

Research Article

Chapa Sirithunge*, Gayashan Porawagamage, Nikolas Dahn,
A. G. Buddhika P. Jayasekara, and D. P. Chandima

Recognition of arm and body postures as social cues for proactive HRI

<https://doi.org/10.1515/pjbr-2021-0030>

received May 24, 2021; accepted September 18, 2021

Abstract: Artificial agents can uplift the living standards of domestic population considerably. One hindrance for this is that the robot is less competent to perceive complex human behaviors. With such perceptive skills in the robot, nonexpert users will find it easier to cope with their robot companion with less and less instructions to follow. Perception of the internal state of a user or “user situation” before interaction is crucial in this regard. There are a variety of factors that affect this user situation. Out of these, posture becomes prominent in displaying the emotional state of a person. This article presents a novel approach to identify diverse human postures often encountered in domestic environments and how a robot could assess its user’s emotional state of mind before an interaction based on postures. Therefore, the robot evaluates posture and the overall postural behavior of its user throughout the period of observation before initiating an interaction with its user. Aforementioned user evaluation is non-verbal and decisions are made through observation as well. We introduced a variable called “valence” to measure how “relaxed” or “stressed” a user is, in a certain encounter. The robot decides upon an appropriate approach behavior accordingly. Furthermore, the proposed concept was capable of recognizing both arm and body postures and both

postural behaviors over time. This leads to an interaction initiated by robot itself in a favorable situation so that the scenario looks more intelligent. Hence more humanlike. The system has been implemented, and experiments have been conducted on an assistive robot placed in an artificially created domestic environment. Results of the experiments have been used to validate the proposed concept and critical observations are discussed.

Keywords: body language, posture recognition, postural behavior, human–robot interaction, social robots

1 Introduction

Service robots play an interesting co-operative role in making human lifestyle comfortable and efficient. As in many of the today’s applications, service robots are used in interaction scenarios that are often encountered in social environments in addition to service tasks. Conversations and small talk with the robot companion when the user is in relaxing situations can be stated as examples in this regard [1,2]. Therefore, the capability of service robots toward collaborative interaction other than performing a specialized task is a pleasing factor for its human user. Therefore, in addition to supporting daily tasks such as cooking, cleaning, and therapy [3–5], the robot is expected to help maintain a healthy mental condition in user by accompanying when the user seeks association nonverbally. In order to achieve a socially intelligent behavior, robots must be capable of perceiving user situation. It becomes humanlike if the robot could perceive nonverbal behaviors of users. User situation depends highly on the activity he is engaged in. As a user’s posture varies in accordance with his/her current activity, posture could be analyzed as the foundation. This is due to the fact that almost all the human postures change when the current task of that particular human changes. As a result, posture is task dependent and at the same time emotion dependent [6].

Identification of arm and body postures is widely used for applications where human activity recognition

* **Corresponding author: Chapa Sirithunge**, Department of Electrical Engineering, Intelligent Service Robotics Group, University of Moratuwa, Katubedda 10400, Sri Lanka, e-mail: ra-chapa@uom.lk

Gayashan Porawagamage: Department of Electrical Engineering, University of Moratuwa, Moratuwa, 10400 Sri Lanka, e-mail: gayashandasun@gmail.com

Nikolas Dahn: Robokind Foundation, Hannover 30167, Germany, e-mail: nikolas.dahn@robokind.de

A. G. Buddhika P. Jayasekara: Department of Electrical Engineering, Intelligent Service Robotics Group, University of Moratuwa, Moratuwa, 10400 Sri Lanka, e-mail: buddhikaj@uom.lk

D. P. Chandima: Department of Electrical Engineering, Intelligent Service Robotics Group, University of Moratuwa, Moratuwa, 10400 Sri Lanka, e-mail: chandimadp@uom.lk

is required [7]. This has a wide range of applications other than service applications in domestic environments, such as surveillance [8], ergonomics, and behavior monitoring of workers [9]. In addition, body language plays an important role in displaying one's inner state of mind to the outside world, without the help of words. Posture recognition considering body movements has become a challenging and exhaustive objective for computer vision. This becomes even harder when humans use awkward postures during an activity. Most existing methods of posture evaluation use complex mechanisms or are based on a lengthy stream of data that require computer intensive preprocessing. Due to this reason, deploying such methods becomes difficult when a quick assessment of a scenario has to be made. Existing real-time methods lack recognition accuracy when intermediate postures and postures with high self-occlusion are considered. Hence, these systems are difficult to be used for occasions where decision making on robot-initiated interaction initiation was necessary.

Because of the above reasons, simple methods for posture recognition and effective initiation of interaction are emerging requirements at present. As an effort toward this, we present an approach based on body geometry for arm and body posture recognition, which was then used to determine the user situation. Finally, appropriate behaviors such as approach behavior and verbal responses from a robot were generated for a robot-initiated human-robot interaction (HRI).

2 Related work

2.1 Recognition of body posture

Modern approaches for posture recognition often involve video stream analysis or tracking of body joint locations. Most of the literature focuses on motion capture of body parts during posture recognition. The other methods involve pose estimation using depth images. Recognition of postures through depth video streaming has also received recent attention. But such methods require a lot of pre-processing. Thus, this has become a problem to overcome in real time posture recognition applications.

In ref. [10], authors presented an approach using top view image of the human in order to avoid self-occlusion. Postures with similar top view, but slightly deviated from each other in the arrangement of body joints, are difficult to be identified in this manner. Furthermore, a top view camera is inapplicable for a moving robot in a domestic environment. Cucchiara *et al.* [11] used posture classification

with the help of multiple cameras based on projection histograms implemented using Hidden Markov Model (HMM). Postures that are slightly deviated from each other but fall under a common major posture group (e.g., standing straight, standing while torso bent forward, standing while leaning to a wall, etc.) are difficult to be differentiated during this approach.

Pisharady and Saerbeck [12] have implemented a posture recognition algorithm based on geometric features of the body. In some postures, independent position of body parts as seen by a coordinate frame outside the body is required to identify a posture with a satisfactory accuracy. Buccolieri *et al.* [13] have used a recognition technique with active contours for moving action detection with HMM in fall detection. This requires intersection of inputs from all the cameras. Both the methods mentioned above require a large number of features for accurate functioning. A similar approach is explained in ref. [14], and this method analyzes an ellipse drawn around the human shape. Deviations in the ellipse are used to identify a certain posture. Identification of intermediate postures with similar geometry in the human shape is difficult with these systems. Visual surveillance of shapes is also a projection of this concept [15]. However, optimal matching of a silhouette will meet some difficulties when complex postures are considered. In such occasions, similarities between shapes will sometimes be misleading. A comparison of conventional 2D approach and modern 3D models for posture recognition has been presented in ref. [8]. The 2D model depends on the camera point of view. This gives erroneous results when the camera is in motion. Similar methods which use 3D models recognize only standard postures with high accuracy. Therefore, additional measurements have to be taken to identify intermediate postures accurately.

In summary, there are several body postures that cannot be recognized with a satisfactory accuracy. Hence, a requirement of approaches, which are able to recognize a higher number of postures, emerges.

2.2 Recognition of arm posture

We use gestures frequently to communicate nonverbally or enhance what is communicated verbally [16]. Moreover, human gestures can be considered a resemblance of actions where most of the gestures in nonhuman primates lack the representational elements to perceive gestures [17]. Perceiving a gesture is important in exploring emotional aspects in an encounter as well as a physical activity of an individual [18].

Most of the existing literature to determine the hand pose related to fingers and gripping as in refs [19–21] was unable to identify the overall arrangement of the hand including upper and forearms. A hand posture recognition system which used discrete Adaboost learning algorithm with Lowe's scale invariant feature transform (SIFT) features has been proposed in ref. [22]. Here, the authors have taken an effort to reduce the issues posed by background noise in existing approaches.

In ref. [23], an effort was taken to anticipate future activities of a human to generate reactive responses from a robot. Spatio-temporal behavior of selected subjects has been observed by a robot over time to anticipate future trends in human and object behavior. Each possible tendency has been predicted using an anticipatory temporal conditional random field (ATCRF) which simulated rich spatial-temporal relations through object affordances. Similar approaches can be utilized to identify spatial behavior of certain body parts such as hands which form gestures.

According to the literature, color cues contribute mostly to recognize hand pose. But the spatial orientation of limbs seems to be helpful in determining the overall body language of hands. This fact was suggested by ref. [24]. The study conducted in ref. [25] is an example for the effectiveness of perceiving nonverbal behaviors over voice during human–robot encounters in hospitals. In the work explained in ref. [26], the robot understands the meaning of human upper body gestures and expresses itself by using a combination of body movements, facial expressions, and verbal language. Twelve upper body gestures including four gestures with the association of objects have been considered here. One drawback of the approach is the deployment of wearable sensors to recognize gestures and people in domestic environment will not prefer being disturbed by this equipment for a long time. But here, the robot's reaction toward a user was generated based on the emotions associated with the upper body gestures of its user. A robot teacher which could perceive affect of students and evaluate how that influences the engagement in a classroom is presented in ref. [27]. Affect was mostly deduced from four different facial expressions: “confused,” “distracted,” “engaged,” and “thinking,” and each expression was assigned with a positive/negative/neutral mood. This was combined with behavior recognition as well. For instance, bodily gestures such as “lower head,” “raise hand,” “sit still,” and “support chin” have been utilized to “map” people. People were mapped for their overall “state” of affect to be “confused,” “distracted,” “engaged,” and “thinking” in the end. This approach testifies to the fact that gestures contribute considerably to the internal state

of a person, although a limited number of gestures have been considered for the study. Furthermore, this work is an example where a combination of cues interprets emotional state of people better than a single cue alone.

Most of the gestures involving hands were detected with a considerable accuracy using body worn sensors as in ref. [28]. But humans prefer other methods as they will find it cumbersome to wear sensors in daily encounters. In ref. [29], authors used HAAR-like features which associate digital image features similar to Haar wavelets. This method requires a relatively smaller set of training data compared to other similar methods. Even so, context-free grammar is required to reach a high level of accuracy to analyze the syntactic structure. Most of the probabilistic approaches have shown proven accuracy in gesture recognition [30]. But these involve complex algorithms which slow down the decisioning process which has to be made in real time, consuming more computational power. The spatial behavior of joints in the human hand has been monitored over time to identify friendly gestures in ref. [31]. However, this method requires monitoring several body joints for a minimum duration of time to confirm a gesture. Hence, much simpler mechanisms to recognize arm posture for socially interactive applications are required for present robotic systems.

Still there are several arm postures that could not be recognized with a satisfactory accuracy. Hence, a requirement of approaches that are able to recognize a higher number of arm postures emerges.

2.3 Nonverbal interaction initiation

An assistive robot is expected to possess intelligence to determine when to interact with the user and when not to. This natural human-like behavior enhances the relationship between robot and its nonexpert user. Therefore, the robot must replicate complex human behavior in decision making [32]. However, the focus of these approaches is to maintain the interaction by analyzing user feedback rather than initiating the interaction. Furthermore, postures are beneficial for a robot in assessing interaction demanding of a user prior to a conversation. Even so postures are excluded as a factor to determine user situation in most approaches.

Work explained in ref. [33] states an affect recognition technique based on body language and vocal intonation. This approach has been used to interpret human affective cues and respond appropriately via display of their own emotional behavior. But it conducts diet and fitness counseling during HRI, not an HRI scenario in general. Somehow this can be considered a mechanism where the robot could

accurately guess the emotions associated with a situation. Mechanisms used in refs. [34,35] have evaluated emotions, gestures, pose, and movements of the human in his environment while interaction to measure interactive behavior. In ref. [36], nonverbal user engagement has been measured using initiator and responder gaze times, face orientation and feedback times for the two speakers. Bodily posture which is a critical parameter before and during an interaction is omitted in this method. In ref. [37], user's pose and movements have been taken into account before interaction initiation. But this system lacks any input about the changes in user's attention toward robot's presence. Therefore, there is an emerging requirement for modern cognitive systems to perceive human situation prior to an interaction. In addition, although torso and lower body are good displays of internal state of humans, literature related to the assessment of these are scarce. Analysis of arm and body postures can be an excellent mediator to analyze human state in designing cognitive social robots.

The study conducted in ref. [25] is an example for the effectiveness of perceiving nonverbal behaviors over voice during human-robot encounters in hospitals. In the work explained in ref. [26], the robot understands the meaning of human upper body gestures and expresses itself by using a combination of body movements, facial expressions, and verbal language. Twelve upper body gestures including four gestures with the association of objects have been considered here. One drawback of the approach is the deployment of wearable sensors to recognize gestures and people in domestic environment will not prefer being disturbed by this equipment for a long time. But here, the robot's reactions toward a user was generated based on the emotions associated with the upper body gestures of its user. A robot teacher which could perceive students' affect and evaluate how it affects the engagement in a classroom is presented in ref. [27]. Affect was mostly deduced from four different facial expressions: "confused," "distracted," "engaged," and "thinking," and each expression was assigned with a positive/negative/neutral mood. This was combined with behavior recognition as well. For instance, bodily gestures such as "lower head," "raise hand," "sit still," and "support chin" have been utilized to "map" people. People were mapped for their overall "state" of affect to be "confused," "distracted," "engaged," and "thinking" in the end. This approach testifies to the fact that gestures contribute considerably to the internal state of a person, although a limited number of gestures have been considered for the study. Furthermore, this work is an example where a combination of cues interprets emotional state of people better than a single cue alone.

2.4 Problem statement and the proposed mechanism

Techniques for arm and body posture recognition which could recognize an adequate number of postures with a higher accuracy are scarce. Furthermore, the aforementioned systems evaluate a number of parameters for evaluating the possibility of an interaction. Even so the possibility of integrating posture and gaze factors in monitoring user behavior is not adequately considered within these works. Therefore, this article presents a model to interpret user situation before interaction by analyzing postural behavior and this behavior is recognized by means of arm and body-based geometry derived through postures encountered during each scenario.

The proposed system adopts a novel geometry-based approach for recognizing a set of human arm and body postures. The standard body postures considered here are sitting, standing, and bending, which are most frequently adopted by humans in domestic environments. In addition to that frequent relaxing postures are considered as well. The same approach was applied to recognize often encountered arm postures as well. The system can also be utilized in self-employee environments where the same posture prevails for extended periods. In addition to that, postural behavior of the user is monitored by the system to identify user's interest toward interaction by evaluating his/her body language. This approach enables the robot make intuitive decisions during an encounter rather than adopting a behavior only based on a predefined set of actions.

This mechanism uses body joints and spatial arrangement of limbs to establish patterns in their arrangement. Spatial arrangement or the geometry of selected limbs with respect to each other is analyzed mathematically to differentiate various postures considered. This system is implemented on various activities which involve a number of postures to assess its functionality. A vision-based technique is used for posture identification in the system. Image processing techniques were equipped with extraction of feature points in 3D space using color and depth information. Spatial orientation of postures is mapped to the corresponding posture using a multi-layered feed forward neural network. Arm and body postures observed throughout a period were used to assess the emotional state of a certain encounter. This evaluation determines the appropriate behaviors for the assistive robot. We considered the nature of conversation and approach distance as the "approach behaviors" of the robot.

This article has two contributions.

- (1) A novel method to recognize a number of arm and body postures.
- (2) Evaluation of the effect of considering arm and body postures to generate human-aware responses by a robot prior to an interaction. In this context, the robot's responses are related to its approach behavior which includes mutual distancing and the type of conversation to initiate with its user.

3 System overview

The functional overview of the overall system is shown in Figure 1. This system evaluates factors that can be used to describe the nonverbal interaction demanding of a user, such as postural behavior. This “postural behavior” includes the changes in body posture and arm posture of a person over time. It is intended to enhance the capability of understanding user situation for deciding whether or not an interaction is appropriate at that particular moment.

This system takes spatial orientation of limbs as input to determine a particular posture. Limbs critical in forming a certain posture are used as vectors that make the posture arrangement. Joint information required to form such vectors is extracted from an individual by the Information Extraction Unit through RGB-D camera. Extracted information and calculated parameters are stored in the Data Recorder (DR) for further analysis at a later stage. After observing the user for a predefined period of time, Posture Identifiers (Body Posture Identifier and Arm Posture Identifier) are accessed to determine the current posture of its user and postural changes made by that user during the period of observation, T . Arm and body postures determined by Arm and Body Posture Identifiers are fed into the Interaction Decision Making Module (IDMM) to determine whether the postural behavior of the human is favorable for interaction. In addition to the outputs from Posture

Identifiers, data recorded in DR are used to take decisions regarding interaction initiation, by the IDMM.

Decisions regarding the initiation of interaction with the user are then transferred to Navigation Controller and Voice Response Generation Modules to take basic steps toward interaction such as; moving toward the user and greeting. Navigation Controller is used to achieve a socially interactive distance between the two conversants. In this case, conversant will be the robot and its user. Maps required for navigation within the specified environment are held in Map Repository. Voice responses are generated from the Voice Response Generation Module as the final stage of initiating the interaction. Sentences for the conversation between the user and robot are generated here.

4 Rationale behind using body language for situation-awareness

People undergo different states of mind in different situations. Happiness, insecurity, anxiety, relaxation, stress, confusion, fear, calmness, and anger, to name a few. There are physical means of displaying such mind sets involuntarily, such as the body language. For instance, we cross our arms when we undergo self-restraint or frustration. In contrast, we cross our legs and tilt the head, when we are interested of something.

The facts put forward in ref. [6] reveal fascinating trends in nonverbal human behaviors or body language. In this way, transmission of information is achieved through body-based behaviors such as facial expressions, gestures, touching, physical movements, posture, and body adornment. Nonverbal behavior comprises approximately 60–65% of all interpersonal communication or an interaction [38].

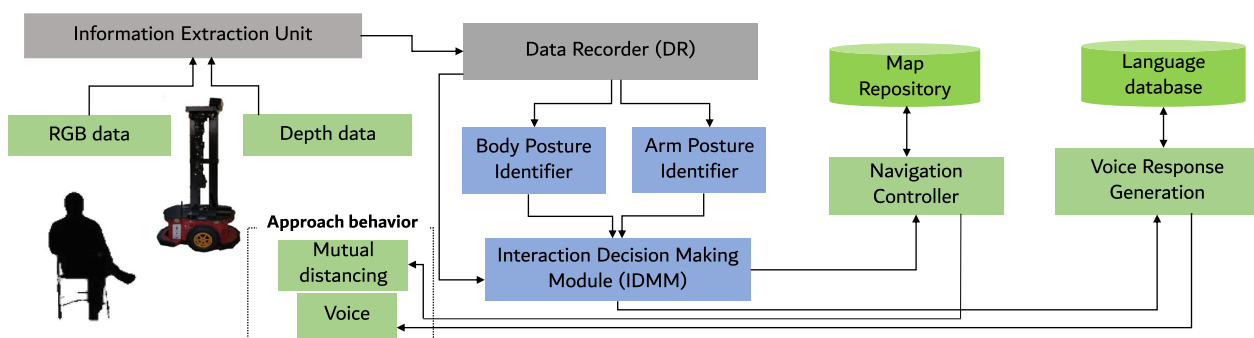


Figure 1: Overview of the system.

As nonverbal behavior can reveal a person's true thoughts, nonverbal behaviors are often referred to as "tells." Because people are not always aware of what they are communicating nonverbally, body language is more honest than a person's voluntary responses which can be crafted as required. Hence, this silent medium can successfully decode the real objectives and expectations of a human-involved scenario. This fact was utilized in a social environment to enhance a robot's perception upon its users before the robot initiates an interaction with them.

According to ref. [6], observing the context is the key to understanding nonverbal behavior. From a robot's perspective, it is important to understand its user's intentions with a careful observation of the person's nonverbals. Learning to recognize and decode nonverbal behaviors is important in this regard. Some bodily behaviors are universal because they are adopted similarly by many. In a typical human-robot encounter, idiosyncratic nonverbal behaviors which are unique to a person cannot be perceived in a single encounter. This will only be possible after gaining experience with that individual for a period of time. Therefore, in this research, we did not consider such unique nonverbal behaviors. But, as the best predictor of future behavior is the past behavior, it is important for a robot to observe its user's immediate past behavior to predict the state of that person in that instance. In order to identify the baseline behaviors of people with whom a robot regularly interacts with, the robot requires to identify changes in ordinary behavior. This behavior might take different forms such as postures, gestures, and facial expressions. These changes in behavior are common among people during interactions with fellow humans. Therefore, we evaluated such behavioral responses of a person toward a robot prior to an interaction. But modeling what goes in a person's mind in such an encounter is challenging and lacks conceptual basis.

4.1 Most honest part of our body

People use their faces to bluff or conceal true sentiments and truthfulness reduces from feet to head [6]. Lower body must receive a significant level of attention when collecting nonverbal cues. For instance, we cross our legs when we feel comfortable. Then we uncross our legs in the presence of someone (e.g., a stranger or an official). When the whole body including legs is considered, individuals

lean forward when they are thinking of something and they lean backward when relaxed. In addition, the degree to which we move arms is a considerable and accurate indicator of an individual's attitudes and sentiments. Normally, arms move effortlessly when they interact. In contrast, when upset or exited we restrict arm movements. Gravity-defying arm movements are often generated from joy and excitement. Hence, arms are plausible cues to access for moods and feelings. Certain arm behaviors gesture attention and some gesture denial. Therefore, in this research, we consider various arm postures to be perceived by a robot for their emotional interpretations as well.

4.2 Changes in behavior

Changes in behavior can help reveal how a person is processing information or adapting to emotional circumstances. Changes in behavior can further reveal a person's interest in an event. Careful observation of such changes can allow a robot predict the intentions upon an encounter in order to align robot's behavior accordingly. Hence, the robot's actions would not cause any intrusion or disturbance to the users. Knowing to distinguish between comfort (e.g., content, relaxation, happiness, excitement) and discomfort (e.g., displeasure, unhappiness, anxiety, stress) will help a robot decode user behavior and generate a response that will comfort that person. Learning to read cues of comfort and discomfort will help a robot decipher what their mind and the body are saying. For instance, people tend to keep their arms on hip, when there are issues or to show dominance.

To decode user behavior, the robot will observe them at a distance in our approach. Hence, neither the observation nor the interaction will be intrusive. For example, a person may distance herself from someone by leaning away. All such behaviors are controlled by the brain [39] and indicates whether a person is interested in or getting ready to avoid an encounter. Another similar example is that a person may rub his face or cheeks when in need of getting rid of a situation. Therefore, hands and the body are important to be considered in reading body language. In this research, we consider a number of arm and body postures to be distinguished by a robot before initiating an interaction with a human. A set of often encountered body and arm postures are shown in Figures 2 and 3, which we tried to recognize using the proposed concept. Hence, a robot may be able to interact with a human



Figure 2: A set of body postures encountered in daily chorus are shown. Each posture is called by an abbreviated form of their actual name and their actual names are explained in Section 4.2.

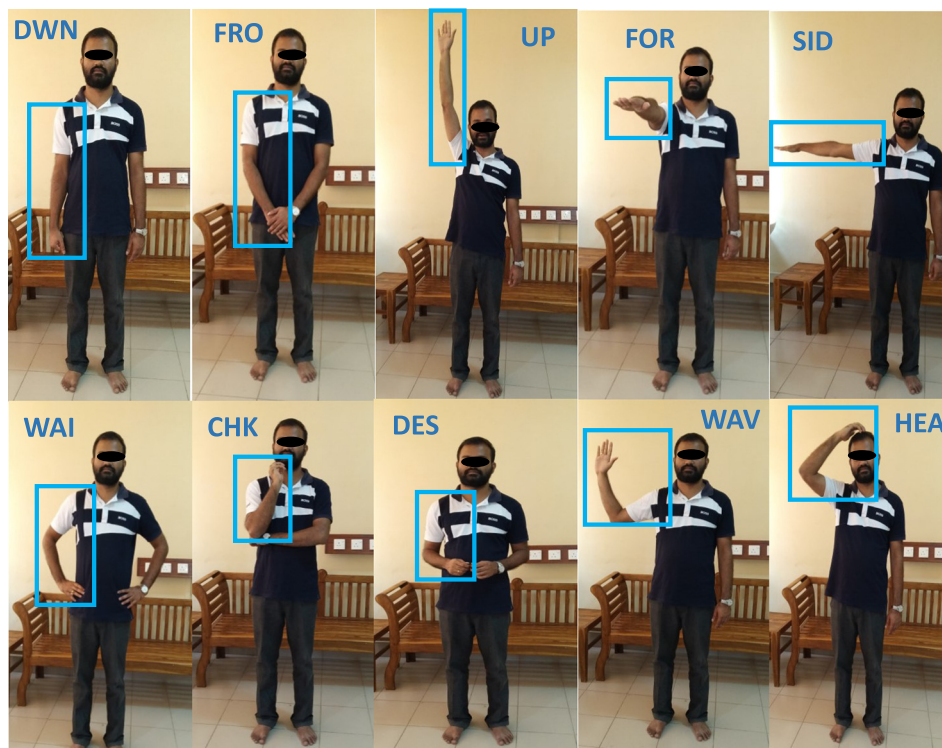


Figure 3: A set of arm postures encountered in daily chorus are shown. Each posture is called by an abbreviated form of their actual name and their actual names are explained in Section 4.2.

without territorial violations and disturbing behaviors. In addition, there are pacifying behaviors such as soothing necks or stroking necks. When we are in discomfort, our hands involuntarily respond to make us comfortable again. In addition, touching face, head, neck, and arms

can originate in response to a negative stimulus such as stress.

Set of body postures shown in Figure 2 identified by the proposed approach are abbreviated for the ease of referencing as follows.

STA	Standing
STDUP	Standing, with one leg up
LEA	Leaning forward
PIK	Picking up something
CRO	Crouch
SEA	Seated
SEAF	Seated, with spine bent forward
SEAB	Seated, with spine bent backward
MED	Meditation
SEAUP	Seated, with one leg up
BEN	Seated, with legs up and bent

Similarly, the set of arm postures shown in Figure 3 identified by the proposed approach are abbreviated as follows.

DWN	Hand down
FRO	Hand in front
UP	Hand up
FOR	Hand forward
SID	Hand to side
WAI	Hand on waist
CHK	Hand on cheek
DES	Hand as in a desk activity
WAV	Waving hand
HEA	Hand on head

5 Posture recognition

We selected a set of limbs in our body, which make up most of the arm and body postures. We tracked each of the limbs using the Information Extraction Unit discussed under Section 3. The RGB-D camera tracks the joints and the vectors joining these joints will make up the limbs. The process of creating these limbs is given in detail below.

5.1 Mechanism used to differentiate postures

To detect a posture uniquely from a set of postures, orientation of vectors drawn along selected limbs in human body is considered. In other words, each limb that makes a posture, becomes a vector. A vector is formed by combining two adjacent joints in the skeleton. Once a person is tracked, his/her skeletal arrangement is extracted to analyze the positioning of a selected set of vectors in

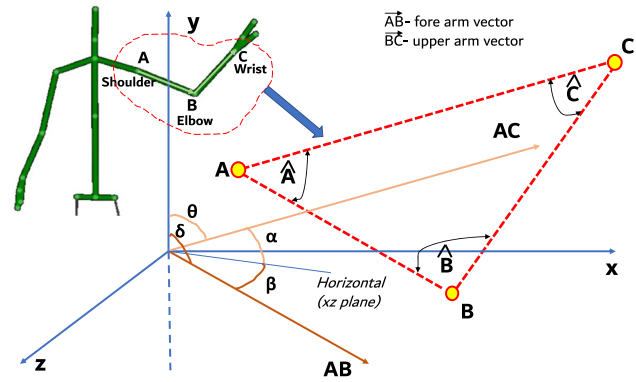


Figure 4: Body angles derived from the limbic vectors of a person with stretched hands are shown. Coordinate system as seen from the camera is shown as X-Y-Z. The angles formed by the three limbic vectors with the XZ-plane are marked as α and β , while the angles made with the XY-plane are marked as θ and δ . These angles are used to recognize the arm posture. A, B, and C denote shoulder, elbow, and wrist joints.

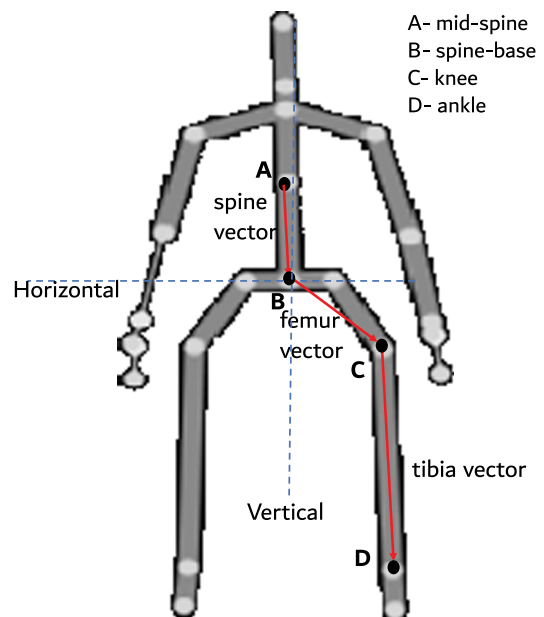


Figure 5: Limbic vectors of a seated person are shown here. The limbic vectors are derived by joining A (spine-mid), B (spine-base), C (knee), and D (ankle) joints.

3D space. This is demonstrated in Figures 4 and 5. Figure 4 shows two vectors: forearm vector and upper arm vector, which are used to determine arm posture. To determine the spatial orientation of the body, three major limbs, spine, femur bone, and knee, are considered. A vector is drawn across the limb, joining the end joints of that limb, according to the extracted skeleton. The angles this vector make with the XZ plane (horizontal) and Y are

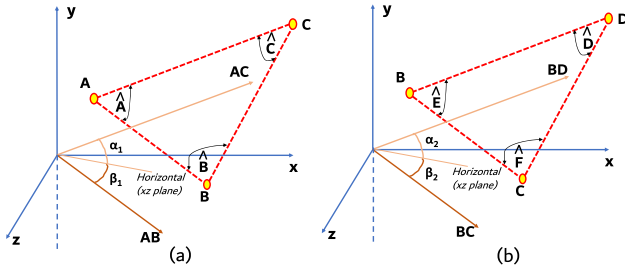


Figure 6: Body angles derived from limb vectors of a seated person as in Figure 5 are shown here. Coordinate system as seen from the camera is shown on bottom left. The angles formed by the three limb vectors with the xz plane are marked as $\alpha_1, \beta_1, \alpha_2, \beta_2$. (a) Orientation of the upperbody vectors is shown by the triangle ABC, (b) orientation of the lowerbody vectors is shown by the triangle BCD.

also measured. Y direction is marked along the vertical of the camera of the robot, as marked in Figures 4 and 6. Spine, femur, and tibia vectors shown in Figure 5 are the vectors used to determine the body posture. Front view of a skeletal while sitting is shown in Figure 5.

5.1.1 Arm postures

Three limbic vectors used to derive arm posture are marked in Figure 4. Following equations are used to calculate each vector considered. A, B , and C denote the coordinate positions of shoulder, elbow, and wrist joints, respectively.

$$\vec{AB} = (x_{\text{elbow}} - x_{\text{shoulder}})\vec{i} + (y_{\text{elbow}} - y_{\text{shoulder}})\vec{j} + (z_{\text{elbow}} - z_{\text{shoulder}})\vec{k}, \quad (1)$$

$$\vec{BC} = (x_{\text{wrist}} - x_{\text{elbow}})\vec{i} + (y_{\text{wrist}} - y_{\text{elbow}})\vec{j} + (z_{\text{wrist}} - z_{\text{elbow}})\vec{k}, \quad (2)$$

where AB and BC denote the vectors drawn along the upper arm and the forearm. x, y , and z denote the coordinates in i, j , and k directions, respectively, in usual vector notation. The labels shoulder, elbow, and wrist denote the shoulder, elbow, and wrist joints. The set of inputs required to identify the particular arm posture is represented by H as follows.

$$H_i = \{\hat{A}, \hat{B}, \hat{C}, \delta, \theta, \alpha, \beta \mid i = 1, 2, \dots, n\}, \quad (3)$$

where, i denotes the i th posture considered out of n number of postures. $\hat{A}, \hat{B}, \hat{C}, \delta, \theta, \alpha$, and β are measured as marked in Figure 4. δ and θ are the angles measured from the vertical (XY plane) drawn through the spine-base joint of the skeletal. α and β are angles measured in three-

dimensional space from the horizontal (XZ plane) drawn through the spine-base joint of the skeletal.

For example, the set of angles made by limbic vectors corresponding to a standing human posture can be represented as follows.

$H_i = \{18^\circ, 148^\circ, 14^\circ, 2^\circ, 17^\circ, -70^\circ, -87^\circ\}$ if the i th arm posture is considered as an ordinary standing posture: STA with hands down: DWN in Figure 3.

5.1.2 Body postures

Angles related to knee and femur bones are calculated for the right-hand side of the body. Orientation of these three vectors is different from posture to posture. Even though other limbic vectors that change with respect to posture are present, these three vectors are critical in differentiating two postures that are slightly deviated from each other. Furthermore, majority of the human body responsible for a posture is included in the above three vectors. Vectors along the limbs in arms are not considered as hands are often moved by humans even without a specific need to do so. Therefore, arms are not considered for body posture recognition during this study. The set of angles denoted by P_i in (7) includes the angles used to calculate the orientation of limb vectors shown in Figure 5.

Limbic vectors required to calculate these angles are shown in Figure 6 and these vectors are calculated as follows.

$$\vec{AB} = (x_{\text{spine_base}} - x_{\text{mid_spine}})\vec{i} + (y_{\text{spine_base}} - y_{\text{mid_spine}})\vec{j} + (z_{\text{spine_base}} - z_{\text{mid_spine}})\vec{k}, \quad (4)$$

$$\vec{BC} = (x_{\text{knee}} - x_{\text{spine_base}})\vec{i} + (y_{\text{knee}} - y_{\text{spine_base}})\vec{j} + (z_{\text{knee}} - z_{\text{spine_base}})\vec{k}, \quad (5)$$

$$\vec{CD} = (x_{\text{ankle}} - x_{\text{knee}})\vec{i} + (y_{\text{ankle}} - y_{\text{knee}})\vec{j} + (z_{\text{ankle}} - z_{\text{knee}})\vec{k}, \quad (6)$$

where AB, BC , and CD denote the vectors drawn along the spine, femur and tibia limbs. x, y , and z denote the coordinates in i, j , and k directions, respectively, in usual vector notation. Labels mid_spine, spine_base, knee, and ankle denote the spine-mid, spine-base, knee, and ankle joints marked in Figure 6.

$$P_i = \{\hat{A}, \hat{B}, \hat{C}, \alpha_1, \beta_1, \hat{D}, \hat{E}, \hat{F}, \alpha_2, \beta_2 \mid j = 1, 2, \dots, m\}, \quad (7)$$

where $\hat{A}, \hat{B}, \hat{C}, \alpha_1, \beta_1, \hat{D}, \hat{E}, \hat{F}, \alpha_2, \beta_2$ are as marked in Figure 6. j denotes the j th posture considered out of m

number of postures. A set of actual values obtained for a specific posture are shown below.

$$P_i = \{6^\circ, 162^\circ, 12^\circ, -82^\circ, -87^\circ, 5^\circ, 167^\circ, 8^\circ, -77^\circ, -73^\circ\}$$

if, i th body posture is considered as “standing” or STA in Figure 2.

Identifying postures in the above manner is a function of the Body Posture Identifier and Arm Posture Identifier. A deep neural network is used to relate the above body angles in P_i and H_i to a particular posture. Inputs to the neural network are $\hat{A}, \hat{B}, \hat{C}, \alpha_1, \beta_1, \hat{D}, \hat{E}, \hat{F}, \alpha_2$, and β_2 in Body Posture Identifier. The output from this neural network is the closest posture out of the specified m postures. This output is considered for all sets of angles recorded each second throughout T s. This is similar in Arm Posture Identifier as well. Inputs to the neural network are $\hat{A}, \hat{B}, \hat{C}, \delta, \theta, \alpha$, and β in Arm Posture Identifier and the output from this neural network is the closest posture out of the specified arm postures.

6 Behavior evaluation model

6.1 Posture change over a timeline

During certain activities, humans change their posture at least once. Therefore, posture changes must be analyzed for an adequate duration. The system observes the user for T seconds from the moment a user’s skeletal was identified. DR records raw information such as coordinate positions of specific joints and calculated parameters such as joint vectors throughout this period. At the end of T s, set of information recorded in DR is analyzed for posture identification and decision evaluation. An example of this evaluation for the period of observation equal to T s is shown in Table 1.

In Table 1, arm and body postures perceived by the model in each second throughout the period of observation are shown. This postural behavior is taken into consideration prior to the interaction. Since different postures have different emotions associated with them, an evaluation of postural behavior makes more sense than

considering the postures observed in the beginning of a scenario. Hence, the decision making is affected by this “postural behavior” of the user. The “initial posture” was recorded at $t = 1$ s and the “final posture” was recorded at $t = T$ s. A separate experiment was conducted to find the optimum value for the period of observation during this type of occasions. This is elaborated in Section 8.5.1.

6.2 Availability of the user

The intention of deploying an assistive robot is to uplift the mental condition of its user, through friendly interaction at user’s leisure or lonely hours. Therefore, the robot’s interaction approach must consider appropriate situations to achieve this outcome. IDMM assesses the availability of the user by analyzing the change of his/her behavior by means of posture over time. The idea behind this evaluation is as follows.

According to behavioral sciences, there is a trend in humans to change their behavior when somebody is around. Therefore, the user’s response toward the robot changes, when he/she sees the robot nearby. A common example for this type of situation is that humans slant their spine (SEAB in Figure 2) when seated for a long duration, but when someone comes near, they adjust their postures to standard sitting posture (SEA in Figure 2) with erect spine. Humans notice such changes in behavior in someone when they intend to interact with that person. Instead of the human, it is the assistive robot who assesses such incidents in domestic environment in this context. This is the motivation behind the use of postural behavior instead of only posture in this approach.

7 Posture-based interaction decision making

A number of postures are found in humans while engaged in various activities in domestic environments. Out of these, often found 11 ($m = 11$) postures shown in Figure 2 were considered in our approach. These include “standing” and “seated” postures and postures in-between. Postures LEA and PIK can also be found while performing a certain task. These are referred to as “intermediate” postures. Bending down to pick a fallen object is an example of a task with such postures. Awkward postures that are unique to humans, which are especially found in children, are not considered here. Postures SEAUP and STAUP have two

Table 1: Observations made through time

Time (s)	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$...	$t = T$
Body posture	CHK	CHK	STA	STA	STA	...	STA
Hand posture	DES	DES	DES	DES	DES	...	DES

different angles for right and left femur and knee vectors, respectively. In such instances, training data sample has postures with both right and left sides included. The reason for this is that only the main postures are responsible to assess the availability of a human for interaction. Limbic vectors vary during each posture considered here. Similarly, there are ten arm postures ($n = 10$) which were selected to be recognized using the proposed approach.

As a prior step before interaction, it is important for the robot to evaluate the user's posture to gain knowledge on current situation of the user. On the one hand, posture is a measure of the emotional condition of a person. On the other hand, postures are related to a user's current activity [41]. Therefore, the robot is expected to evaluate posture as a measure of interactivity of a particular person. Although there are no strict rules on the relationship between posture, emotional state, and physical activities, there are conventional scenarios where user situation can be approximated by analyzing posture information.

Interaction decisions are made after analyzing posture changes according to behavior evaluation model. Information required by the IDMM are given in the set denoted by H_i and P_i in (3) and (7).

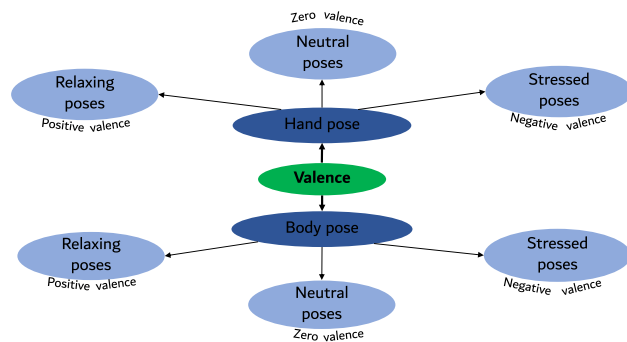


Figure 7: A semantic map illustrating the components that make valence: hand and body poses. Each pose is categorized into three groups: relaxing, neutral, and stressed. This categorization is based on the emotions associated with each pose in general human perception.

7.1 Valence of a situation

Once the postures observed in a single scenario are known, a variable called “valence” is calculated to assess the “internal state” of that person quantitatively. To calculate valence, we categorized postures into three groups namely, “relaxing poses,” “neutral poses,” and “stressed poses.” This is shown in Figure 7. Postures adopted during relaxing situations were categorized as “relaxing poses,” and are assigned a positive valence. Similarly, postures adopted during stressed situations were categorized as “stressed poses” and are assigned a negative valence. Standard postures were assigned the value zero for valence. This categorization was done after analyzing some simulated domestic encounters according to the generally accepted social norms. Assignment of a valence for each posture was influenced by the study shown in ref. [40]. Valence for each posture was determined after this analysis and the values assigned for the valence are given in Table 2. Assigning valence and categorizing postures according to their emotional state were empirical and based on how humans react in social encounters in general. Hence, we simulated such encounters and arranged arm and body postures in an order of their positive emotional state. Values for the valence were assigned after this step.

In a single scenario, a person adopts both body posture and hand posture. In such a case, the valence of the scenario was calculated as follows.

$$\text{Valence} = k_1 * \text{Valence}_{\text{Body Posture}} + k_2 * \text{Valence}_{\text{Hand Posture}} \quad (8)$$

$\text{Valence}_{\text{Body Posture}}$ and $\text{Valence}_{\text{Hand Posture}}$ indicate the valence corresponding to body posture and hand posture as given in Table 2. k_1 and k_2 indicate the weights given to body posture and arm posture, respectively, in determining the valence of a situation.

Interaction decisions made by the robot based on the calculated valence are shown in Figure 8. The values to separate margins between each type of conversation and mutual distancing were empirical and were decided by

Table 2: Valence assigned for each hand and body posture depending on the general interpretation of emotional state behind each posture

Arm posture	DWN	FRO	UP	FOR	SID	WAI	CHK	DES	UP	WAV	HEA
Valence	0	+2	+4	-6	-4	+6	+8	-2	-6	+10	-8
Body posture	STA	BEN	LEA	PIK	CRO	SEA	SEAF	SEAB	MED	SEAUP	STDUP
Valence	0	+8	-8	-6	+2	0	-4	-2	-10	+6	+4

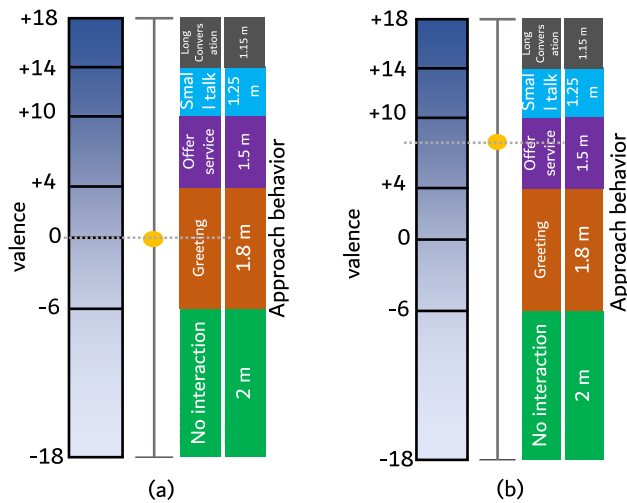


Figure 8: Marginal values for valence to determine appropriate approach behavior of the robot. (a) Initial valence is marked as 0 and this corresponds to “Greeting” at a distance of 1.8 m as the appropriate approach behavior. (b) Final valence increased to +8 which resulted in an approach behavior of “asking for a service” and keeping a mutual distance of 1.5 m.

repeating a few trial experiments before implementing the IDMM. Corresponding type of conversation and the mutual distancing appropriate for a particular instance were determined according to the value of the valence as shown in Figure 8. The margins for each approach behavior are set by simulating a number of occasions in domestic encounters and human responses if a human approaches another human instead of the robot. This simulation helped in determining the values that were used for the valence in each arm and body posture shown in Table 2. Postures were listed in order of their relaxing or stressed nature as per the human perception and then a numerical value for the valence is assigned.

7.2 Decision-making in dynamic user behavior

In a dynamic user behavior, several changes in body and arm postures could be observed. In such occasions, the initial and final postures observed during the period of observation, T were considered in decision-making. If the difference between the two valences: initial valence and final valence, is an increment, the approach behavior is promoted by one set. The conversation types used in this work are listed as follows:

- Greeting
- Offering a service
- Small talk

- Long conversation
- No interaction

For instance, if the valence was increased during a certain occasion and the approach behavior corresponding to the final valence is “Greeting” and 1.5 m, this is promoted to “Offering a service” and 1.8 m. For this promotion, we referred to the chart shown in Figure 8. A combination of a mutual distance and the corresponding conversation is referred to as a “set.” Similarly, if the difference between the initial and final valences is a decrement, the approach behavior is demoted by one set. For instance, if the final valence corresponds to {Greeting, 1.5 m}, this is demoted to {No Interaction, 2 m}. This decision making criterion is given in the following algorithm.

Require: Initial valence, final valence

Ensure: Mutual distance, type of conversation
(Approach behavior)

Arm and body posture transformations known

if final valence = initial valence **then**

 Select the corresponding set of approach behavior

else

if final valence < initial valence **then**

 Select the corresponding set of approach behavior for the final valence

 Demote corresponding approach behavior by 1 set

else

if initial valence < final valence **then**

 Select the corresponding set of approach behavior for the final valence

 Promote corresponding approach behavior by 1 set

end if

end if

end if

8 Experiment and results

A set of experiments were conducted in support of the proposed concepts of posture recognition and robot’s decisioning.

The robot was placed at a predefined location in the map and was allowed to wander within the specified map. The routes that were covered by the robot are determined by a teleoperator. Once a body was tracked, the robot stops its motion and starts observing. During this observation, DR stores information and Body Posture

Identifier and Arm Posture Identifier are initiated. After the recognition of postures, IDMM is initialized to take interaction decisions. Once the robot decides an appropriate type of conversation and the mutual distance, this distance and the conversation were implemented by the teleoperator as in wizard-of-oz experiments. Continuing a conversation by understanding user's voice is out of the scope of this research. Therefore, the teleoperator handled the implementation of the set of interaction decisions. Hence, the teleoperator instructs the Voice Response Generation Module to utter the statements that she types. Navigation Controller takes care of maintaining the mutual distance decided by the IDMM. We implemented the proposed concept to the right side of the body and the corresponding gesture was performed by the right hand. Therefore, we selected only right-handed people as subjects of this research.

The set of experiments are explained as follows.

8.1 Determination of T (Experiment 01)

Since a period of observation was necessary for the robot to continue with the interaction decisions, the first experiment was conducted to determine T in the first place. Therefore, a set of domestic activities were selected and a person was allowed to engage in a particular activity for a period of time. During that period, posture changes and the time gap between two consecutive different postures was recorded. For the experiment, 11 users aged from 26–59 years (SD: 12.83 and mean: 31.63) participated. The participants were given a task to perform, e.g., standing, without further instructions on how to perform the activity. This

was because we tried not to influence people's postures by instructing on how to perform a task. We observed the postures that they performed and therefore we could identify variations in posture while performing the same activity. This step was taken to retain the natural behavior of people although we conducted it in a simulated environment. The participants were given the chance to stop the activity whenever they feel like it. Details of the activities selected are given in Table 3. Each participant was asked to perform at least three tasks which resulted in 33 activities observed altogether. Average values for the variables observed during each activity are given. In Table 3, "minimum and maximum times between two consecutive postures" refer to the average time observed between two postures, if posture changes were observed during the period. Here, a single user was observed for 30 min unless the user walked away in the middle of observation. To determine T , we observed for only body postures because the hands are dynamic and change faster than the body in general. To determine T , postures were recognized using the mechanism explained in Section 5.

8.2 Recognition of arm and body postures (Experiment 02)

A feed forward neural network was trained to map the set of angles made by the limbic vectors to the corresponding **arm posture**. We used a feed forward neural network of three layers and 50 neurons in the first and second layers. Training data set consisted of posture information of 21 human subjects (SD: 11.25 and mean: 29.1). We tried several other hyper parameters such as more and lesser

Table 3: Results of Experiment 01

Activity	Postures observed	Average time between two consecutive body postures (s)	Minimum and maximum times between two consecutive body postures (s)
1 Standing, waiting	STA, STDUP	3.8	2, 8
2 Sitting, relaxing	SEA, SEAF, SEAB, SEAUP, BEN	5.1	3, 9
3 Desk activity (reading/ studying)	SEA, SEAF	6.4	3, 12
4 Cooking	STA	6.1	4, 16
5 Making a phone call, sitting	SEA, SEAF, SEAB, SEAUP, BEN	8.4	5, 76
6 Making a phone call, standing	STA	—	—
7 Standing, engaged in lab work	STA, LEA	5.3	3, 16

neurons and hidden layers as well. But in the end, we could observe an optimum behavior of the algorithm for recognizing these postures at the values given here. The inputs and the output of this network can be stated as follows:

Inputs: $\hat{A}, \hat{B}, \hat{C}, \delta, \theta, \alpha, \beta$

These are the inputs shown by the set H_i in (3). We exempted the two angles \vec{AC} makes with the vertical and the horizontal, from the current set of inputs as we did not observe an improvement in the training and testing accuracies of the dataset with the additional data. This holds true for the inputs selected for the recognition of body postures as well. Therefore, we observed the current set of variables as the optimum set of inputs in our approach.

Output: Corresponding hand posture out of {DWN, FRO, UP, FOR, SID, WAI, CHK, DES, WAV, HEA}

The module called Arm Posture Identifier performs the aforementioned functions.

Another feed forward neural network was used to map the set of limbic vectors to the corresponding **body posture**. This feed forward neural network also consisted of three layers and 50 neurons in the first and second layers. Data related to body posture were also taken from the same subjects as for the hand posture.

Inputs: $\hat{A}, \hat{B}, \hat{C}, \alpha_1, \beta_1, \hat{D}, \hat{E}, \hat{F}, \alpha_2, \beta_2$

These are the inputs shown by the set P_i in (7).

Output: Corresponding body posture out of {STA, STDUP, LEA, PIK, CRO, SEA, SEAF, SEAB, MED, SEAUP, BEN}

Different seating arrangements make a body posture slightly different from each other. Furthermore, people tend to keep their legs at different angles when seated. Figure 9 shows three examples for this. Such instances were also considered while collecting data for the training data set.

The module called Body Posture Identifier performs the above functions. In the end, we used this trained algorithm to recognize posture changes in Experiment 01.

8.3 Implementation of IDMM (Experiment 03)

After implementing the recognition algorithm and determining T , we used these findings to determine the internal state of the people before an interaction. Throughout the article, we refer to robot-initiated interactions in this regard. We instructed the participants to perform a selected set of activities listed under “description” in Table 4. IDMM was implemented and the interactivity of the situation was

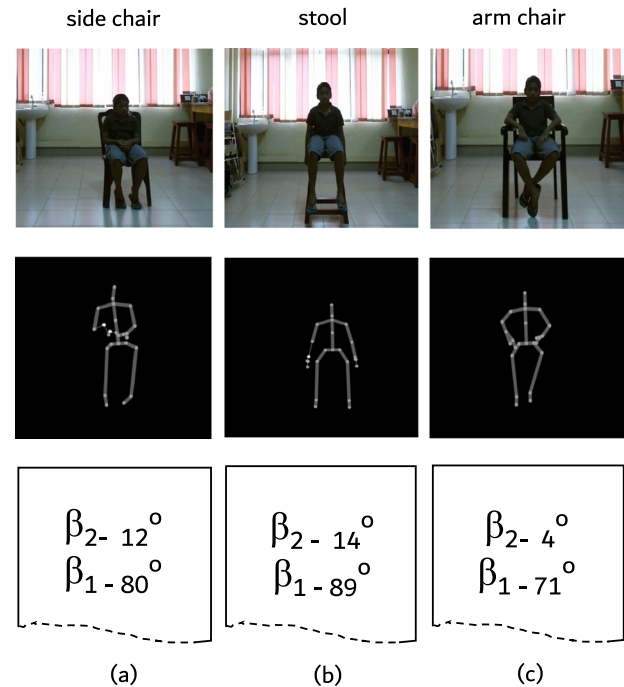


Figure 9: A comparison of the limbic arrangement for regular sitting posture (SEA) when the seated chair is changed. This occurs due to the difference in dimensions of each chair type. (a), (b), and (c) represent the user posture, the corresponding skeletal image, and two angles β_1 and β_2 for side chair, stool and arm chair, respectively.

evaluated at this stage while the participants were performing these activities. The same set of participants in Experiment 02 participated in Experiment 03 as well. We asked each participant to perform at least two activities and altogether we considered 40 different scenarios for Experiment 03. Twenty scenarios out of these are shown in Table 4. All the participants were selected outside the university and over 80% of the participants had no technical background.

Before implementing IDMM, the robot was remotely navigated toward the participant adopting the approach behavior as “Offering a service” and 1.5 m. User was asked to give a feedback score for the robot’s behavior by considering user satisfaction. This is used as the ground truth for this experiment. Then the entire scenario was performed again and the valence of the set of postures was calculated for each scenario according to (8). Then the robot was commanded by a teleoperator to adopt the approach behavior determined by the IDMM. IDMM decided upon an approach behavior by evaluating the *valence*. For the experiments, we assigned 1 for k_1 and k_2 in (8), assuming that the arm and body postures contribute equally for the current user situation. Then the participant was asked to give a feedback score (rating

Table 4: Results of Experiment 03

Initial postures (Body pose Hand pose)	Transformation (Body pose Hand pose)	Description	Valence	Robot responses	Feedback score (IDMM)	Feedback score
1 (STA DWN)	—	Person walking slowly	0	Greeting, 1.8 m	9	5
2 (STD UP WAI)	—	Person thinking, relaxed	+8	Offering a service, 1.5 m	9.5	9.5
3 (LEA DWN)	—	Cleaning the floor	-8	No interaction, 2 m	5	4
4 (PICK WAV)	—	Picking something up, then stalled to wave at robot	+4	Offering a service, 1.5 m	10	10
5 (CRO DWN)	—	Cleaning the floor	+2	Greeting, 1.8 m	7.5	4.5
6 (SEA DES)	—	Seated, relaxed	-2	Greeting, 1.8 m	9	5.5
7 (SEAF DES)	—	Seated, relaxed	-6	Greeting, 1.8 m	9.5	6
8 (SEAF WAV)	—	Seated, waves at robot	+6	Offering a service, 1.5 m	10	10
9 (MED FRO)	—	Meditating	-16	No interaction, 2 m	9	6.5
10 (SEA UP CHK)	—	Seated relaxed with one leg up Hands on cheek	+14	Long conversation, 1.15 m	10	5
11 (BEN FOR)	—	Seated, hand forward	+2	Greeting, 1.8 m	7	8
12 (STA DWN)	(LEA FRO)	Was seated, then stands and relaxes	0 ≥ -6	No interaction, 2 m	10	6.5
13 (SEA DWN)	(STA FRO)	Was seated, then stands.	0 ≥ +2	Greeting, 1.8 m	10	6
14 (SEA DWN)	(SEAF DES)	Seated, then bends forward.	0 ≥ -6	No interaction, 2 m	7.5	7.5
15 (SEA DWN)	(SEAB DES)	Seated, then bends backward and relaxes	0 ≥ -4	No interaction, 2 m	9.5	7
16 (STA DWN)	(STA WAV)	Standing, then waves at robot	0 ≥ +10	Long conversation, 1.15 m	10	6
17 (LEA DWN)	(STA UP)	Exercising	-8 > +4	Small talk, 1.25 m	4	9
18 (STA DWN)	(STA HEA)	Standing, then scratches head	0 ≥ -8	No interaction, 2 m	7	6.5
19 (STA UP)	(STA FOR)	Exercising	+4 ≥ -6	No interaction, 2 m	10	4
20 (STA DWN)	(STA WAI)	Standing, then engaged in deep thought	0 ≥ +6	Small talk, 1.25 m	8	5.5

from 0 to 10) for the approach behaviors of the robot during the two occasions. Feedback score was given by users based on how they preferred or how comfortable were the decisions of the robot during that particular situation. Users were asked the reason to give their corresponding feedback score and important remarks during reasoning are highlighted in the discussion.

8.4 Research platform

Proposed concept has been implemented on Mirob platform attached with a Microsoft Kinect sensor. This is a Pioneer 3DX MobileRobots platform equipped with a Cyton Gamma 300 manipulator. The robot required an initial map of the environment for navigation and path planning. Therefore, we initially defined a map of the simulated environment with the placement of furniture, doors, and walls. Dynamic obstacle avoidance was possible with the MobileRobots platform. Navigation maps required for locomotion were created with Mapper3 Basic software. Skeletal representation of human body was extracted as 3D co-ordinates of feature points in Kinect sensor. The experiment was conducted in an artificially created domestic environment with the participation of users in a broad age gap. Mirob is shown in Figure 10.

8.5 Results and discussion

8.5.1 Determination of T (Experiment 01)

According to the results presented in Table 3, several postures could be observed during a single activity. Results obtained for a single user during the listed activities are shown in the table. All the postures that were observed from each user are listed under “Postures observed.” Number of postures encountered during an activity increases with the duration of the activity. Therefore, when the person was allowed to engage in an activity for a long duration, up to six postures could be observed. In contrast, during short-term tasks such as standing/waiting or making a phone call, only a few postures were observed. A number of posture changes were observed in “making a phone call while sitting.” This is because the person finds more relaxing postures when the duration of the call is longer.

As a whole, the number of posture changes in seated activities was greater than that of standing activities. In the meantime, the duration between two consecutive

postures was also recorded “lesser” during seated activity. An interesting fact revealed during the study was that the initial posture change takes place short after the start of the activity if the person was sitting. This is because of the numerous sitting postures available to choose. Often, the users did not see another person approaching him, when he/she was deeply engaged in the task. Such scenarios are omitted in this experiment. Sometimes, the users tend to talk to the person approaching them without looking. This type of scenarios cannot be used for the assessment of nonverbal interaction demanding. Such complications in human behavior are yet to be solved. This is not observed in standing postures due to the limited number of comfortable standing postures available. Out of the results obtained, it was at the eighth second, the person started to change the posture. That is the average time between two postures was 8 s. Therefore, 8 s was used as the practical value for the period of observation, T .

This value was used for the period of observation in rest of the experiments.

8.5.2 Recognition of arm and body postures (Experiment 02)

The model developed to recognize hand posture resulted in a training accuracy of 97.05% and a testing accuracy of 96%. The model to recognize body posture resulted in a training accuracy of 98.5% and a testing accuracy of 97.5%. These two models were integrated to achieve the outcomes of the IDMM. That is, we used these two algorithms separately to recognize arm and body postures of a person during an encounter. One algorithm was run after the first and then recognized arm and body postures were stored in the IDMM for decision making. At the beginning of the observation, the initial postures were recognized and the final postures were determined at the end of T . Then the transformation of the participant from the initial set of postures to the final set of postures was identified. This transformation is given under “Initial Posture” and “Transformation” columns in Table 4.

8.5.3 User responses prior to interaction (Experiment 03)

Table 4 shows the results of Experiment 03: Implementation of the IDMM. “Initial Postures” column indicates the set of body and hand postures observed by the robot at the beginning of time T . “Transformation” column gives the set of body posture and hand posture at the end of T .

The column “Description” gives a detailed image of the user situation. The valence of each scenario is given in the “Valence” column. The type of conversation and the mutual distance are given as a set in the “Robot Responses” column. “Feedback score (IDMM)” shows the feedback (out of 10) received by the user in each scenario when the IDMM was implemented. “Feedback score” gives the score given by the same user for the same occasion when the ground truth was implemented. When the user was stationary, no transformation was found in the initial set of postures and therefore the column is kept empty. When the users had posture changes over the period of observation, the finally observed set of postures was given under the “Transformation.” Typical scenarios and our critical observations which are deviated from the expected outcome during the experiment are discussed below.

In scenario 1, the initial postures were (STA|DWN) and there were not any transformation in the postures for the entire duration. During the scenario, the person was walking slowly. Therefore, the person was standing (STA- valence 0), holding hands down (DWN- valence 0). This recorded a valence of 0 when both hand and body postures were considered, which resulted in a greeting and a mutual distance of 1.8 m. This is shown in Figure 8. As the user was walking, he preferred a greeting rather than a conversation. But once the robot spoke, he stopped to greet back. This occasion is shown in Figure 10(a). Hence, the approach behavior received a high feedback score of 9.

In the next scenario, the person was standing, doing nothing. Body and hand postures corresponding to this scenario were STDUP and WAI that resulted in an overall

valence of +8. This resulted in robot asking for a service at a distance of 1.5 m. As the person was relaxed and willing to talk, a feedback score of 9.5 was received.

The feedback received for the third scenario was low compared to the previous occasions. Here, the person was cleaning the floor, where he was engaged, as perceived by the robot. Therefore, the robot’s decision was not to have any interaction, even though the user preferred to have a friendly conversation in such a task.

In the 12th scenario, a dynamic user behavior was observed. First the user was sitting and then he stood up. Hence, the initial postures transformed from (STA|DWN) in to (LEA|FRO). This resulted in a drop of valence from 0 to −6. This resulted in a response of no interaction and a mutual distance of 2 m from the robot. This behavior of the robot received a feedback score of 10 due to higher user satisfaction because of not getting disturbed.

In the 17th scenario, (LEA|DWN) transformed in to (STA|UP) as the user was exercising. But the valence increased from −8 to +4 during posture transformation. This resulted in a small talk and a mutual distance of 1.25 m according to the decision-making criteria in dynamic user behavior. This resulted in low user satisfaction as the user was distracted by robot’s behavior. Figure 10(b) and (c) corresponds to a situation where user situation changed from (SEA|DES) to (SEAB|FRO). Valence changed from −2 to 0 that corresponds to a small talk at a distance of 1.5 m. Here the valence had an increment within the transfer of initial to final occasion and 0 corresponds to “greeting” at a distance of 1.8 m. When this approach behavior is promoted by 1 set, the new approach behavior became “greeting” at 1.5 m.

Table 5 compares the feedback scores received after implementing the IDMM with the ground truth. These feedback scores are obtained from Table 4 which shows the results of Experiment 03. We compared the feedback received for implementing the ground truth (without IDMM) and the IDMM (with IDMM). These feedback are given under “feedback score” and “feedback score (IDMM)” columns in

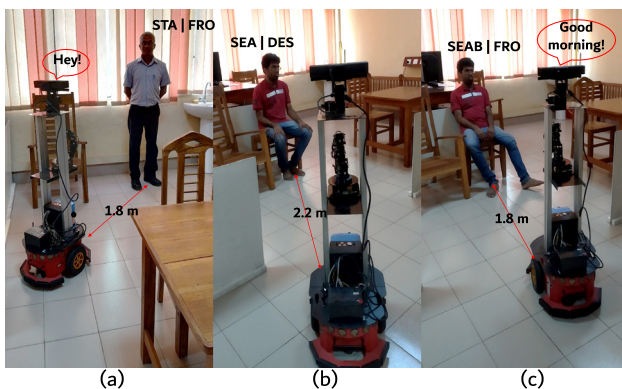


Figure 10: Two occasions during Experiment 03 is shown. (a) Occasion 1: Responses of the user when the approach behavior: “greeting at 1.8 m” was adopted by the robot. (b) Occasion 2: Robot observes the user at a distance. (c) Final position of the robot and the behavior of the user. Recognized postures and mutual distances are marked.

Table 5: Results of the *t*-test performed upon the feedback scores received for the approach behaviors without and with the IDMM

T-scores	Feedback scores with IDMM	Feedback scores without IDMM
Mean	8.56	5.53
Variance	1.99	2.28
dof	39	
<i>P</i>	3.97×10^{-10}	
<i>t</i>	1.685	

Table 4, respectively. In the beginning, we assumed that there is no significant difference between the feedback scores received for the robot's behavior with and without the IDMM. This becomes the null hypothesis. Table 5 gives the results of a *t*-test performed to test our null hypothesis. Test output is a *P* value of 3.97×10^{-10} (<0.05). Hence, the null hypothesis cannot be accepted with a confidence interval of 95%. Hence, we can conclude that there is a significant difference between the feedback scores received by the two systems. Looking at the figurative value of the feedback scores, we can say that the system with the IDMM received a higher score and resulted in a higher user satisfaction.

Facts observed during these occasions can be summarized as follows. Interactivity of a human strongly depends on the activity that the particular human is engaged in during that situation. As a clue of the activity as well as the emotional state of mind, posture changes were considered during this work. Using posture changes as behavioral measurements of human readiness toward interaction showed positive results toward interaction decision making by a robot. In some occasions where the feedback scores were average, the users preferred to speak more even though they were engaged. For instance, in the tasks such as cleaning or just doing nothing. In such tasks, recognition of the task becomes important. This is an implication derived from this study. Sometimes, they did not prefer to be distracted. Furthermore, certain individuals gave a low feedback score upon the robot's response, preferred more interaction with the robot. This was a personality trait and was not evaluated within the scope of this study. Therefore, it can be stated as another implication of this study. Other major implications we derived from this study are elaborated in Section 8.6.

User feedback was low in some occasions because it is not only the pose that contributes to a human's state. According to the behavior of the model it can be seen that postures such as sleeping and running could also be identified using this approach. We excluded such postures for the ease of participants in our experiments.

Besides, there should definitely be other responses expected by users from a robot. Therefore, in the future, integrating multiple responses from the robot is important, other than interactive distance and voice responses.

We implemented the algorithms for the right-hand side of the body. There are situations where we inherit different poses in our right and left hands. This fact holds true for the legs as well. Hence, we can adopt the same recognition techniques and then calculate valence of the entire body considering both sides of the body in the future.

8.6 Implications derived from the set of experiments

Implications derived from the set of experiments can be stated as follows. These implications could be utilized in the development of conceptual basis for response generation and the design of future social robots.

It can be seen from the experiments that an individual changes his/her behavior every once in a while. Therefore, it is important for robots to observe their users for at least a short duration before initiating an interaction with them. This can be proposed as the first design guideline derived from our experiments.

The implementation of IDMM received a higher feedback score when compared with a nonadaptive approach behavior used as the ground truth. Hence, considering the variables in the context such as postures could be advantageous for a robot in order to receive a higher user acceptance. Therefore, considering the factors within humans before making decisions can be stated as the second design guideline for intelligent robot assistants. Even though we considered only arm and body posture as a signal of internal state of human mind, there can be other factors that affect the internal state of mind. Therefore, a conceptual basis must be laid to identify such factors and use them for AI (Artificial Intelligence) systems in a theoretical manner. This can be stated as the third design guideline derived from these experiments.

There are some situations at which humans nonverbally seek attention from outside for various reasons. Therefore, unveiling the meanings of nonverbal behavior is an important aspect in present HRI. This can be considered the fourth design guideline proposed by these sets of experiments.

It can be seen that users preferred different approach behaviors at different occasions. The difference between the user feedback scores received for a user-aware approach behavior and static approach behavior testifies to this. Therefore, generating adaptive approach behavior which matches with the context can be proposed as the fifth design guideline derived from our experiments.

9 Conclusion

A method has been presented to recognize often encountered human postures using a simplified method rather than using hardware intensive complex mechanisms. The proposed method is based on the orientation of limbs and proved to be an efficient method to recognize a selected set of postures with a considerable accuracy. These postures cover a majority of domestic postures that are

encountered during daily chorus and this can be stated as an advantage of the proposed concept.

After recognizing user's arm and body postures, the robot goes through a logical argument about user's behavior before deciding upon an action. Prior to a direct conversation, the robot will be able to estimate the emotional state of its user and whether a conversation is appropriate in that particular situation. Hence, one major improvement of the system is the implementation of a novel vector-based approach for posture recognition using spatial orientation of limbs. Second, the utilization of a nonverbal mechanism for evaluation of user situation is an added advantage for a proactive robot. Over the existing systems, presented system has the ability to perceive user situation using uncountable features such as posture and postural changes. Furthermore, these features are deviated from often used cues to measure the user's attention such as facial expressions and emotional responses.

As a whole, a robot could be able to generate adaptive responses based on a user's postural behavior. An unadaptive approach behavior was compared with the user-aware approach behavior proposed in this article to evaluate the performance of the robot's adaptive approach behavior. Furthermore, the proposed system proved to be a convenient mechanism to identify a defined set of postures and to perceive a person's nonverbal demanding for an interaction, at a distance. This fact was confirmed by the experiments and the results presented. Hence, considering the variables in the context such as postures could be advantageous for a robot in order to receive a higher acceptance from users.

Acknowledgments: The authors are thankful to the participants who volunteered the study.

Funding information: The authors state that no funding is involved in conducting this research.

Conflict of interest: Authors state no conflict of interest.

Informed consent: Informed consent was obtained from all individuals included in this study.

Data availability statement: The datasets generated during the current study are available from the corresponding author on reasonable request.

References

- [1] G. Hoffman, O. Zuckerman, G. Hirschberger, M. Luria, and T. Shani Sherman, "Design and evaluation of a peripheral

- robotic conversation companion," in: *Proceedings of the 10th Annual ACM/IEEE International Conference on Human-Robot Interaction*, Portland, USA: ACM, 2015, pp. 3–10.
- [2] S. Satake, T. Kanda, D. F. Glas, M. Imai, H. Ishiguro, and N. Hagita, "How to approach humans? – Strategies for social robots to initiate interaction," in: *4th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, California USA. New York, United States: ACM, pp. 109–116, 2009.
- [3] C. Ø. Laursen, S. Pedersen, T. Merritt, and O. Caprani, "Robot-supported food experiences," in: *Cultural Robotics: First International Workshop, CR 2015, held as Part of IEEE RO-MAN 2015*, Lecture Notes in Computer Science, vol. 9549, Kobe, Japan: Springer, 2016, pp. 107–130.
- [4] S. Liu, L. Zheng, S. Wang, R. Li, and Y. Zhao, "Cognitive abilities of indoor cleaning robots," in: *12th World Congress on Intelligent Control and Automation (WCICA)*, Guilin, China: IEEE, 2016, pp. 1508–1513.
- [5] M. Coeckelbergh, C. Pop, R. Simut, A. Peca, S. Pintea, et al., "A survey of expectations about the role of robots in robot-assisted therapy for children with ASD: Ethical acceptability, trust, sociability, appearance, and attachment," *Science and Engineering Ethics*, vol. 22, no. 1, pp. 47–65, 2016.
- [6] J. Navarro and M. Karlins, *What Every Body is Saying*, New York, USA: William Morrow Paperbacks, 2008.
- [7] C. C. Clark, C. M. Barnes, G. Stratton, M. A. McNarry, K. A. Mackintosh, and H. D. Summers, "A review of emerging analytical techniques for objective physical activity measurement in humans," *Sports Medicine*, vol. 47, no. 3, pp. 439–447, 2017.
- [8] T. B. Moeslund and E. Granum, "A survey of computer vision-based human motion capture," *Computer Vision and Image Understanding*, vol. 81, no. 3, pp. 231–268, 2001.
- [9] C. Roossien, J. Stegenga, A. Hodseldmans, S. Spook, W. Koolhaas, et al., "Can a smart chair improve the sitting behavior of office workers?," *Applied Ergonomics*, vol. 65, pp. 355–361, 2017.
- [10] N. Hu, G. Englebienne, and B. Krose, "Posture recognition with a top-view camera," in: *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2013, pp. 2152–2157.
- [11] R. Cucchiara, A. Prati, and R. Vezzani, "Posture classification in a multi-camera indoor environment," in: *2005 IEEE International Conference on Image Processing*, vol. 1, 2005, p. I-725-8.
- [12] P. K. Pisharady and M. Saerbeck, "Kinect based body posture detection and recognition system," *International Conference on Graphic and Image Processing (ICGIP 2012)*, vol. 8768, International Society for Optics and Photonics, 2013, art. 87687F.
- [13] F. Buccolieri, C. Distanto, and A. Leone, "Human posture recognition using active contours and radial basis function neural network," in: *IEEE Conference on Advanced Video and Signal Based Surveillance*, 2005, pp. 213–218.
- [14] H. Foroughi, B. S. Aski, and H. Pourreza, "Intelligent video surveillance for monitoring fall detection of elderly in home environments," in: *11th International Conference on Computer and Information Technology*, 2008, pp. 219–224.
- [15] Y. Li, S. Ma, and H. Lu, "Human posture recognition using multi-scale morphological method and Kalman motion estimation," *Proceedings of 14th International Conference on Pattern Recognition (Cat. No. 98EX170)*, vol. 1, 1998, pp. 175–177.

- [16] N. Jacobs and A. Garnham, "The role of conversational hand gestures in a narrative task," *Journal of Memory and Language*, vol. 56, no. 2, pp. 291–303, 2007.
- [17] E. A. Cartmill, S. Beilock, and S. Goldin-Meadow, "A word in the hand: action, gesture and mental representation in humans and non-human primates," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 367, no. 1585, pp. 129–143, 2012.
- [18] V. Gentile, F. Milazzo, S. Sorce, A. Gentile, A. Augello, and G. Pilato, "Body gestures and spoken sentences: a novel approach for revealing user's emotions," in: *IEEE 11th International Conference on Semantic Computing (ICSC)*, San Diego, CA, USA: IEEE, 2017, pp. 69–72.
- [19] S. F. Chevtchenko, R. F. Vale, and V. Macario, "Multi-objective optimization for hand posture recognition," *Expert Systems with Applications*, vol. 92, pp. 170–181, 2018.
- [20] X. Yin and X. Zhu, "Hand posture recognition in gesture-based human–robot interaction," in: *1st IEEE Conference on Industrial Electronics and Applications*, Singapore: IEEE, 2006, pp. 1–6.
- [21] S. Malassiotis and M. G. Strintzis, "Real-time hand posture recognition using range data," *Image and Vision Computing*, vol. 26, no. 7, pp. 1027–1037, 2008.
- [22] S. Lee, I. H. Suh, and M. S. Kim, "Recent progress in robotics: viable robotic service to human: an edition of the selected papers from the 13th International Conference on Advanced Robotics (ICAR)," *Lecture Notes in Control and Information Sciences*, Vol. 370, Berlin: Springer, 2008.
- [23] H. S. Koppula and A. Saxena, "Anticipating human activities for reactive robotic response," in: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2013, p. 2071.
- [24] C. Sirithunge, A. Buddhika, P. Jayasekara, and D. Chandima, "Effect of activity space on detection of human activities by domestic service robots," *Region 10 Conference (TENCON 2017)*, Penang, Malaysia: IEEE, 2017, pp. 344–349.
- [25] S. Parekh and J. C. Lim, "Receptivity and interaction of social robots in hospitals," *ACM/IEEE International Conference on Human-Robot Interaction*, 2020, pp. 389–391.
- [26] Y. Xiao, Z. Zhang, A. Beck, J. Yuan, and D. Thalmann, "Human-robot interaction by understanding upper body gestures," *Presence*, vol. 23, no. 2, pp. 133–154, 2014.
- [27] M.-L. Bourguet, Y. Jin, Y. Shi, Y. Chen, L. Rincon-Ardila, and G. Venture, "Social robots that can sense and improve student engagement," in: *2020 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)*, Takamatsu, Japan: IEEE, 2020, pp. 127–134.
- [28] C. Matuszek, L. Bo, L. Zettlemoyer, and D. Fox, "Learning from unscripted deictic gesture and language for human–robot interactions," in: *Association for the Advancement of Artificial Intelligence (AAAI 2014)*, 2014, pp. 2556–2563.
- [29] Q. Chen, N. D. Georganas, and E. M. Petriu, "Real-time vision-based hand gesture recognition using haar-like features," in: *Instrumentation and Measurement Technology Conference Proceedings*, Piscataway, NJ: IEEE, 2007, pp. 1–6.
- [30] Z. Ren, J. Yuan, and Z. Zhang, "Robust hand gesture recognition based on finger-earth mover's distance with a commodity depth camera," in: *Proceedings of the 19th ACM International Conference on Multimedia*, New York, NY, United States: ACM, 2011, pp. 1093–1096.
- [31] C. Sirithunge, P. Arjuna, S. Srimal, A. Buddhika, and P. Jayasekara, "Identification of friendly and deictic gestures in a sequence using upper body skeletal information," in: *Region 10 Conference (TENCON 2017)*, Penang, Malaysia: IEEE, 2017, pp. 3051–3056.
- [32] S. Strohkorb and B. Scassellati, "Promoting Collaboration with Social Robots," in: *The 11th ACM/IEEE International Conference on Human Robot Interaction*, Christchurch, New Zealand, IEEE Press, 2016, pp. 639–640.
- [33] A. Hong, N. Lunscher, T. Hu, Y. Tsuboi, X. Zhang, et al., "A multimodal emotional human–robot interaction architecture for social robots engaged in bidirectional communication," *IEEE Transactions on Cybernetics*, vol. 51, no. 12, pp. 5954–5968, 2021.
- [34] M. P. Michalowski, S. Sabanovic, and R. Simmons, "A spatial model of engagement for a social robot," in: *9th IEEE International Workshop on Advanced Motion Control*, Istanbul, Turkey: IEEE, 2006, pp. 762–767.
- [35] C. H. Chen and P. S.-P. Wang, *Handbook of Pattern Recognition and Computer Vision*, 3rd edition, River Edge, NJ: World Scientific, 2005.
- [36] Z. Zhang, "Microsoft kinect sensor and its effect," *IEEE multimedia*, vol. 19, no. 2, pp. 4–10, 2012.
- [37] C. Sirithunge, M. A. Viraj, P. Jayasekara, and D. P. Chandima, "Interpretation of interaction demanding of a user based on nonverbal behavior in a domestic environment," *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, IEEE, 2017, pp. 1–8.
- [38] J. K. Burgoon and D. B. Buller, "Interpersonal deception: III. Effects of deceit on perceived communication and nonverbal behavior dynamics," *Journal of Nonverbal Behavior*, vol. 18, no. 2, pp. 155–184, 1994.
- [39] A. J. Gruber and R. J. McDonald, "Context, emotion, and the strategic pursuit of goals: interactions among multiple brain systems controlling motivated behavior," *Frontiers in Behavioral Neuroscience*, vol. 6, art. 50, 2012.
- [40] S. C. G. Tan and A. Nareyek, "Integrating facial, gesture, and posture emotion expression for a 3D virtual agent," *Proceedings of the 14th International Conference on Computer Games: AI, Animation, Mobile, Interactive Multimedia, Educational and Serious Games (CGames 2009 USA)*, 2009, pp. 23–31.
- [41] T. Fu and A. Macleod, "Intellichair: An approach for activity detection and prediction via posture analysis," in: *2014 International Conference on Intelligent Environments (IE)*, Rome, Italy: IEEE, 2014, pp. 211–213.