12 Automatic Analysis of Bodily Social Signals

Ronald Poppe

The human body plays an important role in face-to-face interactions (Knapp & Hall, 2010; McNeill, 1992). We use our bodies to regulate turns, to display attitudes and to signal attention (Scheflen, 1964). Unconsciously, the body also reflects our affective and mental states (Ekman & Friesen, 1969). There is a long history of research into the bodily behaviors that correlate with the social and affective state of a person, in particular in interaction with others (Argyle, 2010; Dittmann, 1987; Mehrabian, 1968). We will refer to these behaviors as *bodily social signals*. These social and affective cues can be detected and interpreted by observing the human body's posture and movement (Harrigan, 2008; Kleinsmith & Bianchi-Berthouze, 2013). Automatic observation and analysis has applications such as the detection of driver fatigue and deception, the analysis of interest and mood in interactions with robot companions, and in the interpretation of higher-level phenomena such as mimicry and turn-taking.

In this chapter, we will discuss various bodily social signals, and how to analyze and recognize them automatically. Human motion can be studied on many levels, from the physical level involving muscles and joints, to the level of interpreting a person's full-body actions and intentions (Poppe, 2007, 2010; Jiang et al., 2013). We will focus on automatically analyzing movements with a relatively short time scale, such as a gesture or posture shift. In the first section, we will discuss the different ways of measurement and coding, both from motion capture data and images and video. The recorded data can subsequently be interpreted in terms of social signals. In the second section, we address the automatic recognition of several bodily social signals. We will conclude the chapter with a discussion of challenges and directions of future work.

Measurement of Body Motion

Body movement can be observed and described quantitatively, for example, in terms of joint rotations or qualitatively with movement labels. While social signals are typically detected and identified as belonging to a certain category, body motion is typically described quantitatively. Therefore, the detection of bodily social signals is often based on a quantitative representation of the movement. From the perspective of computation, body motion is most conveniently recorded and measured using motion capture (mocap) devices. However, their obtrusive nature, cost, and the fact that they typically cannot be used outside the laboratory has limited their employment. Therefore, many researchers have turned to common, unobtrusive cameras for action recognition. Recently, the availability of cheap depth cameras provides opportunities as well. Bodily social signals can be detected directly from videos and depth sequences or, indirectly, from recovered body poses and movement.

We first discuss the manual and automatic measurement and common ways to represent human body movement. Next, we summarize the recording of motion capture, video and depth images, and the processing needed to transform raw outputs into body movement descriptions.

Manual and Automatic Measurement

The systematic analysis of body movement dates back to the early photography experiments of Marey and Muybridge (see Klette & Tee, 2007 for a historical background). By analyzing successive photos, they were able to analyze patterns of movement. Later, the introduction of video recording and play-back equipment allowed researchers to analyze behavior on a finer time scale (Condon & Ogston, 1966; Eisler, Hersen, & Agras, 1973). Initially, such analyses were used to investigate patients with mental diseases, but these methods soon found their way to the more general study into (communicative) nonverbal behavior.

Together with the increasing sophistication of recording and play-back devices, the opportunities for analysis developed. From videos, researchers coded specific behaviors that they were interested in. Evaluative coding relies on researchers that code their recorded material for the occurrence of particular forms of nonverbal behavior (Rozensky & Honor, 1982). These specific qualitative schemes have led to models of turn-taking (Sacks, Schegloff, & Jefferson, 1974) and gesturing (Lausberg & Sloetjes, 2009), amongst others. While it has been found that many bodily behaviors can be coded reliably (Baesler & Burgoon, 1987), evaluative schemes require interpretation of the observed behavior. This is especially true for bodily social signals. The variation in the performance of nonverbal behavior in magnitude, form, and direction requires that boundaries on the labels are set, which is an arbitrary task (Scherer & Ekman, 2008).

To address this issue, researchers have been looking at ways to describe human motion quantitatively. They developed schemes including the Bernese system for time series notation (Frey & von Cranach, 1973) and the Laban movement analysis (von Laban, 1975), which evolved into Labanotion (Hutchinson Guest, 2005). These systems describe body part positions and motion in terms of angles and velocities (Bente, 1989; Hirsbrunner, Frey, & Crawford, 1987) and have been found to be generally applicable and sufficiently detailed to animate computer characters (Bente et al., 2001). The recently introduced body and action posture (BAP; Dael, Mortillaro, & Scherer, 2012) coding system includes both quantitative aspects such as orientation and magnitude of body part movement, and functional descriptions, following Ekman & Friesen (1969). The system differentiates between posture units and action units, of which the latter are more subject to interpretation.

Both the qualitative and quantitative approaches have led to insights into bodily behavior. However, manually coding data is time consuming, meaning that there is often an inherent trade-off between the number of coded actions and the amount of coded material (Poppe et al., 2014). With the increasing availability of technology to record and analyze human motion, researchers have begun to address the automatic analysis of recorded data (Poppe, 2007, 2010). We will discuss advances in this direction.

Human Body Representation

Body movement can be described in terms of body mass displacement, muscle activations, or joint positions, to name a few. Describing the movement at the skeleton level is convenient, given that motion takes place at the joints (Poppe et al., 2014). The skeleton can be considered as a set of body parts (bones) connected by joints. Body poses can be represented as instantiations of joint positions. All joints and body parts in the human body together form a kinematic tree, a hierarchical model of interconnectivity. Typically, joints in the spine, hands, and feet are omitted. The joint at the top of the tree, usually the pelvis, forms a root to which all other joints are relative. When two joints are connected to a body part, the one higher in the tree hierarchy is considered the parent and the other the child. Movement in a parent joint affects the child joints. For example, movement of the left shoulder affects the position of the left elbow and wrist joints. Joint positions can be described globally with reference to a global axis system and origin. Alternatively, they can be described relative to their parent in the tree. Global and local representations each have their relative advantages. The former are most convenient when comparing full-body poses, as distances between pairs of joints can be calculated in a straightforward manner. When analyzing the movement of a single body part or joint, local representations enable the analysis of the motion in isolation.

The global or local positions of all joints form an adequate description of the body pose, especially when normalized for global position, orientation, and differences in body sizes (Poppe et al., 2014). Poses encode the positions of body parts, but do not reveal anything about their motion. To this end, the velocity of the joints can be used. Pose and motion information are often complementary and are both used in the analysis of bodily social signals.

Motion Capture

Motion capture technology employs either markers or wearable sensors to determine a subject's body pose. Marker-based mocap setups record the positions of markers attached to the body using many cameras. With proper calibration, these sensor positions can be translated to the positions of the joints. The advantage of such systems is their high accuracy. However, the space in which the movement can take place is limited, and marker occlusions, especially in the presence of other subjects, require additional post-processing. Inertial devices eliminate the need for visible markers as the sensors are worn on the body, possibly underneath clothing. This allows for their use in larger spaces and they perform more robustly when recording interactions between multiple subjects. Their acceleration measurements can be converted to 3-D positions of the joints. See (Poppe et al., 2014) for an overview and discussion of motion capture approaches. Both global and local joint positions can be obtained from mocap devices.

Video Recordings

The use of video for the study of nonverbal behavior is appealing as the recording is unobtrusive, both inside and outside the lab. In contrast to mocap devices, video cameras are cheap and widely available. Moreover, the abundance of available recordings portraying human behavior motivates the research efforts aimed at automatically analyzing them.

The analysis of human motion from video is challenging because of several factors. An image is a projection of a 3-D scene in which the depth information is lost. Moreover, determining which parts of the image represent the human figure is challenging, especially in the presence of background clutter and partial occlusion of the body. Nuisance factors such as variations in lighting, clothing, body sizes, and viewpoint add further to the challenge (Poppe, 2010).

In general, there are two main approaches to analyzing human movement from video. First, a body movement representation in terms of joint positions can be extracted, as described in the section on human body representation. Second, the characteristics of the image or movement in the image can be used directly for analysis. The results of these two approaches are pose-based and feature-based representations, respectively. We will discuss them in the following sections.

Pose-based Representations

There is a large volume of published research on estimating human body poses from video. A comprehensive discussion appears in Poppe (2007). Here, we will outline the most common approaches: model-based and discriminative.

In the first approach, model-based human pose estimation algorithms match an articulated model of a human to an image frame in a video. The model consists of a kinematic structure (see the section on human body representation) and a function that projects the model to the image. The image projection function determines how a pose of the model appears in an image, for example, in terms of image edges, silhouette, or color. Given that a body pose is a particular joint parameter instantiation, pose estimation becomes the process of finding the parameters that result in the best match between the image and the model projection. This match is evaluated in terms of image feature distance, usually in an iterative manner. This process is computationally expensive, but allows for the evaluation of a large number of parameters of the pose as well as the shape of the person (Guan et al., 2009).

This estimation process can be top-down, starting with the torso and working down the kinematic chain until the pose of the limbs is found. Deutscher, & Reid (2005) match the edges and silhouette information of a model with cylindrical body parts to those extracted from an image. They gradually reduce the amount of change in the pose to arrive at the final body pose estimate. Usually, the refinement of the pose is guided by a priori information on how humans move, including typical poses (Vondrak, Sigal, & Jenkins, 2013).

Alternatively, the process of estimating body poses can be bottom-up by first detecting potential body part locations in the image. Detectors are templates of a body part, often encoded as edge representations with additional cues such as color and motion (Eichner et al., 2012). In recent years, deformable part models have become popular due to their ability to simultaneously detect different parts of the body and reason which body poses are physically feasible and plausible (Felzenszwalb et al., 2010). Their output is a set of 2-D joint positions, which can be lifted to 3-D when sufficient assumptions about the observed motion have been made.

The second approach is the discriminative approach. Rather than iteratively fitting a human model to the data, one can learn a mapping from image to body poses from training data. Such a mapping can be implemented by regression models (Bo & Sminchisescu, 2010). Typically, training data consists of image features and an associated description of pose and viewpoint. Body poses can be recovered from test videos by first extracting image features and then applying the mapping. These discriminative, or learning-based, approaches are computationally much faster than model-based algorithms but can only reliably recover body poses if there is training data available with similar poses and viewpoints. This requires a lot of training data to sufficiently cover the range of poses. Given the large number of possible body poses, this has typically led researchers to concentrate their training data on common activities, although more recent approaches have targeted less constrained motion domains (Shotton et al., 2011).

Feature-based Representations

In contrast to pose-based representations, feature-based representation are less semantically meaningful but can be extracted efficiently from video images. Comparing an image of a scene with people to an image of the same scene without people will reveal one or a number of regions of differences that correspond to the locations of the people. The locations, sizes and movements of these regions are informative of their positions in the scene and can be used to investigate proximity and interaction patterns of small groups, such as from top-down views (Veenstra & Hung, 2011).

By analyzing differences between subsequent frames, one can analyze motion at a finer scale. While such differences can be the basis for the estimation of the locations of body parts (Fragkiadaki, Hu, & Shi, 2013), they can also be used directly. For example, the amount of movement, the direction of the movement, or the relative location of the movement (upper-body or lower-body) can be informative of the social signals that a person produces. Moreover, when looking at the movement of several people simultaneously, one can analyze the degree of mimicry in their interaction (Paxton & Dale, 2013).

When analyzing bodily social signals, often there is a specific interest in the locations of the hands and face. This is especially true for the analysis of gestures. Estimating the 2-D or 3-D positions of the hands and head is often less complex than estimating a full-body pose, especially when relying on skin color detection. By detecting skin-colored pixels and grouping them into connected regions, one can recover the location of the hands and face.

Depth Images

Time-of-flight (Ganapathi et al., 2010) and structured light cameras such as Microsoft's Kinect (Shotton et al., 2011), can estimate the distance between the camera and points in the scene. The availability of cheap devices has sparked the interest to use them to observe and analyze human movement. Nuisance factors that occur when using videos, including cluttered backgrounds and variation in lighting, are significantly reduced and the additional availability of depth information aids in labeling body parts and their orientation.

Recognition of Bodily Social Signals

In this section, we will discuss the recognition of various bodily social signals from the representations described in the first section. Recognizing, or classifying, social signals is the process of assigning a (semantic) label to an observed sequence of bodily movement. In general, the detection (in time) and recognition of bodily social signals are challenging due to the variations in the temporal and spatial performance, both between and within subjects. Social signals can have different bodily manifestations. Reversely, one distinct bodily behavior can have different meanings. For example, raising a hand can be a greeting or a sign to take the floor. The context in which the behavior is performed is important to disambiguate between the different meanings. We will discuss this in the next section.

Both the detection and recognition of social signals from body movement representations are often implemented with machine learning techniques (Vinciarelli, Pantic, & Bourlard, 2009). Given training data, which is a collection of body movement instances with associated social signal labels, a mapping from the former to the latter is learned. This mapping can take many forms, including state-space models such as hidden Markov models (HMM), or discriminative classifiers, such as the support vector machine (SVM). To deal with challenges, such as the diversity of the observed behavior, the inherent ambiguity of the observed behavior, and the typically limited amount of available training data, many different variants of machine learning algorithms have been introduced. Other chapters address these techniques for the understanding of social signals. In this section, we will focus on the potential and challenges in recognizing certain social signals from body movement. We will subsequently discuss the interpretation of a person's position relative to others (see Proxemics) and the analysis of social signals from the body (see Kinesics).

Proxemics

The way people use the space around them in relation to others is referred to as proxemics. Hall (1966) defines four zones of interpersonal distance with different characteristics in how people interact in terms of the way of gesturing and positioning the body. Moreover, these zones correspond to the relation between the people, such as friend or stranger. For small groups, people have been found to arrange themselves in so-termed F-formations in which each person has equal, direct, and exclusive access to the others (Kendon, 1990). When analyzing groups of people, the notion of relative orientation and proximity have been found good cues to determine who is part of a subgroup (Groh et al., 2010) and to predict mutual interest (Veenstra & Hung, 2011). Most of the work on automatic analysis of proxemics has been carried out in social surveillance setting in which body movement representations typically are feature-based. The automatic analysis of proxemics has also been studied at a closer distance by Mead, Atrash, and Matarić (2013). They considered a range of body movement features, including (relative) body position and elements of the pose. We will discuss the analysis of full-body movement in social interaction in the next section.

Kinesics

Kinesics refers to the study of body poses and movements as a mode of communication (Birdwhistell, 1952). The research on the automatic analysis of kinesics has focused mainly on conversational settings, such as meetings, interviews, and other small group interactions. The body has been found to communicate attitudes toward others in the interaction (Ekman, 1965). Okwechime et al. (2011) have addressed the automatic recognition of interest in interaction partners by analyzing gross body motion. Body shifts can be easily detected from pose-based and feature-based body movement representations, have been found to be indicative of disagreement (Bousmalis, Mehu, & Pantic, 2013), and play a role in the turn-taking process (Scheflen, 1964), to name a few. Moreover, mimicry in gross body motion can be a sign of rapport. It can be analyzed from pose-based representations, from simple frame-differencing techniques (Paxton & Dale, 2013) or from the detected position of the face in the image (Park et al., 2013). Closer analysis of the body also allows for the analysis of respiration, which can be a sign of anxiety. Burba et al. (2012) estimate the rate using a depth camera. Laughing can be considered a more discrete bodily signal, and different types of laughter can be recognized from mocap data (Griffin et al., 2013).

The hands are particularly informative of a subject's social and affective state, given that hand movements are closely tied to a person's speech (McNeill, 1985). Gestures and their co-occurrence with speech have been studied in great detail (Ekman & Friesen, 1969). The amount of gesturing has been found indicative of a user's attitude and mental state (Bull, 1987). For example, fidgeting behaviors have been shown to correlate with an increased experience of distress (Scherer et al., 2013) and can be extracted robustly from mocap representations (Burba et al., 2012). Similarly, self-touching has been found to be a sign of self-confidence as well as anxiety (McNeill, 1992). Marcos-Ramiro et al.

(2013) analyze self-touching in conversations from body pose representations obtained from a depth camera.

Especially in conversational settings, the pose and movement of the head is indicative of the subject's attention and serves several functions in the turn-taking process (Heylen, 2006). The analysis of head pose over time from pose-representations is straightforward. When the camera view covers a larger area and the subjects in the view are smaller, head orientation estimation based on both the subject's pose and head detection can be used (Bazzani et al., 2013). This allows investigating the role of head movement in the process of group formation and the evolvement of small group interactions.

One line of research has focused on estimating a subject's affective state from full-body poses and movements. The relation between specific body part positions and movements has been analyzed, for example, by Wallbott (1998). Recently, the automatic analysis has been attempted from pose-based, mainly recorded with mocap equipment, and feature-based representations. The reader is referred to Kleinsmith, Bianchi-Berthouze, and Steed (2011) for an overview of research in this area.

Challenges and Opportunities

The research into automatic recognition of bodily social signals and the study of social, nonverbal behavior are not isolated but rather benefit from each other. A better understanding of how humans behave informs the design and implementation of better recognition algorithms and, in turn, these advances in the automatic recognition help to better understand human behavior.

Apart from their use in understanding the principles of human behavior, automatic analysis of human body motion will continue to provide opportunities for online applications. The analysis of body movement can be used to analyze the outcome of negotiations and debates, to help practice public speaking and as a quick way to automate border control surveillance, to name a few. While initial work along these lines has already begun, there are some challenges that need to be addressed.

Measurement

Mocap equipment allows for the accurate measurement of body motion, but not unobtrusively. As such, it is not suitable for many applications outside the lab. Advances in computer vision algorithms and the recent introduction of depth cameras allow for the measurement outside the lab without the need of markers or wearable sensors, but their accuracy and robustness is still limited.

Given that many of the systematics of human nonverbal behavior are expressed in qualitative terms, a challenge is faced in converting the quantitative body movement measurements to these human-understandable, qualitative terms. This would allow for the adoption of the large body of literature of bodily behavior. Velloso, Bulling, and Gellersen (2013), among others, address this challenge by automatically estimating BAP labels from mocap data. They demonstrate that this is not a straightforward task and future work should be aimed at investigating how such a mapping can be made.

Recognition

Researchers have begun to adopt machine learning techniques that take into account individual differences in the display of bodily signals and the inherent ambiguity of body movement. Learning such models typically requires large amounts of training data for which obtaining ground truth labels is time-consuming. Researchers should look for alternative ways to label their data, for example, using crowdsourcing, implicit tagging, semi-supervised approaches, or by considering correlations between modalities. Moreover, when evaluating recognition algorithms, the optionality and ambiguity of social signals should be taken into account. The detection in time is often not addressed, which effectively avoids issues with the rare occurrence of social signals, and the associated problem of the detection of false positives. Future work should address the simultaneous detection and recognition of social signals from body movement data.

Context

Current work targets the recognition of specific bodily social signals in relative isolation. While the work in this direction progresses, there is an increasing need to understand the behavior more thoroughly. To this end, researchers should look beyond just the body and include other available knowledge, sometimes referred to as context. We distinguish here between the notion of other subjects, the specific task and setting, and cues from other modalities than the body movement.

Other subjects often provide a strong cue of the type of interaction that takes place. People respond to each other in more or less known patterns. Observing certain behavior in one person might aid in automatically understanding that of another person. For example, recognizing that one person sneezes helps in understanding why others turn their heads.

Many social signals are being studied in a restricted domain, such as a negotiation or tutoring setting. Knowledge of this setting helps in reducing the ambiguity in explaining the occurrence of a bodily behavior. When moving to less constrained application domains, it will be necessary to explicitly model the task and setting in order to perform such disambiguation.

We have discussed the analysis of social signals from the body, but there are often correlations between behavior of the body, the face, and voice. By taking a multimodal approach, the ambiguity in a single modality can be reduced and the recognition can accordingly be made more robust. Moreover, taking into account multiple modalities will help in addressing individual differences in the display of social signals across modalities (Romera-Paredes et al., 2013).

Conclusion

In this chapter, we have discussed the measurement and representation of human body motion. We have presented the current state of recognizing several bodily social signals. Finally, we have presented challenges in the automatic detection and recognition of bodily social signals and ways to address these. Given the advances, both in measurement technology and recognition algorithms, we foresee many interesting novel applications that consider social signals from the body. Moreover, the increasing robustness of current algorithms will allow for a wider embedding of such algorithms in multimedia analysis, social surveillance, and in human–machine interfaces, including social robots.

Acknowledgment

This publication was supported by the Dutch national program COMMIT, and received funding from the EU FP7 projects TERESA and SSPNet.

References

Argyle, Michael (2010). *Bodily Communication* (2nd rev. edn). New York: Routledge.

Baesler, E. James & Burgoon, Judee K. (1987). Measurement and reliability of nonverbal behavior. *Journal of Nonverbal Behavior, 11*(4), 205–233.

Bazzani, Loris, Cristani, Marco, Tosato, Diego, et al. (2013). Social interactions by visual focus of attention in a three-dimensional environment. *Expert Systems, 30*(2), 115–127.

Bente, Gary (1989). Facilities for the graphical computer simulation of head and body movements. *Behavior Research Methods, Instruments, & Computers, 21*(4), 455–462.

Bente, Gary, Petersen, Anita, Krämer, Nicole C., & De Ruiter, Jan Peter (2001). Transcript-based computer animation of movement: Evaluating a new tool for nonverbal behavior research. *Behavior Research Methods, Instruments, & Computers, 33*(3), 303–310.

Birdwhistell, Ray L. (1952). *Introduction to Kinesics: An Annotation System for Analysis of Body Motion and Gesture*. Louisville, KY: University of Louisville.

Bo, Liefeng & Sminchisescu, Cristian (2010). Twin Gaussian processes for structured prediction. *International Journal of Computer Vision, 87*(1–2), 28–52.

Bousmalis, Konstantinos, Mehu, Marc, & Pantic, Maja. (2013). Towards the automatic detection of spontaneous agreement and disagreement based on nonverbal behaviour: A survey of related cues, databases, and tools. *Image and Vision Computing, 31*(2), 203–221.

Bull, Peter E. (1987). *Posture and Gesture*. Oxford: Pergamon Press.

Burba, Nathan, Bolas, Mark, Krum, David M., & Suma, Evan A. (2012). Unobtrusive measurement of subtle nonverbal behaviors with the Microsoft Kinect. In *Proceedings of IEEE Virtual Reality Short Papers and Posters* March 4–8, 2012, Costa Mesa, CA.

Condon, William S. & Ogston, William D. (1966). Sound film analysis of normal and pathological behavior patterns. *Journal of Nervous and Mental Disease, 143*(4), 338–347.

Dael, Nele, Mortillaro, Marcello, & Scherer, Klaus R. (2012). The body action and posture coding system (BAP): Development and reliability. *Journal of Nonverbal Behavior, 36*(2), 97–121.

Deutscher, Jonathan, & Reid, Ian (2005). Articulated body motion capture by stochastic search. *International Journal of Computer Vision, 61*(2), 185–205.

Dittmann, Allen T. (1987). The role of body movement in communication. In A. W. Siegman & S. Feldstein (Eds), *Nonverbal Behavior and Communication* (pp. 37–64). Hillsdale, NJ: Lawrence Erlbaum.

Eichner, M., Marin-Jimenez, M., Zisserman, A., & Ferrari, V. (2012). 2D articulated human pose estimation and retrieval in (almost) unconstrained still images. *International Journal of Computer Vision, 99*(2), 190–214.

Eisler, Richard M., Hersen, Michel, & Agras, W. Stewart (1973). Videotape: A method for the controlled observation of nonverbal interpersonal behavior. *Behavior Therapy, 4*(3), 420–425.

Ekman, Paul (1965). Communication through nonverbal behavior: A source of information about an interpersonal relationship. In S. S. Tomkins & C. E. Izard (Eds), *Affect, Cognition, and Personality* (pp. 390–442). New York: Springer.

Ekman, Paul & Friesen, Wallace V. (1969). The repertoire of nonverbal behavior: Categories, origins, usage and coding. *Semiotica, 1*(1), 49–98.

Felzenszwalb, Pedro F., Girshick, Ross B., McAllester, David, & Ramanan, Deva (2010). Object detection with discriminatively trained part-based models. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 32*(9), 1627–1645.

Fragkiadaki, Katerina, Hu, Han, & Shi, Jianbo (2013). Pose from flow and flow from pose. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 2059–2066).

Frey, Siegfried & von Cranach, Mario (1973). A method for the assessment of body movement variability. In M. von Cranach & I. Vine (Eds), *Social Communication and Movement* (pp. 389–418). New York: Academic Press.

Ganapathi, Varun, Plagemann, Christian, Koller, Daphne, & Thrun, Sebastian (2010). Real time motion capture using a single time-of-flight camera. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 755–762).

Griffin, Harry J., Aung, Min S. H., Romera-Paredes, Bernardino, et al. (2013). Laughter type recognition from whole body motion. In *Proceedings of the International Conference on Affective Computing and Intelligent Interaction (ACII)* (pp. 349–355).

Groh, Georg, Lehmann, Alexander, Reimers, Jonas, Friess, Marc Rene, & Schwarz, Loren (2010). Detecting social situations from interaction geometry. In *Proceedings of the International Conference on Social Computing (SocialCom)*. (pp. 1–8).

Guan, Peng, Weiss, Alexander, Bălan, Alexandru O., & Black, Michael J. (2009). Estimating human shape and pose from a single image. In *Proceedings of the International Conference On Computer Vision (ICCV)*.

Hall, Edward T. (1966). *The Hidden Dimension*. New York: Doubleday.

Harrigan, Jinni A. (2008). Proxemics, kinesics, and gaze. In J. A. Harrigan & R. Rosenthal (Eds), *New Handbook of Methods in Nonverbal Behavior Research* (pp. 137–198). Oxford: Oxford University Press.

Heylen, Dirk (2006). Head gestures, gaze and the principles of conversational structure. *International Journal of Humanoid Robotics, 3*(3), 241–267.

Hirsbrunner, Hans-Peter, Frey, Siegfried, & Crawford, Robert (1987). Movement in human interaction: Description, parameter formation, and analysis. In A. W. Siegman & S. Feldstein (Eds), *Nonverbal Behavior and Communication* (pp. 99–140). Hillsdale, NJ: Lawrence Erlbaum.

Hutchinson Guest, Ann (2005). *Labanotation: The System of Analyzing and Recording Movement* (4th edn). New York: Routledge.

Jiang, Yu-Gang, Bhattacharya, Subhabrata, Chang, Shih-Fu, & Shah, Mubarak (2013). High-level event recognition in unconstrained videos. *International Journal of Multimedia Information Retrieval, 2*(2), 73–101.

Kendon, Adam (1990). *Conducting Interaction: Patterns of Behavior in Focused Encounters*. Cambridge: Cambridge University Press.

Kleinsmith, Andrea, & Bianchi-Berthouze, Nadia (2013). Affective body expression perception and recognition: A survey. *IEEE Transactions on Affective Computing, 4*(1), 15–33.

Kleinsmith, Andrea, Bianchi-Berthouze, Nadia, & Steed, Anthony (2011). Automatic recognition of non-acted affective postures. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*: Cybernetics, *41*(4), 1027–1038.

Klette, Reinhard, & Tee, Garry (2007). Understanding human motion: A historic review. In B. Rosenhahn, R. Klette, & D. Metaxas (Eds), *Human Motion: Understanding, Modelling, Capture and Animation* (pp. 1–22). New York: Springer.

Knapp, Mark L., & Hall, Judith A. (2010). *Nonverbal Communication in Human Interaction* (7th edn). Andover, UK: Cengage Learning.

Lausberg, Hedda & Sloetjes, Han (2009). Coding gestural behavior with the NEUROGES-ELAN system. *Behavior Research Methods, 41*(3), 841–849.

Marcos-Ramiro, Alvaro, Pizarro-Perez, Daniel, Romera, Marta Marrón, Nguyen, Laurent, & Gatica-Perez, Daniel (2013). Body communicative cue extraction for conversational analysis. In *Proceedings of the International Conference on Automatic Face and Gesture Recognition (FG)* (pp. 1–8).

McNeill, David (1985). So you think gestures are nonverbal? *Psychological Review, 92*(3), 350–371.

McNeill, David (1992). *Hand and Mind: What Gestures Reveal About Thought*. Chicago: University of Chicago Press.

Mead, Ross, Atrash, Amin, & Matarić, Maja J. (2013). Automated proxemic feature extraction and behavior recognition: Applications in human-robot interaction. *International Journal of Social Robotics, 5*(3), 367–378.

Mehrabian, Albert (1968). Some referents and measures of nonverbal behavior. *Behavior Research Methods, 1*(6), 203–207.

Okwechime, Dumebi, Ong, Eng-Jon, Gilbert, Andrew, & Bowden, Richard (2011). Visualisation and prediction of conversation interest through mined social signals. Pages 951–956 of: *Proceedings of the International Conference on Automatic Face and Gesture Recognition (FG)*.

Park, Sunghyun, Scherer, Stefan, Gratch, Jonathan, Carnevale, Peter, & Morency, Louis-Philippe (2013). Mutual behaviors during dyadic negotiation: Automatic prediction of respondent reactions. In *Proceedings of the International Conference on Affective Computing and Intelligent Interaction (ACII)* (pp. 423–428).

Paxton, Alexandra, & Dale, Rick (2013). Frame-differencing methods for measuring bodily synchrony in conversation. *Behavior Research Methods, 45*(2), 329–343.

Poppe, Ronald (2007). Vision-based human motion analysis: An overview. *Computer Vision and Image Understanding, 108*(1–2), 4–18.

Poppe, Ronald (2010). A survey on vision-based human action recognition. *Image and Vision Computing, 28*(6), 976–990.

Poppe, Ronald, Van Der Zee, Sophie, Heylen, Dirk K. J., & Taylor, Paul J. (2014). AMAB: Automated measurement and analysis of body motion. *Behavior Research Methods, 46*(3), 625–633.

Romera-Paredes, Bernardino, Aung, Hane, Pontil, Massimiliano, et al. (2013). Transfer learning to account for idiosyncrasy in face and body expressions. In *Proceedings of the International Conference on Automatic Face and Gesture Recognition (FG)* (pp. 1–8).

Rozensky, Ronald H., & Honor, Laurie Feldman (1982). Notation systems for coding nonverbal behavior: A review. *Journal of Behavioral Assessment, 4*(2), 119–132.

Sacks, Harvey, Schegloff, Emanuel A., & Jefferson, Gail (1974). A simplest systematics for the organisation of turn-taking for conversation. *Language, 50*(4), 696–735.

Scheflen, Albert E. (1964). The significance of posture in communicational systems. *Psychiatry, 27*(4), 316–331.

Scherer, Klaus R., & Ekman, Paul (2008). Methodological issues in studying nonverbal behavior. In J. A. Harrigan & R. Rosenthal (Eds), *New Handbook of Methods in Nonverbal Behavior Research* (pp. 471–504). Oxford: Oxford University Press.

Scherer, Stefan, Stratou, Giota, Mahmoud, Marwa, et al. (2013). Automatic behavior descriptors for psychological disorder analysis. *Proceedings of the International Conference on Automatic Face and Gesture Recognition (FG)* (pp. 1–8).

Shotton, Jamie, Fitzgibbon, Andrew, Cook, Mat, et al. (2011). Real-time human pose recognition in parts from single depth images. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 1297–1304).

Veenstra, Arno, & Hung, Hayley (2011). Do they like me? Using video cues to predict desires during speed-dates. In *Proceedings of the International Conference on Computer Vision (ICCV) Workshops* (pp. 838–845).

Velloso, Eduardo, Bulling, Andreas, & Gellersen, Hans (2013). AutoBAP: Automatic coding of body action and posture units from wearable sensors. In *Proceedings of the International Conference on Affective Computing and Intelligent Interaction (ACII)* (pp. 135–140).

Vinciarelli, Alessandro, Pantic, Maja, & Bourlard, Hervé (2009). Social signal processing: Survey of an emerging domain. *Image and Vision Computing, 27*(12), 1743–1759.

Vondrak, Marek, Sigal, Leonid, & Jenkins, Odest Chadwicke (2013). Dynamical simulation priors for human motion tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 35*(1), 52–65.

von Laban, Rudolf (1975). *Laban's Principles of Dance and Movement Notation* (2nd edn). London: MacDonald and Evans.

Wallbott, Harald G. (1998). Bodily expression of emotion. *European Journal of Social Psychology, 28*(6), 879–896.