sign of wishing for luck, or the Chinese finger signs for some numbers such as 6, 7, 8). Tilting the head back (Greece) or nodding the head (Bulgarian) are used symbolically for ‘no’, but would certainly not be interpreted that way in many other cultures. Cultures also differ in their types and scope of movement in (class 2) expressive gestures: Consider, for example, the differences of rhythm, prosody, hand motions, eye contact, and facial expressions accompanying speech between British, Italian, Japanese, and French speakers.

Within cultures, differences between different individuals’ uses of gestures can be regional, restricted to particular

CLASSIFICATION OF GESTURAL CLASSES AND ASSOCIATED (LIMITED) CATEGORIES OF HUMAN INTENT		
CLASS	NAME	DEFINING CHARACTERISTICS AND ASSOCIATED INTENT
1	'IRRELEVANT' AND MANIPULATIVE GESTURES	INFLUENCE ON NON-ANIMATE ENVIRONMENT OR HUMAN'S RELATIONSHIP TO IT; manipulation of objects, side effects of motor behavior, body motion
2	SIDE EFFECT OF EXPRESSIVE BEHAVIOUR	EXPRESSIVE MARKING, (NO SPECIFIC DIRECT INTERACTIVE, SYMBOLIC, REFERENTIAL ROLE) associated to communication or affective states of human
3	SYMBOLIC GESTURES	CONVENTIONALIZED SIGNAL IN COMMUNICATIVE INTERACTION; communicative of semantic content (language-like)
4	INTERACTIONAL GESTURES	REGULATION OF INTERACTION WITH A PARTNER; INFLUENCE ON HUMAN (OR OTHER ANIMATED) AGENTS IN ENVIRONMENT BUT GENERALLY WITH LACK OF ANY SYMBOLIC/REFERENTIAL CONTENT used to initiate, maintain, regulate, synchronize, organize or or terminate various types of interaction
5	REFERENTIAL/POINTING GESTURES	DEIXIS; INDICATING OBJECTS, AGENTS OR (POSSIBLY PROXY) LOCI OF DISCOURSE TOPICS, TOPICS OF INTEREST; pointing of all kinds with all kinds of effectors (incl. eyes): referential, topicalizing, attention-directing

TABLE I

Five Classes of Gesture. See text for explanation, details and examples. Note that some occurrences of the same physical gesture can be used in different classes depending on context and interactional history; moreover, some gestures are used in a manner that in the same instance belongs to several classes (see text for examples).

social groups within the culture, and vary in particularities (such as speed, repertoire, intensity of movement, etc.) between individuals according to preference or ontogeny. Elderly and young may employ gestures in different ways.

VI. INFERRING THE INTENT OF GESTURE

Being able to identify details of gestural kinematics and even to classify into one of the above classes gives us only starting points for inferring the intent of the person making the gesture due to frequent ambiguity. Resolving this points to the important roles of context and interactional history. Thus, it is necessary to develop operational methods for

recognizing the class of gesture in a particular context.⁴ If the interactional context of recent activity in which a gesture occurs is known, this can suggest possibilities for which classes (and subtypes) of gesture might be involved. Information on the state of human (e.g. working, thirsty, talking, ...) often can limit the possibilities. Data on the following could help the robot classify the gesture and infer the intent of the human:

- (a) the activity of the gesturer is known,
- (b) previous and current interaction patterns are remem-

⁴Knowledge of the immediate context in some cases needs to be augmented by taking into account of the broader *temporal horizon* of interactional history (cf. [14]).

bered to predict the likely current and next behaviour of the particular person,

- (c) objects, humans and other animated agents in the environment are identified and tracked.
- (d) the scenario and situational context are known (e.g. knowing whether a gesture occurs at a tea party or during a card game).

A programme to apply the above classification can be developed as follows: (1) Identify the many, particular gestural motions that fit within each of the five classes. Some gestural motions will appear in more than one class. For example, the same mechanical motion of putting a hand and arm forward with the forearm horizontal and the hand open could indicate preparation to manipulate an object in front of the human (class 1), to show which object is being referred to (class 5), or to greet someone who is approaching, or to ask for an object to be handed over (both class 4). (2) Gestural motions identified as belonging to several classes need to be studied to determine in which contexts they occur: determining in which class(es) particular a instance of the gesture is being used may require consideration of objects and persons in the vicinity, the situational context, and the history of interaction. (3) Systematic characterizations of a physical gestural motion together with interactional contexts in which they are occur could then be used to determine the likely class. (4) Deploy on-board characterizations of the relationships between classes and kinematic gestural motions for a range of typical interactional contexts to infer intent and guide robot behaviour. (5) Updating the Interaction History: Attribution of intent related to gesture can then feedback into understanding of the situational context, including motivational state of the human performing the gesture, and becomes part of the updated interaction history, which can then help in inferring intent from ensuing gestures and activity.

VII. HEURISTIC-BASED FAST RECOGNITION AND DISAMBIGUATION OF GESTURES WITH A TIME-OF-FLIGHT DEPTH SENSOR

There are mainly three methods to recognize and detect gestures: model-based approaches fit a kinematics model into the scene observed by sensors, recognition based on classifiers use learning algorithms to label gestures, and heuristic-based methods which directly search for hints related to a gesture. Depending on the context of the overall robotic control system, all of these may be of use. The model-based approach is followed by researchers in the Cognirion project, and, as shown above, this must be augmented by contextual and situational knowledge. The goal is to develop algorithms that geometrically fit a model maintained by the robot into the current scene observed by stereo vision systems and a time-of-flight depth sensor proposed in [15]. Apart from this exhaustive approach there is also work related to a computationally much cheaper heuristic-based method only using data delivered by the depth sensor. The motivation for this is two-fold. Firstly, a ‘quick’ check of the existence of humans in the close

vicinity of the robot and a first basic evaluation of possibly important gestures can be used directly for communication. Secondly, outputs of a fast algorithm related for instance to body, head or arm positions can serve to trigger more detailed investigation by e.g. model-based algorithms. Additionally, the data can be used to initialize model fitting.

The heuristic approach first divides the depth scene observed by the sensor into consecutive depth intervals each having a fixed distance and size. The two intervals containing the most measurements are used for binary segmentation of the humans profile. Within the profile the algorithm searches for a human’s center point by summing all pixels belonging to the profile and averaging their co-ordinates. From this point the algorithm searches upwards and determines a bounding box for the head including the neck by incorporating estimates of the shoulder end points. They can be found as being the left and right extremes of the profile at a height roughly at the bottom of the head. Interestingly, the height of the bounding box around the head plus the width of the shoulder can give an estimate of the length of the upper arm as described in medical statistics. Based on this information four cases can be distinguished:

- Outstretched arm away from the body
- Outstretched arm up or down
- Bent arm next to body
- Bent arm in front of body

For each case further heuristic algorithms are used to determine the hand position and orientation. This can be used to recognize and discriminate between basic gestures. See Figure 2 for a visualization of how the program finds a WAVE gesture and SHAKE HANDS gesture; both are interactional gestures (class 4). By using additional information such as orientation and distance of the human towards the robot and internal state of the robot, tentative disambiguations between similar gestures have been made.

VIII. CONCLUSIONS, NEXT STEPS AND THE FUTURE

In order to infer the intent of a human interaction partner, it may be useful to employ a classification of gesture according to some major types – five in the tentative classification proposed here – whose intent may be (1) absent / directed to objects or environment, (2) incidentally expressive, (3) symbolic, (4) interactional, or (5) deictic. A summary of the classes is given by Table I.

In order to deploy the inference of intent on robots interacting with humans it is necessary to operationalize the distinctions between these (sometimes overlapping) classes. This may require the use of knowledge of human activity, recognition of objects and persons in the environment, and previous interactions with particular humans, as well as knowledge of conventional human gestural referencing and expression, in addition to specialized signaling codes or symbolic systems.

Work in Cognirion now focuses on the organization of the robot decisional abilities and more particularly on the management of human interaction. There is explicit

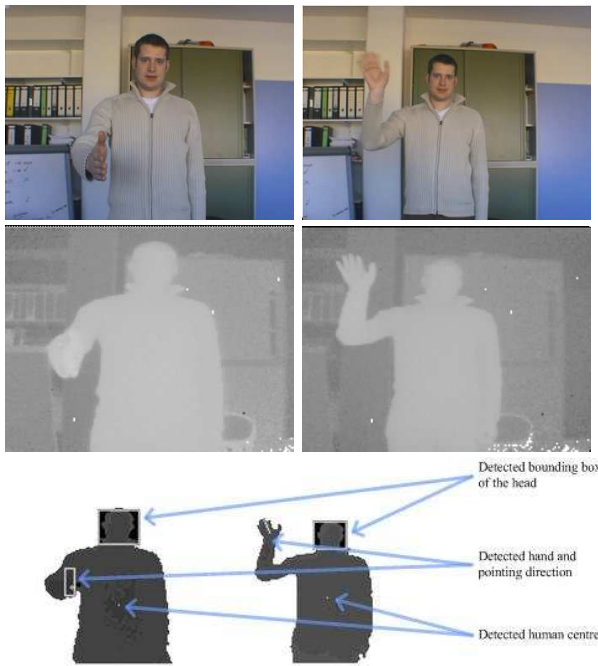


Fig. 2. **Heuristic-based fast recognition and disambiguation of interactional gestures.** Detecting bounding box of head, hand and its orientation, and center of human (top row: color image of gesturing human in robot's vicinity. second row: depth image derived from camera images. Left column: SHAKE HANDS gesture, right column: WAVE gesture.

management of the interactions between the robot companion and its human partners. This requires essentially task-oriented processes that each consist of establishing a common goal, achieving it and verifying commitment of all agents involved during the task performance. Indeed, perception of the human partner is one essential source of information all along the human-robot interaction process from the detection of human presence to the monitoring of human activity and the continuous estimation of its commitment level to a joint goal. This viewpoint is compatible with and served by the classification of gestures proposed here. It also helps us to operationalize use of the classification. Indeed, gestures of type 3 and many of type 1 may be considered as task-oriented and the inference of their intent can be done relative to the task at hand. Gestures of type 4 include generic interactional gestures that may serve to manage the session itself: inviting the robot to start an interaction, suspending or stopping an interaction session, etc. Many gestures of type 4 are consequently task independent.

The classification presented here suggests some requirements for the design and implementation of systems inferring intent from gesture based on this classification. These requirements might be realized in a variety of different ways using, e.g. continuous low-key tracking or more detailed analysis, event-based and/or scenario-based recognition, and prediction of human activity based on models of human activity flows (with or without recognition of particular humans and their previous interactions), depending the particular needs of the given human-robot

interaction design and the constraints and specificity of its intended operational context. Design of a robot restricted to helping always the same user in the kitchen environment would be quite different from one that should be a more general purpose servant or companion in a home environment containing several adults, children and pets, but the classification presented here is applicable in informing the design of gesture recognition for inferring intent in either type of system, and for designing other HRI systems.

Finally, effective human-robot interaction will require generation of gestures and feedback signals by the robot. The classification given here can suggest categories of robotic gestures that could be implemented to improve the *legibility* to humans of the robot's behaviour, so that they will be better able to understand and predict the robot's activity when interacting with it.

REFERENCES

- [1] BERNADIN, K., OGAWARA, K., IKEUCHI, K., AND DILLMANN, R. A sensor fusion approach for recognizing continuous human grasping sequences using hidden markov models. *IEEE Trans. on Robotics* 21, 1 (2005), 47–57.
- [2] BILLARD, A., EPARS, Y., CALINON, S., CHENG, G., AND SCHAAAL, S. Discovering optimal imitation strategies. *Robotics & Autonomous Systems*, Special Issue: Robot Learning from Demonstration 47, 2-3 (2004), 69–77.
- [3] BREAZEAL, C., AND ARYANANDA, L. Recognition of affective communicative intent in robot-directed speech. *Autonomous Robots* 12, 1 (2002), 83–104.
- [4] CONDON, W. S., AND OGSTON, W. D. A segmentation of behavior. *Journal of Psychiatric Research* 5 (1967), 221–235.
- [5] DUBOIS, E., NEDEL, L. P., FREITAS, C. M. D. S., AND JACON, L. Beyond user experimentation: notational-based systematic evaluation of interaction techniques in virtual reality environments. *Virtual Reality* 8, 2 (2005), 118–128.
- [6] *Gesture*. ISSN: 1568-1475. John Benjamins Publishing Co., The Netherlands http://www.benjamins.com/cgi-bin/t_seriesview.cgi?series=GEST.
- [7] HALL, E. T. *The Dance of Life: The Other Dimension of Time*. Anchor Books, 1983.
- [8] IBA, S., PAREDIS, C. J. J., AND KHOSLA, P. K. Interactive multimodal robot programming. In *Proceedings of the 2002 IEEE International Conference on Robotics and Automation, Washington D.C., May 11-15, 2002* (2002).
- [9] KENDON, A. Movement coordination in social interaction: Some examples described. *Acta Psychologica* 32 (1970), 100–125.
- [10] KITA, S., Ed. *Pointing: Where Language, Culture and Cognition Meet*. Lawrence Erlbaum Associates, Inc, 2003.
- [11] MEY, J. *Pragmatics: An Introduction*. Blackwell Publishers, 2001.
- [12] MILLIKAN, R. G. *The Varieties of Meaning: The 2002 Jean Nicod Lectures*. MIT Press/Bradford Books, 2004.
- [13] MINERS, W. B. *Hand Gesture for Interactive Service Robots*. MSc Thesis, The University of Guelph, Faculty of Graduate Studies, August 2002.
- [14] NEHANIV, C. L., POLANI, D., DAUTENHAHN, K., TE BOEKHORST, R., AND CAÑAMERO, L. Meaningful information, sensor evolution, and the temporal horizon of embodied organisms. In *Artificial Life VIII* (2002), MIT Press, pp. 345–349.
- [15] OGGIER, T., AND LANG, G. K. User manual: Swissranger 2, December 2003, Rev. March 2004.
- [16] SHNEIDERMAN, B. *Designing the User Interface: Strategies of Effective Human-Computer Interaction*, 3rd ed. Addison-Wesley, 1998.
- [17] VOLTERRA, V., CASELLI, M. C., CAPIRCI, O., AND PIZZUTO, E. Gesture and the emergence and development of language. In *Beyond Nature-Nurture: Essays in Honor of Elizabeth Bates*, M. Tomasello and D. I. Slobin, Eds. Lawrence Erlbaum Associates, 2004.
- [18] WESTEYN, T., BRASHEAR, H., ATRASH, A., AND STARNER, T. Georgia tech gesture toolkit: Supporting experiments in gesture recognition. In *ICMI 2003: Fifth International Conference on Multimodal Interfaces* (2003), ACM Press.