

G

Gesture, Posture, Facial Interfaces



Dongheui Lee
Technical University of Munich (TUM),
Munich, Germany
German Aerospace Center (DLR), Weßling,
Germany

Synonyms

[Bodily expression](#); [Body language](#); [Kinesics](#);
[Nonverbal communication interfaces](#)

Definition

A *gesture* is a form of nonvocal communication in which visible bodily actions communicate particular messages. Gestures include movement made with body parts, such as hands, arms, fingers, head, and legs. One gesture type is gestures that convey meaning by themselves and are assumed to be deliberately performed by the speaker. They are conventionalized symbols and strongly culture-dependent. Another type is conversational gestures, which accompany the speech but do not deliver the semantic content of the accompanying speech. An example is simple, repetitive, rhythmic movements like the beats.

Posture interface conveys information about interpersonal relations, personality traits (e.g.,

confidence, submissiveness, and openness), social standing, and current emotional states through the position and orientation of specific body parts, which can be expressed either in a fixed coordinate on the own body or in the relative coordinate to the other agent. Often postures are considered as snapshots of movements.

Facial interface is combinations of eye, eyebrow, lip, nose, and cheek movements by moving the muscles beneath the skin of the face. Facial expressions convey emotions of an individual, for instance, a raise of eyebrow for skepticism. They are a primary means of conveying social information between humans.

Overview

In contrast to interacting with machines, humans can master the complexities of interpersonal communication apparently without difficulty. One of the reasons for such human-to-human communication is multimodality. For example, we use gestures, postures, and facial expression. Some researchers conclude that gesture, posture, and facial interfaces can provide a significant amount of important information through nonverbal communication during interpersonal interactions. In robotics, gesture, posture, and facial interfaces appeared as a way to realize intuitive natural communication tools between a human and a robot without using complex artificial interface. These gesture, posture, and facial interfaces are beneficial because these can provide user-friendly

interfaces that do not require knowledge in complex new technology and that are straightforward usable without the need of getting to know a new interface. The goal is to realize natural human-robot interaction by developing robots equipped with human's nonverbal communication interfaces. These nonverbal communication interfaces are employed in order to understand human's states and to convey robot's internal states in a human-robot interaction domain.

Key Research Findings

Key research findings of gesture, posture, and facial interfaces are grouped into two categories in this chapter: control interfaces to robots and robots with affective computing.

Control Interfaces to Robots

One of the most typical interfaces to control machines (e.g., televisions, actuated toys, video games, etc.) is a joystick-like remote controller. Instead of such an external device, gestures, postures, and facial expressions can be also used as commands to control machines. Recently it is not difficult to find commercially available drones equipped with hand gesture commands. Research on gesture recognition, pose estimation, and retargeting of human movements to a machine could help to realize controlling robots using nonverbal cues.

Control Command Recognition

One way for controlling machines is based on command recognition. The basic idea is simple: from the signals of human gestures, postures, and facial movements, human commands are recognized, and then the recognized commands trigger the corresponding movement primitives (e.g., go left/right, go forward/backward, turn clockwise, etc.) of the machines (Ott et al. 2010; Waldherr et al. 2000). In order to recognize command gestures, features are selected (or descriptors are calculated), classifiers are trained, and then commands are recognized using the trained classifiers.

As a first step in this pipeline, appropriate features (like joint angles, 3D coordinates of

body joints, etc.) are selected. Instead of manual selection of features, new descriptors could be developed for robust gesture recognition (Falco et al. 2017, 2018). Invariance properties in terms of viewpoints, translations, and scales (Lee et al. 2017; Saveriano and Lee 2013; Soloperto et al. 2015) are investigated to design new invariant descriptors of human gestures. There have been various studies on gesture recognition (Mitra and Acharya 2007), especially hand gesture recognition (Kumar et al. 2017; Neto et al. 2013; Molchanov et al. 2016), facial expression (Pantic and Rothkranz 2000; Yacoob and Davis 1996; Fasel and Luetttin 2003), and body gestures (Lee and Nakamura 2005). Tools for gesture recognition range from hidden Markov models (HMM) and support vector machine (SVM) to artificial neural networks (ANN) and recurrent neural networks (RNN), which have received a great amount of attention recently using big data. Control interfaces based on command gesture recognition have been applied to many robots, including humanoid robots (Ott et al. 2010), mobile robots (Waldherr et al. 2000), wheel chairs (Fehr et al. 2000), and collaborative robots (Magnanimo et al. 2014; Lee et al. 2009; Parusel et al. 2014).

Retargeting Based on Pose Estimation

The other approach for controlling machines is so-called retargeting, where captured human poses are mapped to robot poses. Several approaches are existing for whole body human motion mapping to humanoids (Ott et al. 2008; Dariush et al. 2008; Hu and Lee 2012; Hu et al. 2014). Most of retargeting approaches use a commercially available motion capture system, in order to measure human movements precisely, including facial muscle movements. Widely used motion capture systems are optical ones, which usually consist of multiple cameras and reflected markers in a controlled environment, as well as inertia sensor-based systems. Such a motion capturing technique is widely applied in animations and movies, where human's body and facial expression is mapped to an animated character, in order to create natural movements and vivid expression. Geminoid,

whose appearance closely resembles a specific human, can imitate human's facial expression, head rotation, and mouth opening (Nishio et al. 2007). Similarly human's hand motion can be captured by a data glove and imitated by a robotic hand (Schmidts et al. 2011).

In contrast to using commercially available camera-based motion capture systems or data gloves, researchers in computer vision community have developed a number of algorithms to estimate human pose from raw camera images: human's whole body pose estimation (Sun et al. 2012; Toshev and Szegedy 2014; Zhang et al. 2012; Liu et al. 2011; Lee and Nakamura 2014, 2007), hand pose estimation (Yuan et al. 2018; Li et al. 2018; Li and Lee 2018; Ge et al. 2018), and face reconstruction (Cao et al. 2012; Dantone et al. 2012). When the estimation results of these methods become robust to noises in sensors and environments and ready to directly apply them to control robots safely, the higher impact for human-robot interaction applications is expected.

Robots with Affective Computing

During interaction, people communicate emotion through gesture, posture, and facial expression. Charles Darwin discussed the role of facial expression and body movement in communicating emotion (Darwin 1872). Facial expression and bodily expression are reported to be congruent when interpreting emotions (Kret et al. 2011).

Affective computing, also called artificial emotional intelligence, is computing that relates to, arises from, or deliberately influences emotion or other affective phenomena (Picard 1997). According to Hook (2009) affective loop is the interactive process in which the user of the system first expresses emotions involving own body (e.g., through gestures); and the system then responds by generating affective expression which "in turn affects the user making the user respond and step-by-step feel more and more involved with the system." Research goes into two directions: first *affective input* to interpret the emotional state of humans and second *affective output* to respond to those human's emotions for appropriate social engagement.

Affective Input

In affective input tools, the emotional state of humans is detected and recognized from various sensory data. The process for affective input can be described as following the four steps (Karg et al. 2013): (1) measuring movements by sensors, (2) segmenting movements into primitives, (3) constructing features, and (4) detecting emotion. At the first step, a wide range of sensors can be employed in order to record affect expressive movements, such as optical motion capture, inertial motion tracking, and data gloves, as well as physiological data like temperature, pulse, heart rate, etc. Also, computer vision algorithms could be applied in order to measure physical movement of a human from sequential camera images. The second step is needed in order to handle online streaming signals. One of simplest approach is sliding time window. Also there are complicated segmentation approaches (Fod et al. 2002; Kulic et al. 2012; Caccavale et al. 2018) which aim at finding the precise boundaries of movement primitives. The third step is preprocessing of signals: features are manually selected directly from sensor inputs, or selected/calculated based on perceptual study in psychology (Velocity, acceleration, duration of the movement, and finger motion range are reported as important features in perception of affect.), or extracted automatically by machine learning algorithms. The features could be further processed in order to lie in a lower dimensional space. In the final step, which is the main part of the affective input, the extracted meaningful features of human movement are fed into machine learning algorithms to detect emotions.

Models of emotion can be categorized into three models: categorical models (Ekman 1972), dimensional models (Russell and Mehrabian 1977; Mehrabian 1996), and componential models (Plutchik 2001). Categorical models (Ekman 1972) are based on the underlying assumption that humans universally express and recognize a set of discrete primary emotions which include happiness, sadness, fear, anger, disgust, and surprise. Herein, emotions are classified into a set of distinct classes that can be recognized by applying standard pattern

recognition procedures. In dimensional models (Russell and Mehrabian 1977; Mehrabian 1996), emotions are modeled along a set of continuous latent dimensions. The PAD model describes an emotional state with values of pleasure (or valence), arousal, and dominance. Therein, pleasure represents how pleasant or unpleasant a feeling is. Arousal corresponds to the level of activation and physical activity. Dominance represents the sense of control over the emotion. Their continuous nature allows to model rich and complex emotions in theory. Categorical labels can be mapped to the continuous PAD space, e.g., happiness, amusement, and contentment are related to high pleasure, whereas anger, fear, and sadness are related to low pleasure. Componential models (Plutchik 2001) emotions are hierarchically represented where complex emotions are combinations of basic emotions.

Categorical models are most widely used for emotion recognition (Gunes and Piccardi 2005; Saha et al. 2014). Machine learning algorithms are applied to classify emotions, for example, six basic emotions (anger, happiness, fear, disgust, sadness, and surprise) proposed by Ekman (1972). Frequently applied classifiers include linear discriminant classifier, support vector machine, hidden Markov model, k-nearest neighbor, artificial neural networks, decision tree, dynamic time warping, naive Bayes, etc. Fewer studies have applied a dimensional representation (Glowinski et al. 2015; Karg et al. 2010a; Kleinsmith and Bianchi-Berthouze 2007). While its continuous nature allows to model rich and complex emotions in theory, the automatic emotion recognition becomes challenging. Thus, when using a dimensional representation, each continuous dimension is divided into a few categories (low, neutral, and high), and a classifier is trained. Alternatively regression technique like support vector regression can be applied.

Facial interfaces are widely studied for emotion recognition, also for person-dependent and person-independent emotion recognition. Other nonverbal signals are also investigated: upper body motion (Gunes and Piccardi 2005), whole body motion like walking (Karg et al. 2010a) and dancing (Park et al. 2004), as well as head

movements (Gunes and Pantic 2010b; Monkarese et al. 2012). For further details, survey papers on emotion detection can be refereed (Gunes and Pantic 2010a; Kleinsmith and Bianchi-Berthouze 2013; Karg et al. 2013).

Affective Output

There have been a wide range of study on tools for affective output. In order to express emotions in robots, researchers tried to apply the *illusion of life*, which is the Disney's 12 principles on how to create believable and expressive characters (Thomas and Johnston 1995), to robots. One of the attempts is enriched hardware features to introduce expressive features, such as actuated eyes, pupils, and muscles. The design of computational devices is proposed to exhibit either innate emotional capabilities or convincingly simulating emotions. Examples are emotion expression for characters in the interface or regulating robot behaviors in order to enrich and facilitate interactivity between human and machine.

Affective output has been investigated for anthropomorphic, zoomorphic, or cartoon-like embodiments. In the design of *embodiment*, forms and expectations are important because it establishes social expectations. For example, a dog-like robot would be treated differently than a humanoid robot. Anthropomorphic robots facilitate tendency to attribute human characteristics to robots, such as NAO from Softbank Robotics, Kobian (Endo and Takanishi 2011), and CB2 (Minato et al. 2007). Arms and hands are designed to exhibit human-like gestures, e.g., pointing, shaking hands, shrugging shoulders, etc. In order to limit human expectation in a low level and still to create social interaction with humans, zoomorphic design, imitating living creatures, is taken into account, such as Paro (Wada et al. 2008), AIBO (Fujita and Kitano 1998), MIT Leonardo, iCat robot from Philips research, etc. For example, Paro therapeutic robot (Wada et al. 2008) is designed followed by appearance of a baby harp seal, a rather nonfamiliar animal, so that people can accept Paro easily without preconceptions. Mechanical faces are used to communicate with humans via facial expressions like

Kobian (Endo and Takanishi 2011) and Kismet (Breazeal 2002).

It is necessary for a social robot to model and select behaviors based on affective goals and conditions. A robot must be able to properly express those actions with an adequate affective display. Research addressing empathy in social robots has studied on automatic and involuntary process, such as mimicking the user's affective states. In Hegel et al. (2006) and Riek et al. (2010) users reported that interaction with robots with affective expression was more appropriate than ones without affective expression. The study in Kishi et al. (2013) reported that integration of dynamic emotion expression through facial and whole body pattern based on external stimuli makes a humanoid robot more appealing to humans. In Gonsior et al. (2012) when a robot adapts its own mood to the user's mood, it increases empathy, and the subject is more willing to help the robot.

Emotions are expressed not only through verbal interfaces, like tone and speech, but also nonverbal interfaces. There have been a lot of research on emotion generation through facial expressions (Zecca et al. 2008). Besides facial expressions (including iris extension and gaze direction), body movements, such as gait pattern and sitting and standing posture, can convey emotions (de Gelder et al. 2010). According to de Gelder et al. (2010), bodily expressions are suitable for communication at over larger distance and facial ones for communication in proximity. Studies in Montepare et al. (1987), Crane and Gross (2007), and Roether et al. (2009) reported that affective states can be expressed during walking and recognized by human observers and that gait parameters and kinematics (e.g., walking speed, shoulder and elbow range of motion, and head orientation) are modulated according to the affective states.

Extensive survey on emotional behavior generation can be found in Gunes et al. (2011), Bethel and Murphy (2008), and Karg et al. (2013). In affective expression generation for robots, given a desired affective state and movement type, the movement is modulated, and trajectory is generated and controlled on a real

robot. For movement modulation for affective expression, extracted movements variables are often linked with emotional state. For example, in Destephe et al. (2013) emotional gait parameters from the captured human gait data were extracted to build emotion mental model and to generate emotional gait. Also, a relationship between motion parameters and affective state in the PAD model can be used for emotional robot behavior generation. In Beck et al. (2010) key poses for affective states are arranged in the valence and arousal space, and then a new expression is synthesized by blending between the joint angles of the key poses. The gait of a hexapod robot could display affective states by linking the gait parameters and low or high pleasure, arousal, and dominance (Karg et al. 2010b). In Nakagawa et al. (2009) a general approach which modifies an arbitrary motion to control affective nuances is discussed. Velocity and expansiveness of the motion correlate with arousal, and the basic posture relates to the expressed level of valence.

Examples of Application

Social Robots

Social robots need to engage us not only on a cognitive level but also on an emotional level. In Breazeal and Scassellati (1999) it is argued that in order to interact socially with a human, a robot must convey intentionality, that is, the human must believe that the robot has beliefs, desires, and intension. Also culture awareness becomes important, especially differentiating universal gesture from culture-specific gestures. There are some gestures which have different meaning in different cultures. For instance, the thumbs-up gesture means good in some countries, while it is insulting in other countries. In most cultures the head nod signalizes agreement, but not in India. Using humanoid robot Kobian, cross-cultural study on human-robot greeting interaction was carried out with Egyptians and Japanese (Trovato et al. 2013). Entertainment robots and educational robots can be included in the category of social robots. Examples of social robots are NAO, Paro, AIBO (Fujita

and Kitano 1998), Leonardo, Kismet (Breazeal 2002), etc. Kismet is capable of conveying intentionality through facial expression and behavior (Breazeal and Scassellati 1999). Kismet can display expression analogous to happiness, sadness, surprise, boredom, anger, calmness, displeasure, fear, and interest. Many social robots have been built and used for a teaching aid for autistic children where kids learn social skills for communication with other people (Dautenhahn 2004).

Animated Characters

People use verbal (voice) and nonverbal channels (gesture, posture, facial expression) in face-to-face interaction. In contrast to face-to-face interaction in a same physical world, we can communicate with others also using virtual avatars and telepresence robots. Such virtual avatars are applicable in computer games, virtual training environments, interactive storytelling systems, etc. Interaction with virtual reality involves often cabled data gloves but recent techniques on vision-based acquisition of hand and figure posture, such as leap motion and RGB-D sensors. Recently it also became common to send messages via mobile phone to others using emoticons and to create personal emoticons, in which an animated character mimics facial expression of the user.

Assistive Robots

Gesture, posture, and facial interfaces are widely used in robotic assistive systems for handicapped people and elderly. Prostheses are categorized into motionless cosmetic, body-powered (operated by a cable running through the shoulder girdle), and self-powered (operated by electric motors) prostheses. In self-powered mobility assistive robots, while some like prosthesis and exoskeleton are controlled by using bio-signals like EMG and EEG, others are controlled by non-bio-signals. Depending on the disabilities of users, the control interface varies from conventional joystick to head pose, eye gaze, tongue pad, facial expression, and gestures (Fehr et al. 2000). Using trained classifiers, for instance, input signals are mapped to steering commands for control of a robotic wheelchair.

Another example of assistive robots is automated sign language recognition. Sign languages are the main media for the daily communication between and with the deaf. Automated sign language recognition will improve communication between the hearing and the deaf.

Besides for handicapped people, there has been a lot of progress also for assistive robots for nonhandicapped. For instance, many functionalities for assistive driving have been developed, such as lane keeping, lane changing, speed control, and distance keeping. Vision-based acquisition of the eye blinks to detect the tiredness of a driver is another example to use facial interface.

Collaborative Robots

Paralinguistic communication plays an important role in coordinating joint action during collaborative tasks. A team establishes common ground, which is mutual beliefs regarding the current state of the tasks and roles and responsibility of each members. Visual perspective taking and shared attention is used to direct attention in order to coordinate joint actions about objects in the world. In Breazeal et al. (2005) Leonard robot communicates with human subjects implicitly (gaze and facial expression) and explicitly (pointing gestures, a short nod to acknowledge success of tasks). In Parusel et al. (2014) a collaborative robot communicates with a human partner through haptic interface and hand gestures in order to accomplish joint assembly tasks. In Medina Hernández et al. (2012) a human and a mobile robot accomplished joint-carrying task through verbal and nonverbal communication.

Future Directions for Research

Social Robots from Lab to Real World

Gesture, posture, and facial expression can improve usability of complex machines. A lot of efforts toward designing intuitive interface in human-robot interaction have been taken for the last decades. Besides their usages as commanding interfaces, these nonverbal interfaces will become essential for us to communicate with social robots. While most existing social robots

are tested in lab environment, social robots must move from lab to real settings. This includes many challenges. First, a robot should be capable of handling dynamic changes in a real daily environment. Dynamic features of emotions might become essential to adapt to changing environments. Affective expression should be studied not only in short terms (e.g., emotion) but also in long terms (e.g., mood). In order to accomplish long-term human-robot symbiosis, life-long learning, adaptation, and memory in both short- and long-terms are needed to be further investigated.

Multimodal Approaches

It is reported that combination of several modalities, through facial expressions, speech, and bodily expressions, provides more robust estimation for emotion recognition and enhances the believability and congruency for emotion generation. Multimodal approaches for robots with affective computing will broaden the range of affective states. To date, the generation of affect-expressive movements is limited to a small set of movement types and focusing on upper body movement. It is desired to broaden the range of affect-expressive movements. Besides affective expression, new types of nonverbal communication can increase connectivities for people, including motor-handicapped populations. These links include physiological signals, brain-machine interfaces, and robotics platforms.

Cognition-Enabled Robots

Most existing robotic approaches for interacting with humans through gesture, posture, and facial expression do not take *context* information into account. However, context knowledge regarding environment, interaction partner, and activities can lead to more cognitive interaction with humans. Learning affective expression and movement primitives from humans (Calinon and Lee 2017) will enable robot to become personalized to specific users.

Cross-References

- ▶ Cognitive Human-Robot Interaction
- ▶ Human Detection and Tracking

- ▶ Human-Robot Collaboration
- ▶ Motion Capture
- ▶ Physical Human-Robot Interaction
- ▶ Social Human-Robot Interaction

References

- Beck A, Hiole A, Mazel A, Canamero L (2010) Interpretation of emotional body language displayed by robots. In: International workshop affective interaction in natural environments (AFFINE), pp 37–42
- Bethel C, Murphy R (2008) Survey of non-facial/non-verbal affective expressions for appearance-constrained robots. *IEEE Trans Syst Man Cybern Part C Appl Rev* 38(1):83–92
- Breazeal C (2002) Designing sociable Robots. The MIT Press
- Breazeal C, Scassellati B (1999) How to build robots that make friends and influence people. IROS
- Breazeal C, Kidd C, Thomaz AL, Hoffman G, Berlin M (2005) Effects of nonverbal communication on efficiency and robustness in human-robot teamwork. In: IEEE/RSJ international conference on intelligent robots and systems (IROS)
- Caccavale R, Saveriano M, Finzi A, Lee D (2018) Kinesthetic teaching and attentional supervision of structured tasks in human-robot interaction. In: Autonomous robots
- Calinon S, Lee D (2017) Learning control. In: Vadakkepat P, Goswami A (eds) *Humanoid robotics: a Reference*. Springer
- Cao X, Wei Y, Wen F, Sun J (2012) Face alignment by explicit shape regression. In: Conference on computer vision and pattern recognition, pp 2887–2894
- Crane E, Gross M (2007) Motion capture and emotion: affect detection in whole body movement. In: International conference on affective computing and intelligent interaction (ACII), pp 95–101
- Dantone M, Gall J, Fanelli G, Van Gool LJ (2012) Real-time facial feature detection using conditional regression forests. In: Conference on computer vision and pattern recognition, pp 2578–2585
- Dariush B, Gienger M, Jian B, Goerick C, Fujimura K (2008) Whole body humanoid control from human motion descriptors. In: IEEE international conference on robotics and automation (ICRA), pp 2677–2684
- Darwin C (1872) *The expression of the emotions in man and animals*. Oxford University Press
- Dautenhahn K, Werry I (2004) Towards interactive robots in autism therapy: Background, motivation and challenges. *Pragmatics & Cognition*, 12:1–35. <https://doi.org/10.1075/pc.12.1.03dau>
- de Gelder B, Van den Stock J, Meeren HK, Sinke C, Kret ME, Tamietto M (2010) Standing up for the body. Recent progress in uncovering the networks involved in the perception of bodies and bodily expressions. *Neurosci Biobehav Rev* 34(4):513–527

- Destephe M, Hashimoto K, Takanishi A (2013) Emotional gait generation method based on emotion mental model – preliminary experiment with happiness and sadness. In: International conference on ubiquitous robots and ambient intelligence (URAI), pp 86–89
- Ekman P (1972) Universals and cultural differences in facial expression of emotion. In: Nebraska symposium on motivation. University of Nebraska Press, Lincoln, pp 207–283
- Endo N, Takanishi A (2011) Development of whole-body emotional expression humanoid robot for ADL-assistive RT services. *J Robot Mechatronics* 23(6). Fuji Press pp. 969–977
- Falco P, Saveriano M, Hasany EG, Kirk N, Lee D (2017) A human action descriptor based on motion coordination. *IEEE Robot Autom Lett (RA-L)* 2(2):811–818
- Falco P, Saveriano M, Shah D, Lee D (2018) Representing human motion with FADE and U-FADE: an efficient frequency-domain approach. In: *Autonomous robots (AURO)*
- Fasel B, Luetttin J (2003) Automatic facial expression analysis: a survey. *Pattern Recogn* 36:259–275
- Fehr L, Langbein WE, Skaar SB (2000) Adequacy of power wheelchair control interfaces for persons with severe disabilities: a clinical survey. *J Rehabil Res Dev* 37(3):353–360
- Fod A, Mataric MJ, Jenkins OC (2002) Automated derivation of primitives for movement classification. *Auton Robots* 12(1):3954
- Fujita M, Kitano H (1998) Development of an autonomous quadruped robot for robot entertainment. *Auton Robots* 5:7–20
- Ge Y, Li B, Yan W, Zhao Y (2018) A real-time gesture prediction system using neural networks and multimodal fusion based on data glove. In: IEEE international conference on advanced computational intelligence (ICACI), pp 625–630
- Glowinski D, Mortillaro M, Scherer K, Dael N, Camurri GVA (2015) Towards a minimal representation of affective gestures. In: IEEE international conference on affective computing and intelligent interaction (ACII), pp 498–504
- Gonsior B, Sosnowski S, Buß M, Wollherr D, Kühnlenz K (2012) An emotional adaption approach to increase helpfulness towards a robot. In: IEEE/RSJ international conference on intelligent robots and systems (IROS)
- Gunes H, Pantic M (2010a) Automatic, dimensional and continuous emotion recognition. *Int J Synth Emot* 1(1):69–99
- Gunes H, Pantic M (2010b) Dimensional emotion prediction from spontaneous head gestures for interaction with sensitive artificial listeners. In: International conference on affective computing and intelligent interaction (ACII), pp 371–377
- Gunes H, Piccardi M (2005) Affect recognition from face and body: early fusion vs. late fusion. In: IEEE international conference on systems, man and cybernetics, pp 3437–3443
- Gunes H, Schuller B, Pantic M, Cowie R (2011) Emotion representation, analysis and synthesis in continuous space: a survey. In: Proceedings of IEEE international conference on FG, workshop emotion synthesis, representation, and analysis in continuous space
- Hegel F, Spexard T, Vogt T, Horstmann G, Wrede B (2006) Playing a different imitation game: interaction with an Empathic Android Robot. In: IEEE-RAS international conference on humanoid robots (Humanoids), pp 56–61
- Hook K (2009) Affective loop experiences: designing for interactional embodiment. *Phil Trans R Soc B* 364:3585–3595
- Hu K, Lee D (2012) Biped locomotion primitive learning, control and prediction from human data. In: Proceedings of the 10th international IFAC symposium on robot control (SYROCO)
- Hu K, Ott C, Lee D (2014) Online human walking imitation in task and joint space based on quadratic programming. In: IEEE international conference on robotics and automation (ICRA), pp 3458–3464
- Karg M, Kuehnlennz K, Buss M (2010a) Recognition of affect based on gait patterns. *IEEE Trans Syst Man Cybern Part B Cybern* 40(4):1050–1061
- Karg M, Schwimbeck M, Kuehnlennz K, Buss M (2010b) Towards mapping emotive gait patterns from human to robot. In: IEEE international symposium RO-MAN, pp 258–263
- Karg M, Samadani A-A, Gorbet R, Kuehnlennz K, Hoey J, Kulic D (2013) Body movements for affective expression: a survey of automatic recognition and generation. *IEEE Trans Affect Comput* 4(4):341–359
- Kishi T, Kojima T, Endo N, Destephe M, Otani T, Jamone L, Kryczka P, Trovato G, Hashimoto K, Cosentino S, Takanishi A (2013) Impression survey of the emotion expression humanoid robot with mental model based dynamic emotions. In: IEEE international conference on robotics and automation, pp 1655–1660
- Kleinsmith A, Bianchi-Berthouze N (2007) Recognizing affective dimensions from body posture. In: International conference on affective computing and intelligent interaction (ACII), pp 48–58
- Kleinsmith A, Bianchi-Berthouze N (2013) Affective body expression perception and recognition: a survey. *IEEE Trans Affect Comput* 4(1):15–38
- Kret ME, Pichon S, Grezes J, de Gelder B (2011) Similarities and differences in perceiving threat from dynamic faces and bodies. An fMRI study. *NeuroImage* 54(2):1755–1762
- Kulic D, Ott C, Lee D, Ishikawa J, Nakamura Y (2012) Incremental learning of full body motion primitives and their sequencing through human motion observation. *Int J Robot Res* 31(3):330345
- Kumar P, Gauba H, Roy PP, Dogra DP (2017) A multimodal framework for sensor based sign language recognition. *Neurocomputing* 259:21–38
- Lee D, Nakamura Y (2005) Mimesis from partial observations. In: IEEE/RSJ international conference on intelligent robots and systems (IROS), pp 1911–1916
- Lee D, Nakamura Y (2007) Motion capturing from monocular vision by statistical inference based on motion database: vector field approach. In: IEEE/RSJ

- international conference on intelligent robots and systems (IROS), pp 617–623
- Lee D, Nakamura Y (2014) Motion recognition and recovery from occluded monocular observations. *Robot Auton Syst* 62(6):818–832
- Lee D, Ott C, Nakamura Y (2009) Mimetic communication with impedance control for physical human-robot interaction. In: *IEEE international conference on robotics and automation (ICRA)*, pp 1535–1542
- Lee D, Soloperto R, Saveriano M (2017) Bidirectional invariant representation of rigid body motions and its application to gesture recognition and reproduction. *Auton Robots (AURO)* 42(1):125–145
- Li S, Lee D (2018) Point-to-pose voting based hand pose estimation using residual permutation. *Equivariant Layer*, arXiv:1812.02050, arXiv
- Li S, Woehlke J, Lee D (2018) Model-based hand pose estimation for generalized hand shape with spatial transformer network. In: *European conference on computer vision (ECCV)*, 4th international workshop on observing and understanding hands in action (HANDS)
- Liu Z, Lee D, Sepp W (2011) Particle filter based monocular human tracking with a 3D cardbox model and a novel deterministic resampling strategy. In: *IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pp 3626–3631
- Magnanimo V, Saveriano M, Rossi S, Lee D (2014) A Bayesian approach for task recognition and future human activity prediction. In: *IEEE international symposium on robot and human interactive communication (RO-MAN)*, pp 726–731
- Medina Hernández JR, Shelley M, Lee D, Takano W, Hirche S (2012) Towards interactive physical robotic assistance: parameterizing motion primitives through natural language. In: *IEEE international symposium on robot and human interactive communication (Ro-Man)*
- Mehrabian A (1996) Pleasure-arousal-dominance: a general framework for describing and measuring individual differences in temperament. *Curr Psychol* 14(4):261–292
- Minato T, Yoshikawa Y, Noda T, Ikemoto S, Ishiguro H, Asada M (2007) CB2: a child robot with biomimetic body for cognitive developmental robotics. In: *IEEE-RAS international conference on humanoid robots*
- Mitra S, Acharya T (2007) Gesture recognition: a survey. *IEEE Trans Syst Man Cybern Part C Appl Rev* 37(3):311–324
- Molchanov P, Yang X, Gupta S, Kim K, Tyree S, Kautz J (2016) Online detection and classification of dynamic hand gestures with recurrent 3D convolutional neural network. In: *IEEE conference on computer vision and pattern recognition*, pp 4207–4215
- Monkarezi H, Calvo RA, Hussain MS (2012) Automatic natural expression recognition using head movement and skin color features. In: *International working conference on advanced visual interfaces*, pp 657–660
- Montepare J, Goldstein S, Clausen A (1987) The identification of emotions from gait information. *J Nonverbal Behav* 11:33–42
- Nakagawa K, Shinozawa K, Ishiguro H, Akimoto T, Hagita N (2009) Motion modification method to control affective nuances for robots. In: *IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pp 5003–5008
- Neto P, Pereira D, Pires JN, Moreira AP (2013) Real-time and continuous hand gesture spotting: an approach based on artificial neural networks. In: *IEEE international conference on robotics and automation (ICRA)*, pp 178–183
- Nishio S, Ishiguro H, Hagita N (2007) Geminoid: tele-operated android of an existing person. In: *Humanoid robots: new developments*. I-Tech Education and Publishing, pp 343–352
- Ott C, Lee D, Nakamura Y (2008) Motion capture based human motion recognition and imitation by direct marker control. In: *IEEE international conference on humanoid robots*, pp 399–405
- Ott C, Baumgartner C, Mayr J, Fuchs M, Burger R, Lee D, Eiberger O, Albu-Schaffer A, Grebenstein M, Hirzinger G (2010) Development of a biped robot with torque controlled joints. In: *IEEE international conference on humanoid robots*, pp 167–173
- Pantic M, Rothkranz LJM (2000) Automatic analysis of facial expressions: the state of the art. *IEEE Trans Pattern Anal Mach Intell* 22(12):1424–1445
- Park H, Park J-I, Kim U-M, Woo W (2004) Emotion recognition from dance image sequences using contour approximation. In: *Joint IAPR international workshops structural, syntactic, and statistical pattern recognition*, pp 547–555
- Parusel S, Widmoser H, Golz S, Ende T, Blodow N, Saveriano M, Maldonado A, Kresse I, Weitschat R, Lee D, Beetz M, Albu-Schaeffer A, Haddadin S (2014) Human-robot interaction planning. In: *The 28th AAAI conference on artificial intelligence (AAAI)*
- Picard RW (1997) *Affective computing*. The MIT Press
- Plutchik R (2001) The nature of emotions human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *Am Sci* 89(4):344–350
- Riek LD, Paul PC, Robinson P (2010) When my robot smiles at me: enabling human-robot rapport via real-time head gesture mimicry. *J Multimodal User Interfaces* 3(1–2):99–108
- Roether C, Omlor L, Christensen A, Giese M (2009) Critical features for the perception of emotion from gait. *J Vis* 9(6):1–32
- Russell J, Mehrabian A (1977) Evidence for a three-factor theory of emotions. *J Res Pers* 11:273–294
- Saha S, Datta S, Konar A, Janarthanan R (2014) A study on emotion recognition from body gestures using kinect sensor. In: *IEEE international conference on communications and signal processing (ICCSP)*, pp 056–060
- Saveriano M, Lee D (2013) Invariant representation for user independent motion recognition. In: *IEEE international symposium on robot and human interactive communication (RO-MAN)*, pp 650–655

- Schmidts A, Lee D, Peer A (2011) Imitation learning of human grasping skills from motion and force data. In: IEEE/RSJ international conference on intelligent robots and systems (IROS), pp 1002–1007
- Soloperto R, Saveriano M, Lee D (2015) A bidirectional invariant representation of motion for gesture recognition and reproduction. In: IEEE international conference on robotics and automation (ICRA), pp 6146–6152
- Sun M, Kohli P, Shotton J (2012) Conditional regression forests for human pose estimation. In: Conference on computer vision and pattern recognition, pp 3394–3401
- Thomas F, Johnston O (1995) The illusion of life: Disney animation. Hyperion
- Toshev A, Szegedy C (2014) Deeppose: human pose estimation via deep neural networks. In: Conference on computer vision and pattern recognition
- Trovato G, Zecca M, Sessa S, Jamone L, Ham J, Hashimoto K, Takanishi A (2013) Cross-cultural study on human-robot greeting interaction: acceptance and discomfort by Egyptians and Japanese. *J Behav Robot* 4(2)
- Wada K, Shibata T, Musha T, Kimura S (2008) Robot therapy for elders affected by dementia. *IEEE Eng Med Biol Mag* 27(4):53–60
- Waldherr S, Romero R, Thrun S (2000) A gesture based interface for human-robot interaction. *Auton Robots* 9(2):151–173
- Yacoob Y, Davis L (1996) Recognizing human facial expressions from long image sequences using optical flow. *IEEE Trans Pattern Anal Mach Intell* 18(6):636–642
- Yuan S, Garcia-Hernando G, Stenger B, Kim T-K, Moon G, Chang JY, Lee KM, Molchanov P, Ge L, Yuan J, Chen X, Wang G, Yang F, Akiyama K, Wu Y, Wan Q, Madadi M, Escalera S, Li S, Lee D, Oikonomidis I, Argyros A (2018) 3D hand pose estimation: from current achievements to future goals. In: IEEE international conference on computer vision and pattern recognition (CVPR), pp 2636–2645
- Zecca M, Endo N, Momoki S, Itoh K, Takanishi A (2008) Design of the humanoid robot KOBIAN – preliminary analysis of facial and whole body emotion expression capabilities. In: IEEE-RAS international conference on humanoid robots (Humanoids), pp 487–492
- Zhang L, Sturm J, Cremers D, Lee D (2012) Real-time human motion tracking using multiple depth cameras. In: IEEE/RSJ international conference on intelligent robots and systems (IROS), pp 2389–2395