Automatic Emotion Recognition Based on Body Movement Analysis

A Survey

Haris Zacharatos ■ *University of Cyprus*

Christos Gatzoulis ■ *Bahrain Polytechnic*

Yiorgos L. Chrysanthou ■ *University of Cyprus*

Recent scientific findings indicate that emotions are essential in decision making, perception, and learning; hence, they influence the mechanisms of rational thinking. Emotional imbalance can impair decision making. According to Rosalind Picard, if we want computers to be genuinely intelligent and interact naturally with us, we must give them the ability to recognize, understand, and express emotions.[1] Nevertheless, despite the significance of user affect in computing, technologists have largely ignored emotion, often resulting in frustrating experiences for users.

Affective computing is the study and development of systems and devices that recognize, interpret, process, and simulate human affect. This research combines engineering and computer science with psychology, cognitive science, neuroscience, sociology, education, psychophysiology, value-centered design, ethics, and more. Detecting emotional information begins with active or passive sensors that capture data of the user's physical state or behavior without interpreting the input. The gathered data is analogous to the cues humans use to perceive emotions in others. For example, a video camera might capture facial expressions, body posture, and gestures, whereas a microphone captures speech. Other sensors detect emotional cues by measuring physiological data, such as skin temperature and galvanic resistance.

Researchers have successfully explored some of these modalities, and new modalities are arising. (For a look at some of them, see the sidebar.) One promising modality that's receiving researchers' attention is body posture and movement. Two recent surveys have covered in depth the issues regarding human perception, automatic recognition, and the generation of affect through body expressions.[2,3] This article provides an overview of some of these issues and raises others we consider important.

One of the most expressive modalities for humans is body posture and movement, which researchers have recently started exploiting to perform emotion recognition. Automatic emotion recognition is becoming increasingly important in areas such as healthcare and gaming. However, such recognition based on body movement analysis poses challenges.

Some Application Areas

Various application areas can benefit from emotion recognition based on body movement. Intelligent tutoring systems can use it to adapt the presentation style when a learner is bored, interested, or frustrated or to detect student motivation.

Emerging Emotion Recognition Modalities

Scientists have extensively researched emotion recognition using various modalities; the most popular ones are facial expressions and voice. Besides body expressions and posture (see the main article), the following modalities have been receiving attention and show potential.

Brain Activity

Over the last two decades, researchers in cognitive and affective neuroscience have focused on emotion recognition through brain–computer interfaces. These interfaces are based on users' neural activity; the most common technique is electroencephalography (EEG) signal analysis. Neuroscientists are developing brain-imaging techniques that help map the neural circuitry underlying emotions. The cost, time resolution, and complexity of setting up protocols usable in real-world activities limit application development. Nevertheless, researchers have developed signal processing and classification algorithms for EEG in the context of building brain–computer interfaces.[1] For a look at how such interfaces can be linked to body emotion recognition, see the section "Brain and Body Emotion Recognition" in the main article.

Physiological Signals

You can also detect a user's emotional state by monitoring and analyzing physiological signals such as the pulse and heart rate, blood volume pulse (BVP), galvanic skin response (GSR), and minute contractions of the facial muscles. Photoplethysmography measures BVP; it produces a graph indicating blood flow through the extremities. GSR indicates skin conductivity, which depends on the skin's moisture level. Because the sweat glands produce this moisture and the nervous system controls the glands, a correlation exists between GSR and the body's arousal state. The more aroused a subject is, the greater the skin conductivity and GSR reading.[2] Facial electromyography (EMG) measures the facial muscles' electrical activity by amplifying the tiny electrical impulses that muscle fibers generate when they contract.

Muscle Activity

You can also detect a user's emotional state by analyzing muscle patterns. To perform muscle EMG in a nonlaboratory situation, wearable sensors are essential. The measurements are categorized into emotional states. The Body Action Coding System for muscle activation is investigating the connection between muscles and emotional behavior.[3] This can be an interesting research tool for psychological or neurological disorders such as autism or anxiety or for developmental disorders.[4,5]

Touch Behavior

With the increasing number of individuals using touchscreens, touch behavior can serve to detect the user's emotional state. For example, you can measure the fingers' movement on a mobile device, such as scrolling or stroking patterns.[6] So, touch behavior can serve as an evaluation indicator for application designers or can be used in real time to detect the user's emotional state as he or she scrolls down or plays a game.

References

1. A. Bashashati et al., "A Survey of Signal Processing Algorithms in Brain–Computer Interfaces Based on Electrical Brain Signals," *J. Neural Eng.*, vol. 4, no. 2, 2007, pp. R32–R57.
2. R. Picard, *Affective Computing*, MIT Press, 1998.
3. E. Huis in 't Veld, V. Boxtel, and B. Gelder, "The Body Action Coding System I: Muscle Activations during the Perception and Expression of Emotion," *Social Neuroscience*, vol. 9, no. 3, 2014, pp. 249–264.
4. M. Karg et al., "Body Movements for Affective Expression: A Survey of Automatic Recognition and Generation," *IEEE Trans. Affective Computing*, vol. 4, no. 4, 2013, pp. 341–359.
5. A. Kleinsmith and N. Bianchi-Berthouze, "Affective Body Expression Perception and Recognition: A Survey," *IEEE Trans. Affective Computing*, vol. 4, no. 1, 2012, pp. 15–33.
6. Y. Gao, N. Bianchi-Berthouze, and H. Meng, "What Does Touch Tell Us about Emotions in Touchscreen-Based Gameplay?," *ACM Trans. Computer-Human Interaction*, vol. 19, no. 4, 2012, article 31.

To do this, the system monitors the user's upper body poses, estimating the level of interest and engagement on the basis of the theory of mind. This theory emphasizes the principles and techniques humans deploy to understand, predict, and manipulate other humans' behavior.

In games, players' affective state is a significant factor in their motivation and engagement. Often, players lose interest and stop playing owing to negative emotions such as frustration or anger. In contrast, players who experience positive emotions when playing will more likely continue playing. As natural-motion interfaces such as Microsoft's Kinect become popular, a system that recognizes player emotions from body movements can be useful for game designers. It lets them employ AI behaviors in response to the user's state.

In social robotics, the main objective is natural interaction between humans and robots. Robots can incorporate a camera and motion-tracking sensors that detect emotion from body movements. Social robots are important in healthcare, with interactive robots that interact socially with humans and noninteractive robots for surgery, rehabilitation, and medication delivery.

Medical applications also use emotion recognition based on body movement to detect the depression levels of patients with dementia, schizophrenia, and autism.

Models of Emotion

There are two general approaches to representing emotions. The first represents them as distinct (happiness, sadness, fear, anger, surprise, disgust, and so on). The second approach measures and contextualizes them according to some dimensional space.

As an example of the latter approach, Figure 1 represents emotions in the dimensions of *valence* and *arousal*. Each emotion is a point in the space defined by these dimensions. (According to Beatrice Gelder, the body not only provides information about valence and arousal but also directs attention to a person's actions by focusing on body expressions.[5] Researchers found that by using unsupervised modeling, they can derive someone's action tendency from motion capture data.[6])

Similarly, you can use the dimensions of *activation* and *evaluation*. Activation is a person's tendency to execute an action according to his or her emotions; evaluation reflects the global appraisal of the positive or negative feeling.

You can also view emotions on the basis of psychological models of nonverbal communication such as the *Pleasure-Arousal-Dominant* (PAD) model, which has been used to research head and body movements.[7] In the PAD model, the pleasure–displeasure scale measures how pleasant an emotion is, the arousal–nonarousal scale measures the emotion's intensity, and the dominance–submissiveness scale represents the emotion's controlling and dominant nature.[8]

Healthcare professionals often look for distinct emotions to indicate depression or concentration. However, by using dimensional models, you can combine distinct emotions, resulting in higher recognition rates. Dimensional approaches can be used in applications such as games, in which a dimension of valence or arousal is enough to differentiate excitement from boredom.

An alternative representation of emotions is based on *appraisal theory*. This theory posits that emotions derive from our evaluations (appraisals) of events that cause specific reactions in different people. This theory offers a more flexible framework than discrete and dimensional models, accounting for differences among individuals and different responses to the same stimulus by the same individual at different times. The most frequently implemented model of appraisal is the *OCC model* (OCC stands for Ortony, Clore, and Collins). It states that a given emotion's strength depends primarily on the events, agents, or objects in the environment of the agent exhibiting the emotion.[9]

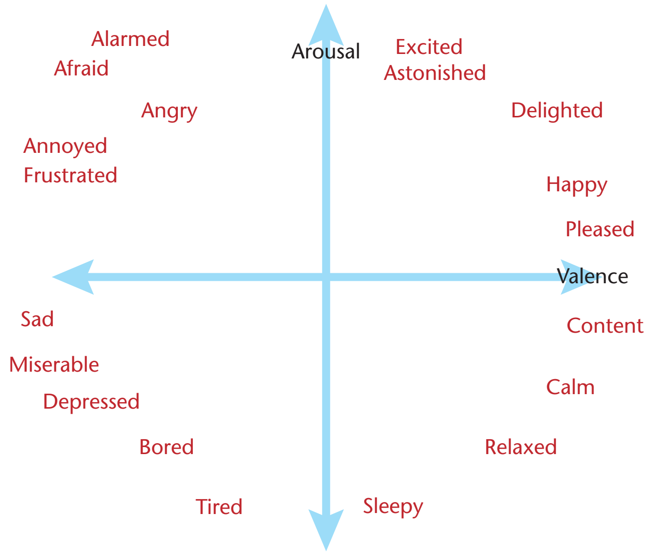


Figure 1. The valence–arousal space.[4] Each emotion is a point in the space defined by these dimensions.

Emotion Recognition from Body Movement

Darwin first described the association between body language and posture and emotions in humans and animals. Posture can offer information that's unavailable from the face or speech. For example, you might be able to recognize someone's emotional state from a distance by analyzing only his or her posture. Here, we describe the processes required to develop an automatic emotion recognition system.

Capturing Body Movement

Motion capture can accurately record body movements; it requires a controlled environment with cameras and sensors (see Figure 2). Systems based only on video cameras measure body movement less intrusively (for example, through silhouette extraction) but can't always accurately describe the movement of all body parts. Depth cameras such as the Microsoft Kinect can compute more complex features, but they're still not as accurate as traditional motion capture devices.

To obtain precise information on motor movement, you can apply physiological instruments to the body. Such instruments employ wearable sensing technology that measures acceleration, the body's six degrees of freedom, rotational information, and muscle electrical activity.[10] They're particularly useful outdoors, where other techniques have limitations. Recently, researchers have used

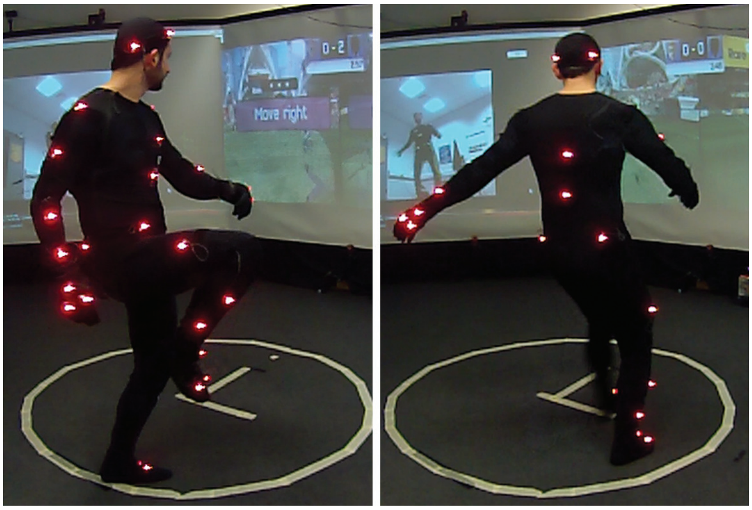


Figure 2. Images of a motion capture session. Motion capture requires a controlled environment with cameras and sensors.

Table 1. The motion factors of the effort component in Laban Movement Analysis.

Motion factor	Cognitive processes	Extremes
Space	Attention, thinking	Indirect–Direct
Weight	Intention, sensing	Light–Strong
Time	Decision, intuiting	Sustained–Sudden
Flow	Progression, feeling	Free–Bound

Arduino-based gyroscopes and electromyography (which measures muscle electrical activity) to measure emotions, with lower cost than commercial solutions.

Body Expression Notation Systems

Notation systems have been devised for observing, describing, notating, and interpreting human motion. Choreographers have used such systems since the 17th century. Some of these systems can describe only foot positions and symbols for each step denoting an aspect such as rhythm. Others can describe how dynamic body postures and gestures convey emotions, attitude, and personality. These latter systems might contain information about the body and limp positions of the extremities, orientation of the different body limbs expressed in spherical coordinates, and spatiotemporal coding that describes position and movement on the three Cartesian axes (sagittal, vertical, and transverse).

In psychology, coding systems describe various body positions and actions, orientation of postures, upper-body postures, and hand actions to check on attitudes and emotions such as boredom or interest and agreement or disagreement. They also correlate different parts of the body with emotion and movement qualities such as activity, expansiveness, and dynamics.

Theories of nonverbal behavior suggest that people use body language in everyday communication and that body language can be studied similarly to language. For example, in kinesics, a *kineme* is the smallest meaningful set of body movements, such as raising the eyebrows or looking upward. Using kinesic theory, researchers have developed a complex system containing kinegraphs for annotation that can be used to research body language.[11]

A widely used movement notation system is *Laban Movement Analysis* (LMA),[12] which dance artist and theorist Rudolf Laban proposed in the early 20th century. LMA focuses on the relationships between internal state, intention, and attention and their effects on all human motions. LMA can describe the expressive content of movements, making it appropriate for emotion and behavior analysis.

One component of LMA is *effort*. Laban saw effort as the inner impulse—a movement sensation, a thought, a feeling, or an emotion from which movement originates. This impulse constitutes the interface between movement's mental and physical components. Every human movement can potentially engage the effort component's four motion factors: space, weight, time, and flow. Table 1 shows these factors, their associated cognitive processes, and the two extremes for each factor.

Scientists have studied how nonacted (real) and acted emotions are grounded in low-level features, leading to new, context-free annotation approaches.[2,13] By using low-level features and unsupervised statistical methods, researchers have showed emotional clusters that reflect observers' labeling.[14]

Another notation system suited for coding nonverbal emotion expressions is *Body Action and Posture* (BAP).[15] BAP describes body movement on anatomical, form, and functional levels and separates 12 categories of behavior variables.

For a list of research using datasets of nonacted emotions, see Table V in the supplement to this article at http://doi.ieeecomputersociety.org/10.1109/MCG.2014.106. For a list and brief description of notation systems, see Table II in the article supplement.

Establishing the Ground Truth

A major issue with emotion recognition is its subjectivity and the difficulty in establishing a benchmark for successful recognition. Most early automatic recognition systems relied on corpora that had been acted. Recent studies are using nonacted natural data with body posture and movements, validating the results by using human observers. For any result to be valid, it should be compared with a database of valid postures or movements set as the ground truth. This ground truth is critical because even self-assessment can be manipulated.

Emotional expressions' believability and authenticity usually increase if all modalities express the same state. Incongruence correlates with lying. In this case, expressions of body movements seem more reliable than facial expressions because facial expressions can be controlled and hence faked.

Researchers have suggested various ways to compute the ground truth and deal with the difficulty of doing this in naturalistic situations. *Preference learning* employs machine learning; it exploits a set of specific features of an individual to predict his or her preferences.[16] Given a set of expressed preferences and statistical features, preference learning approximates the function between selected feature subsets and those preferences. If that function is accurate enough, you can use it to define the required user models for the ground truth.

Multiscore learning associates an expression with multiple values representing the scores of each available emotion label.[17]

Lately, researchers have used *crowdsourcing* for expert and nonexpert annotations. Recruiting annotators through Internet services (for example, Amazon Mechanic Turk) is an appealing option that allows the bulk outsourcing of multiple labeling tasks, typically with low overall costs and fast completion rates.[18]

However, annotations aren't always 100 percent accurate and consistent between observers. Techniques such as cross-validation can assess how statistical-analysis results will generalize to an independent dataset. Cross-validation partitions a data sample into two subsets, analyzing one subset (the training set) and validating the analysis on the other subset (the testing set).[19] Scientists use various techniques to validate agreement among annotators (for a list of techniques, see Table IV in the article supplement).

Segmentation

Humans use visual cues to segment motion sequences into gestures that serve as the building blocks of complex motions. Automatic body emotion recognition follows the same principle but employs geometric features (for example, acceleration, angles, and velocity) to segment a motion or posture. Segmentation enables the identification of each motion unit and its representation through a set of values of relevant parameters.

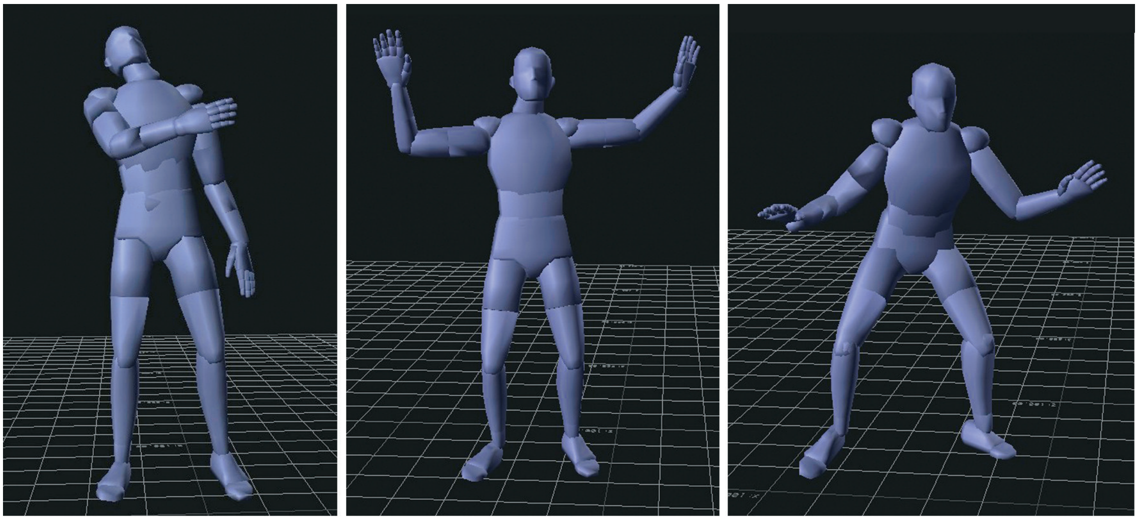
To efficiently create quality motion capture data, recorded sessions typically produce long data streams. Preprocessing breaks the data stream into short segments (motion clips) that are appropriate for an analysis tool. Most studies use manual segmentation in which human observers classify expressive postures and movements.[19] Manual segmentation is laborious and time-consuming. Moreover, it's subjective and depends on the human perception of movement and on the person's cultural background. Segmentation of specific motion data by experts instead of nonexperts, such as LMA experts in dance, gives a better quality of clips.

Daniel Weinland and his colleagues described ways to create automated segmentation tools.[20] An automatic program will produce the same segmentation given the same input data. On the other hand, different people will produce different segmentations from the same motion capture data.

Researchers have proposed two main approaches for automatic motion segmentation. In the direct approach, temporal segmentation (for example, using velocity, acceleration, and trajectory curvature) precedes recognition.[21] The other approach uses motion parameters (for example, human body activity) to identify candidate gesture boundaries.[22] Although both approaches indicate gesture boundaries accurately, they're successful only if a nongesture interval precedes and follows each gesture, a requirement not satisfied for continuous gesturing.

Various indirect methods detect gesture boundaries by finding end points and comparing the recognition likelihood score to a threshold. These kinematic methods are extremely efficient but produce simple low-level segmentations. Other methods use the projection error, resulting from principal component analysis (PCA), or use tracking changes on data that fit to a Gaussian distribution model.[23] These time-based methods produce a higher-level segmentation than the kinematic methods, but they don't use the semantic content of motion.

In addition, some methods use naive Bayes[22] to find segmentation profiles for dance motion capture sequences or employ implicit segmentation using hidden Markov models.[24] Both types of methods have shown that supervised learning can identify the complexity of decision making in much the way that the manual approach does. Implementing supervised learning for general motion is difficult, owing to the enormous amount of training data needed to create a true general classifier. A solution could be a classifier that works



(a) (b) (c)

Figure 3. Emotion recognition from whole-body movements. (a) Frustration. (b) Excitement. (c) Concentration. This approach employs techniques from Laban Movement Analysis.

well on smaller set of classes and general motions.

As we described earlier, you can use LMA's effort component to represent a classifier for emotion segmentation.

Another way to segment expressive motion automatically uses the *quantity of motion*—the amount of detected movement and its evolution over time. Researchers have depicted this as a sequence of bell-shaped curves to perform segmentation between pause and motion phases. To segment a motion, this method computes a list of curves and their features, using empirical thresholds.

A different method uses the concept of *motion energy* to segment motion primitives for emotion recognition. It computes the body's motion energy as a weighted sum of the rotational limb speeds, which will be large during periods of energetic motion and small during periods of low motion energy.[25]

Table III in the article supplement lists studies on automatic segmentation.

Automatic Emotion Recognition

Research on automatic emotion recognition using human-motion data has been based on training the system with low-level feature data (such as a given joint's rotation on a given axis).[2] Some of this research has used[26,27]

- posture-related features (end effector positions, end effector orientations, bounding boxes, and body directionality for the spine and head),
- dynamic features (velocity, acceleration, jerk, and fluidity of movement), and
- frequency-based features, such as the output of a fast Fourier transform for each position trajectory of the end effectors.

Table I in the article supplement lists studies using low-level movement features for emotion recognition.

Balance and postural-control variables can be used for automatic recognition. Human balance is assessed by analyzing the center of gravity or the center of pressure displacements. The combination of low-level 3D postural features with high-level kinematic and geometrical features (body movement activity and power, symmetry, and bending) has achieved high recognition rates.

Automatic recognition employs notation systems for motion analysis of expressive movements. As we described earlier, motion analysis can employ LMA components. Most research using LMA in emotion recognition focuses on the expressive aspect of the effort component, which describes the movement's dynamic qualities and the performer's inner attitude toward using energy. By selecting a set of suitable features from the trajectories described by the hands, foot, and head, you can use the effort component to describe expressive movements.

LMA techniques dealing with hand and arm movements are used to define and extract low-level features such as the hands' direction. They also use whole-body movements to extract features for the body's extremities (arms, legs, and head), its inclination, and the area it covers to classify the emotional set—for example, frustration, excitement, and concentration (see Figure 3).

Nele Dael and her colleagues adopted BAP to examine the types and patterns of body movement from actors expressing 12 emotions.[15] Using PCA, they reduced BAP's 49 behavior variables to 16 and applied a two-step clustering algorithm that revealed natural groupings. This algorithm was

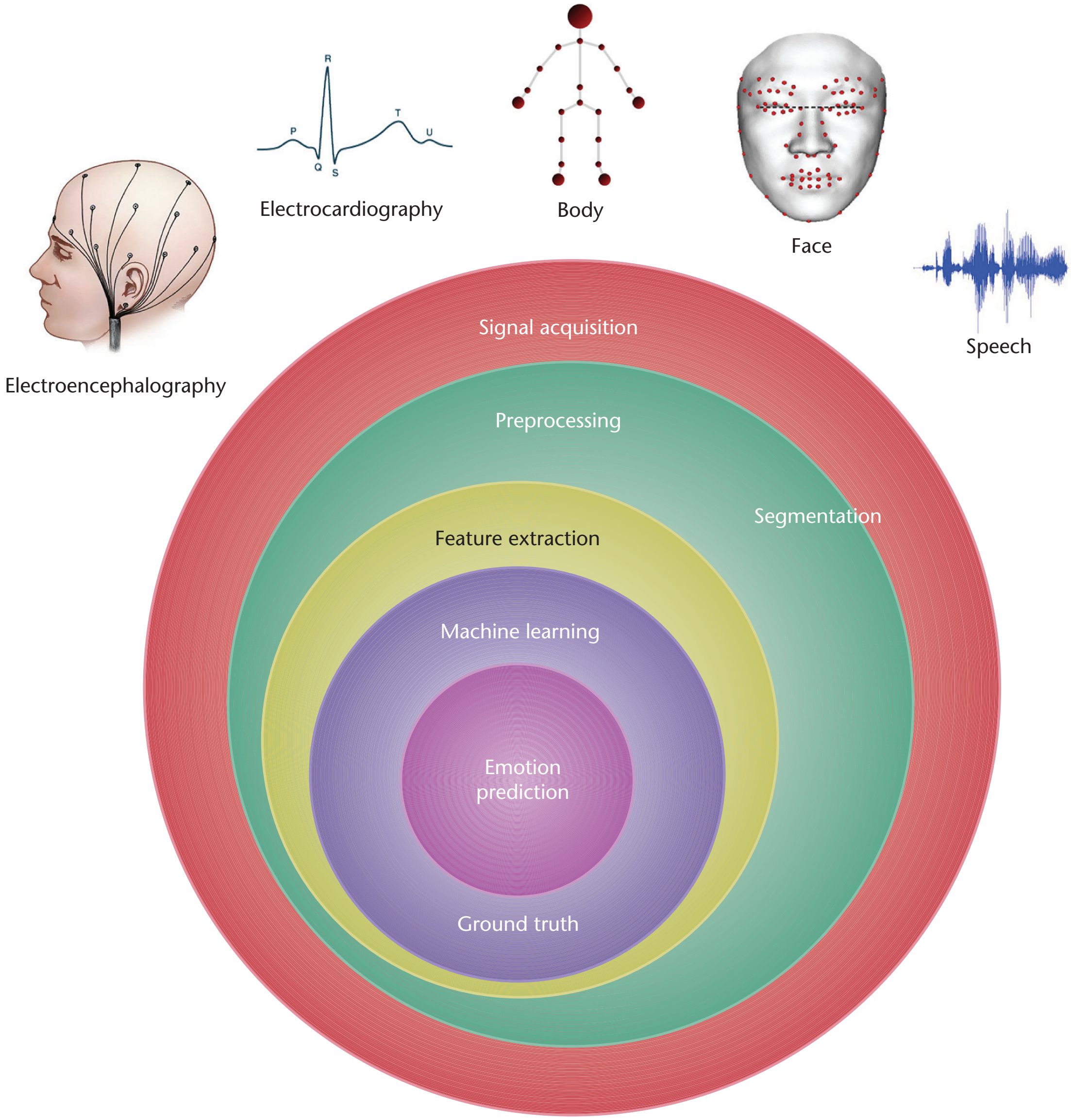


Figure 4. A multimodal automatic emotion recognition system. A multimodal approach produces higher recognition rates.

useful for large datasets with both categorical and continuous variables. Even though some emotions were characterized by a specific behavior pattern, most of these were encoded by a combination of clusters that resulted in considerable expressive variability and overlapping response profiles.

Multimodal Emotion Recognition

No one modality can represent the full range of emotional states; humans often use all their senses to understand other people's emotions. This motivates the development of multimodal emotion recognition systems. Most approaches for automatic recognition investigate the categorical (distinct) emotions—in particular, the basic emotions or a subset of them. Fewer studies refer to emotional dimensions. Nevertheless, recognition systems based on a dimensional model for emotions seem more appropriate for multimodal emotion recognition because modalities differ in their intensity and distinctness in expressing various emotions. Figure 4 illustrates the stages needed to build a multimodal emotion recognition system.

Several multimodal approaches that fuse facial, speech, and body features show that explicit detection of the temporal phases can improve affect recognition accuracy. They also show that recognition that fuses face and body features performs better than recognition using just the face or body. In addition, synchronized feature-level fusion performs better than decision-level fusion.

In cognitive and affective neuroscience, emotional body language is a rapidly emerging research area. Scientists have known about the amygdala's involvement in emotional behavior for some time. Also, biological movement patterns that people experience as pleasant activate subcortical structures including the amygdala. The visual perception of biological movement activates two areas in the occipital and fusiform cortex. This indicates that the areas known for processing faces also process larger properties associated with human bodies. High-field functional-magnetic-resonance-imaging techniques show that exposure to body expressions of fear, as opposed to neutral body postures, activates the fusiform gyrus and amygdala. These two areas' association with facial expressions suggests synergies between facial and body emotional expressions.

Research Directions

Here we present four different but interconnected directions for future research.

Recognition Using High-Level Movement Notation

Even though some studies in emotion recognition have used movement notation systems, the task's complexity allows further investigation. Most approaches have focused on specific aspects or subcomponents of a notation system. In many cases, results that use a specific context are promising but can't generalize to a broader one. A common problem is that the notation systems were created to describe dance movement, not emotion recognition.

Specifically, researchers have tested LMA's suitability for emotion recognition (as we described earlier). Although research has produced encouraging results regarding emotion recognition in presegmented clips in a specific context, the LMA computational models have several limitations in their experimental data. No evidence exists that recognition will be successful in other applications and contexts. It would be interesting to see how LMA can offer a more comprehensive, holistic system for detecting emotions in different contexts. Such an approach would require studying a set of emotions and their bodily manifestations across different contexts and then deriving correlations from those observations.

Perhaps one way to build more robust automatic recognition systems is to define a notation system for body movements, similar to the Facial Action Coding System, that fully defines the relation between action units and emotions. For example, an extension of BAP could lead to a notation system containing a one-to-one relationship between body action units and emotion. Such a system would need to be able to capture the full variety of naturalistic expressions.

Automatic Emotion Segmentation

A challenge in automatic emotion recognition is to segment continuous body motion signals so that recognition can occur using the segmented motion clips. Automating segmentation can yield significant benefits by reducing expensive human intervention and avoiding ambiguity due to the subjectivity of human perception and decision. A rational approach would be to establish automatic emotion recognition based on automatic gesture segmentation.

Researchers have employed techniques that use continuous signals of motion parameters and look for sudden changes to those parameters,[21] but these techniques detect only nongesture signals. Also, such techniques usually won't work for real applications. To address this problem, we need research to initially determine a real-time separation of gesture and nongesture signals and then test emotion recognition techniques that use immediate, short-term memories and continuous-frame windows.

Also, systems should be able to distinguish context periods within a signal. For example, in games, it would helpful if systems could distinguish between gameplay movements and non-gameplay movements. Each type of movement would require a different technique for emotion recognition. Emotion detection during gameplay could look into properties such as arousal and energy of movement for gestures. For non-gameplay, the recognition of more standard gesture properties could provide insights.

Another possible extension of existing research would be to compare the success of high-level and low-level segmentation in different emotion recognition techniques.[23] For instance, LMA could be tested on both high- and low-level segmentation owing to the correlations found between LMA qualities and kinematic features.

Moreover, it would be worth examining the use of fuzzy logic to determine the membership of emotions for each clip. Such membership can be used both for emotion recognition and to study the richness of motion with the observations and annotations of experts in expression. Furthermore, unsupervised learning could detect outliers in frame windows to distinguish context from non-context signal segments or expressive from non-expressive clips.

Supervised techniques can also provide solutions to this problem. With a large collection of datasets of human ratings, you could create a predictive model to select candid motion clips for two different approaches. The first one would use the clips for automatic selection and extraction of expressive clips without having to determine a definite emotion for annotation. Observers could then annotate those clips, using the techniques we described earlier. The second approach would establish automatic emotion recognition using a set of different features, as in the emotion recognition literature.[25] The section "Automatic Emotion Recognition" provides more information on this topic.

Context Knowledge

Context knowledge describes information about the surrounding environment and, more specifically, interaction items such as objects or other humans or avatars. Analyzing this context when a movement is performed can lead to more accurate emotion recognition, something that's mostly missing from body-movement-driven computational models. Successful recognition of emotional states in human–computer interfaces is linked to individual human characteristics and behavior types. So, emotion recognition systems must take into account the interaction's context as well as the type of user.

Affective user models are representational structures that store information about the user's affective profile. This information can be important for

- designing adaptive interactive applications for personalized experiences and
- globalizing systems by transferring a user's model among different applications.

Adaption and learning can be enhanced by context-sensitive frameworks for emotion classification, and context can be defined as the emotional context of past and future observations.

Another way to define context-aware emotions is by using domain ontologies that can be personalized according to the user, language, and culture. Ontologies can use cognitive input to apply context and user profiles in emotion recognition. Such ontologies can be tested on other types of input such as body motion data and movement notation systems to allow emotion recognition based on context information. A recent study, of particular interest to this survey, incorporated the context of the bodily reaction and cognitive input to demonstrate how context influences how humans determine and interpret other humans' emotions.[28] Because humans are the standard benchmark for testing emotion recognition for nonacted data, this study's results show that such an ontology offers another step toward more realistic and successful emotion recognition based on body motion data.

Context is important during not only the actual recognition process but also the creation of the emotion information databases used to train systems. In that case, and with the use of nonacted data, the context might influence not only the emotion expression style but also the opinion of the person annotating a dataset with an emotion label. In such scenarios, the expressions are uncontrolled and annotation becomes subjective, with varying agreement levels being the major benchmark measurement. Nevertheless, emotion recognition experiments show that invalid annotation can lead to poor results. Moreover, expressions themselves might vary with context, producing poor results if you try to generalize by employing emotion captured in one context as training data for other contexts.

Successful recognition of emotional states in human–computer interfaces is linked to individual human characteristics and behavior types.

Ingo Siegert and his colleagues investigated how context information influences the assignment of labels with realistic data.[29] They used video and audio channels to present data to annotators. The results show that contextual information might be more important for specific emotions. However, the results were limited and didn't provide a definite set of rules. It would be interesting to determine a set of features or an adaptation of a movement notation that allows datasets that can generalize to a set of emotions in different contexts.

Brain and Body Emotion Recognition

Emotional body language consists of emotions expressed with the whole body. This comprises coordinated movements and meaningful actions that so far have been investigated in isolation and not in the context of the perception of emotional body language. As we described earlier, the fusiform gyrus and amygdala process properties related to emotional body language. The amygdala decodes the affective relevance of sensory inputs and initiates adaptive behaviors through its connections to the motor system. More recently, the discovery of neurons that encode complex movements and actions (mirror neurons) has provided the neurobiological basis for all emotional and social cognition skills. This has changed our view of the motor system's role in the perception of body movement and emotional body language.

Brain activity directly associated with exposure to emotional body language is relatively unexplored. Emotion researchers are trying to get closer to the link between emotion and behavior, to shed light on disorders that combine motor and emotional components, such as autism, schizophrenia, and Huntington's disease. To perceive bodily expressed emotions, humans rely on a mixture of visual form and motion cues. Attempts to emulate emotion recognition in machines will require detailed knowledge of not only how all the brain's subsystems operate but also how they interact, which is a current focus of cognitive-neuroscience research.

Another interesting use of brain technology on body-based emotion recognition would be to compare the success of an emotion recognition system trained from clips annotated by brain signals with a system trained using manual annotation. Moreover, because neuron headsets can provide input to robotic controllers and virtual simulations, it would be interesting to see how specific emotions trigger certain body motions and gestures and to determine correlations between the emotions and movements.

In online situations, emotion recognition based on body movement will require a continuous signal. So, automatic emotion recognition will be a necessity. This will require applying and configuring knowledge from motion segmentation to emotion recognition.

Also, emotion recognition rates for areas such as medical applications will need to be high because errors could have a serious negative impact. (In contrast, in computer games, recognition rates could vary.) Nevertheless, we should use emotion recognition systems even if they can't reach perfection for their intended application. Perhaps we'll be able to improve them by fusing them with other modalities such as facial expressions and speech.

References

1. R. Picard, *Affective Computing*, MIT Press, 1998.
2. A. Kleinsmith and N. Bianchi-Berthouze, "Affective Body Expression Perception and Recognition: A Survey," *IEEE Trans. Affective Computing*, vol. 4, no. 1, 2012, pp. 15–33.
3. M. Karg et al., "Body Movements for Affective Expression: A Survey of Automatic Recognition and Generation," *IEEE Trans. Affective Computing*, vol. 4, no. 4, 2013, pp. 341–359.
4. J.A. Russel, "A Circumplex Model of Affect," *J. Psychology and Social Psychology*, vol. 39, no. 6, 1980, pp. 1161–1178.
5. B. Gelder, "Why Bodies? Twelve Reasons for Including Body Expressions in Affective Neuroscience," *Philosophical Trans. Royal Soc.*, vol. 364, no. 3, 2009, pp. 3475–3484.
6. A. Kleinsmith, R. De Silva, and N. Bianchi-Berthouze, "Grounding Affective Dimensions into Postures Description," *Affective Computing and Intelligent Interaction*, LNCS 3784, Springer, 2005, pp. 263–270.
7. B. Lance and S. Marsella, "Emotionally Expressive Head and Body Movement during Gaze Shifts," *Proc. 7th Int'l Conf. Intelligent Virtual Agents* (IVA 07), 2007, pp. 72–85.
8. A. Mehrabian, "Pleasure-Arousal-Dominance: A General Framework for Describing and Measuring Individual Differences in Temperament," *Current Psychology*, vol. 14, no. 4, 1996, pp. 261–292.
9. A. Ortony, G. Clore, and A. Collins, *The Cognitive Structure of Emotions*, Cambridge Univ. Press, 1988.
10. M. Aung et al., "Automatic Recognition of Fear-Avoidance Behaviour in Chronic Pain Physical Rehabilitation," *Proc. 8th Int'l Conf. Pervasive Computing Technologies for Healthcare*, 2014.
11. R. Birdwhistell, "Introduction to Kinesics: An Annotation System for Analysis of Body Motion and Gesture," PhD diss., Univ. Kentucky, 1952.
12. R. Laban, *The Language of Movement: A Guidebook to Choreutics*, Plays Inc., 1974.
13. M. Coulson, "Attributing Emotion to Static Body Postures: Recognition Accuracy, Confusions, and Viewpoint Difference," *J. Nonverbal Behavior*, vol. 28, no. 2, 2004, pp. 117–139.
14. R. De Silva and N. Bianchi-Berthouze, "Modeling Human Affective Postures: An Information Theoretic Characterization of Posture Features," *Computer Animation and Virtual Worlds*, vol. 15, nos. 3–4, 2004, pp. 269–276.
15. N. Dael, M. Mortillaro, and K. Scherer, "The Body Action and Posture Coding System (BAP): Development and Reliability," *Nonverbal Behavior*, vol. 36, no. 2, 2012, pp. 97–121.
16. G. Yiannakakis, "Preference Learning for Affective Modeling," *Proc. 3rd Int'l Conf. Affective Computing and Intelligent Interaction*, 2009, pp. 1–6.
17. H. Meng, A. Kleinsmith, and N. Bianchi-Berthouze, "Multi-score Learning for Affect Recognition: The Case of Body Postures," *Affective Computing and*

Intelligent Interaction, LNCS 6974, Springer, 2011, pp. 225–234.
18. M. Soleymani and M. Larson, "Crowdsourcing for Affective Annotation of Video: Development of a Viewer-Reported Boredom Corpus," *Proc. ACM SIGIR Workshop Crowdsourcing for Search Evaluation*, 2010, pp. 4–8.
19. A. Kleinsmith, N. Bianchi-Berthouze, and N. Steed, "Automatic Recognition of Non-acted Affective Postures," *IEEE Trans. Systems, Man, and Cybernetics*, vol. 41, no. 4, 2011, pp. 1027–1038.
20. D. Weinland, R. Ronfard, and E. Boyer, "A Survey of Vision-Based Methods for Action Representation, Segmentation and Recognition," *Computer Vision and Image Understanding*, vol. 115, no. 2, 2011, pp. 224–241.
21. H. Kang, C. Lee, and K. Jung, "Recognition-Based Gesture Spotting in Video Games," *Pattern Recognition Letters*, vol. 25, no. 15, 2004, pp. 1701–1714.
22. K. Kahol, P. Tripathi, and S. Panchanathan, "Automated Gesture Segmentation from Dance Sequences," *Proc. 6th IEEE Int'l Conf. Automatic Face and Gesture Recognition*, 2004, pp. 883–888.
23. J. Barbic et al., "Segmenting Motion Capture Data into Distinct Behaviors," *Proc. 2004 Graphics Interface Conf.*, 2004, pp. 185–194.
24. T. Starner and A. Pentland, "Visual Recognition of American Sign Language Using Hidden Markov Models," master's thesis, Media Laboratory, MIT, 1995.
25. S. Pianna et al., "Real-Time Automatic Emotion Recognition from Body Gestures," Cornell Univ., 2014.
26. G. Castellano, S. Villalba, and A. Camurri, "Recognising Human Emotions from Body Movement and Gesture Dynamics," *Affective Computing and Intelligence Interaction*, LNCS 4738, 2007, pp. 71–82.
27. G. Cimen, H. Ilhan, and T. Capin, "Classification of Human Motion Based on Affective State Descriptors," *Computer Animation and Virtual Worlds*, vol. 24, nos. 3–4, 2013, pp. 355–363.
28. F. Berthelon and P. Sander, "Emotion Ontology for Context Awareness," *Proc. 4th IEEE Int'l Conf. Cognitive Infocommunications*, 2013, pp. 59–64.
29. I. Siegert, R. Bock, and A. Wendemuth, "The Influence of Context Knowledge for Multimodal Affective Annotation," *Human-Computer Interaction*, part V, LNCS 8008, 2013, pp. 381–390.

***Haris Zacharatos** is a PhD candidate at the University of Cyprus's Affective Computing & Graphics Lab. His research interests include affective computing and machine emotional intelligence, next-generation network architectures and mobile networks, wearable computing, unmanned aerial vehicles, and robotics. Zacharatos received an MSc in computer science from the University of Cyprus, with a specialization in emotion recognition. He's a member of the Association for the Advancement of Affective Computing. Contact him at haris@cellock.com.*

***Christos Gatzoulis** is a program manager for Bahrain Polytechnic's Bachelor of Information and Communications Technology program. His research focuses on artificial intelligence and technical animation for games. He received his PhD in artificial intelligence for computer animation and robotics from the University of Teesside. Contact him at christos.gatzoulis@polytechnic.bh.*

***Yiorgos L. Chrysanthou** is an associate professor in the Computer Science Department of the University of Cyprus, where he heads the Graphics and Hypermedia Lab. His research interest is 3D computer graphics, including real-time rendering, crowd simulation, character animation, reconstruction of urban environments, and affective computing. Chrysanthou received a PhD in computer science from Queen Mary and Westfield College. He's coauthor of* Computer Graphics and Virtual Environments: From Realism to Real-Time *(Addison-Wesley, 2001). Contact him at yiorgos@cs.ucy.ac.cy.*