A Multilevel Body Motion-Based Human Activity Analysis Methodology

Kamrad Khoshhal Roudposhti, Jorge Dias, Paulo Peixoto, Vangelis Metsis, and Urbano Nunes

***Abstract*—Human body motion analysis is an initial procedure for understanding and perceiving human activities. A multilevel approach is proposed here for automatic human activity and social role identification. Different topics contribute to the development of the proposed approach, such as feature extraction, body motion description, and probabilistic modeling, all combined in a multilevel framework. The approach uses 3-D data extracted from a motion capture device. A Bayesian network technique is used to implement the framework. A mid-level body motion descriptor, using the Laban movement analysis system, is the core of the proposed framework. The mid-level descriptor links low-level features to higher levels of human activities, by providing a set of proper human motion-based features. This paper proposes a general framework which is applied in automatic estimation of human activities and behaviors, by exploring dependencies among different levels of feature spaces of the body movement.**

***Index Terms*—Bayesian programming (BP), human activity analysis, Laban movement analysis (LMA), multilevel framework.**

I. Introduction

HUMAN movements can be interpreted to recognize different human activities and behaviors. This paper proposes a multilevel framework for body motion-based human activity analysis. This paper addresses feature extraction, descriptive representation of human body motion, and probabilistic-based human activity analysis with different levels of complexity. Previous studies on automatic human activity analysis can be categorized as follows: 1) individual-based [19], [25], [26]; 2) human–object interaction (H-OI) [13], [27] and human-human interaction (H-HI) [24], [43], [49]; 3) context-based [46], [50]; and 4) social context-based [22], [44]. These studies lead us to conclude that there is a meaningful connection between the different levels of analysis. Thus, we propose a multilevel framework, which deals with all mentioned categories of human activity analysis. The approach is built according to a bottom-up structure, where each level uses outputs from the level preceding it. In the design and implementation of the proposed methodology it is required to define appropriate feature spaces and to use an adequate technique for modeling dependencies among different feature spaces in different layers.

A. Related Work

Automatic monitoring of human activities has application in areas such as surveillance, eldercare, social behavior analysis, and human–robot interaction (HRI). A significant amount of research has been published regarding human activity analysis using body motion-based features. Table I lists a set of relevant research works that address different body motion-based analysis approaches for different levels of human activities.

The lowest level of human motion analysis does not really recognize any type of human activities, but simply identifies motion primitives for specific application contexts, such as: rehabilitation [14], HRI [60], and choreography [55]. These motion primitives can be categorized as: simple context-free actions (*Walking*, *Running*, *Rising*, *Sitting*, etc.), gestures, and sign languages. Each one of these types of motion primitives was solved with respect to different applications: gesture for HRI [40], sign language [23], [53], and simple movements for surveillance [25].

In our daily activities, interaction with objects is very common, so researchers have proposed methods able to perceive this interaction. Up to now this was done always in the context of specific application scenarios, such as: eldercare [13], [17], [20], surveillance [50], scene understanding [27], [46], [58], and HRI [24]. Some of these studies solved the problem by using individual human simple activity level information [17], [46], [50], [58], while others recognized H-OI without individual activity level of analysis.

At the same level as H-OI we can have H-HI analysis, which is a much more complex problem since there are at least two people involved in the analysis. Ryoo and Aggarwal [49]

Manuscript received January 30, 2016; revised May 29, 2016; accepted September 1, 2016. Date of publication September 8, 2016; date of current version March 9, 2017. This work was supported in part by the Institute of Systems and Robotics, Portugal, Khalifa University, Abu Dhabi, UAE, and Texas State University, and in part by the FCT under Project AMS-HMI12: RECI/EEIAUT/0181/2012, cofounded by COMPETE.

K. Khoshhal Roudposhti is with the Department of Computer Engineering, Lahijan Branch, Islamic Azad University, Lahijan, Iran, with the Institute of Systems and Robotics, Department of Electrical and Computer Engineering, University of Coimbra, Portugal, and also with the Department of Computer Science, Texas State University, San Marcos, TX 78666 USA (e-mail: kamrad@txstate.edu).

J. Dias is with the Institute of Systems and Robotics, Department of Electrical and Computer Engineering, University of Coimbra, 3000-214 Coimbra, Portugal, and also with Khalifa University, Abu Dhabi 127788, UAE (e-mail: jorge@isr.uc.pt).

P. Peixoto and U. Nunes are with the Institute of Systems and Robotics, Department of Electrical and Computer Engineering, University of Coimbra, 3000-214 Coimbra, Portugal (e-mail: peixoto@isr.uc.pt; urbano@deec.uc.pt).

V. Metsis is with the Department of Computer Science, Texas State University, San Marcos, TX 78666 USA (e-mail: vmetsis@txstate.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCDS.2016.2607154

addressed H-HI by analyzing individual lower level activities. A similar approach was followed in [46], but considering a 3-D-based analysis. Kelley *et al.* [24] did this analysis without having individual human activity estimation. The highest mentioned level of analysis in this paper is human behavior analysis in a social context. Jayagopi *et al.* [22] estimated social behavior using low-level features (LLFs) extracted from sequence of images of human movements. Roudposhti *et al.* [47] used human primitive motion-based information of people in a mid-level analysis to estimate social behaviors.

The most recent trend in the literature is to analyze different levels of human activities in a single model. One of the reasons for this is the existence of meaningful dependencies between features of different levels of analysis that can be connected in a bottom-up fashion [46]. Outputs of each level can be used as inputs for higher level analysis. However, each level has its own informative description for a specific application.

B. Contribution

The main contribution of this paper is the proposal of a generalized framework for analyzing different types of human activities. In contrast to other works, our approach deals with all the mentioned levels of analysis. We studied a few relevant issues with respect to this framework, addressing feature extraction, human activity, and social behavior modeling. We explain how the body-motion-based features can be analyzed, and we propose a multilevel approach able to deal with the above mentioned levels of human activities.

The idea of having a generalized framework for human activity analysis came from our previous studies on modeling different types of human activity: simple human movement analysis [25], human activity analysis (H-OI and H-HI) [43], and social behavior analysis [47]. Based on these previous studies, we propose the multilevel human activity analysis framework shown in Fig. 1. It is composed of the following main modules:

1) LLFs extraction (e.g., body parts acceleration signals);
2) laban movement analysis (LMA) parameters estimation (a mid-level analysis introduced in Section II-A), (e.g., body parts motion characteristics);
3) human movement estimation (e.g., *Walking* and *Sitting*);
4) interpersonal behaviors (IBs) estimation (e.g., the person is interested or active in a conversation);
5) activity estimation (e.g., *Reaching* and *Hand Shaking*);
6) social role (SR) estimation (e.g., he/she is a leader or not in a conversation).

The remaining of this paper is organized as follows. Section II presents some background materials and methods necessary to understand the proposed framework, while Section III provides a case study that introduces the used experimental setup. Section IV presents the used body motion-based feature extraction methods. The use of the Bayesian programming (BP) formalism (introduced in Section II-B) for modeling the multilevel human activity analysis framework is described in Section V. Finally, Section VI presents some experimental results obtained using the proposed framework.

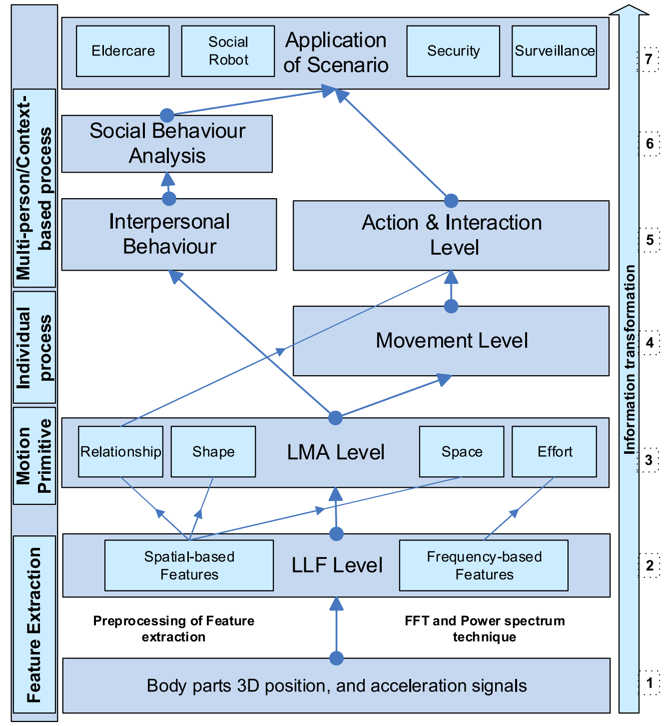


Fig. 1. Diagram of the proposed multilevel framework for human activity analysis that shows the different level of human movement information dependencies in a bottom-up strategy.

II. Background Material

This section presents a brief introduction to two powerful techniques used in the formulation of our proposed multilevel approach: 1) the LMA method, a well-known human movement descriptor and 2) BP.

A. Laban Movement Analysis

The LMA method was proposed by Hutchinson [21] for describing, annotating and interpreting human movement in the field of choreography. The LMA method defines five components: 1) body; 2) space; 3) effort; 4) shape; and 5) relationship. These components can be used for analyzing motion-based human activities, such as communicative gestures [41], [63], individual movements [25], and human action and interaction [43]. *Space* and *Shape* components provide interpretation of human movement in the spatial domain. *Space* describes body parts trajectories in a 3-D reference plane, while *Shape* describes the whole human body (as a blob) deformation in three planes (frontal, horizontal, and sagittal). The *Effort* component embeds the dynamic quality of the human movement, the feeling tone, the texture, and how the energy is being used. For example, it can describe if the motion of a human body part is sudden or sustained (*Effort-Time*), light or strong (*Effort-Weight*), direct or indirect (*Effort-Space*), and bound or free (*Effort-Flow*). The *Relationship* component defines the existent connections between human body and the environment or scene [21]. We will present how this component plays an important

TABLE I
LIST OF RELATED WORK WITH RESPECT TO THEIR LEVEL OF ANALYSIS IN HUMAN-MOTION-BASED ACTIVITIES

Ref.	Different levels of human activity analysis				Description	
	Individual	Multi-person/Context-based				
	Simple movement	Action/interaction (H-OI)	Action/interaction (H-HI)	Social behavior	Motion primitive	Application
[14]	No	No	No	No	LMA	Rehabilitation
[60]	No	No	No	No	Yes	HRI
[55]	No	No	No	No	LMA	Choreography
[40]	Gesture	No	No	No	LMA	HRI
[61]	Gesture	No	No	No	No	Gesture rec.
[23]	Sign language	No	No	No	Phonetic	Sign language rec.
[53]	Sign language	No	No	No	No	Sign language rec.
[25]	Yes	No	No	No	LMA	Human action rec.
[17]	Yes	Yes	No	No	No	Eldercare
[50]	Yes	Yes	No	No	No	Human action rec.
[27]	No	Yes	No	No	No	Scene understanding
[20]	No	Yes	No	No	No	Eldercare
[13]	No	Yes	No	No	No	Human action rec.
[46]	Yes	Yes	Yes	No	LMA	Human action rec.
[58]	Yes	Yes	No	No	No	Scene understanding
[49]	Yes	No	Yes	No	Yes	Human action rec.
[24]	No	Yes	Yes	No	No	HRI
[22]	No	No	No	Yes	No	Social group analysis
[47]	No	No	No	Yes	LMA	Social group analysis
Our model	Yes	Yes	Yes	Yes	LMA	Human activity rec.

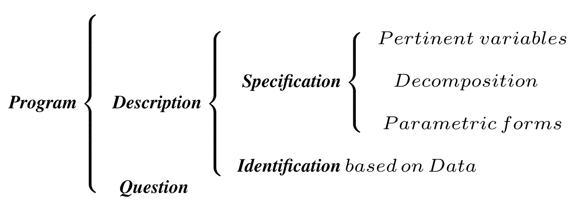


Fig. 2. Structure of a Bayesian program, as described in [5].

role to model the connections of the human body with its surroundings, for estimating higher level information.

B. Bayesian Programming

BP is a methodology proposed by Bessière *et al.* [5], which allows the formulation of probabilistic models in a consistent and easy to interpret manner. Bessière *et al.* [5] showed the capability of the BP methodology to specify different probabilistic models. Fig. 2 presents the BP formalism structure, which consists of the following two main parts.

1) *Description:* Declarative part to present a joint distribution with a set of variables, given previous knowledge and experimental data. It is divided into the following sections:
   a) *specification:* preliminary knowledge:
      i) *pertinent variables:* relevant variables defined in the joint distribution;
      ii) *decomposition:* a probability distribution;
      iii) *forms:* either parametric forms or questions to other Bayesian programs;
   b) *identification:* explains about the dataset;
2) *question:* for estimating a probability distribution of the form $P(Searched \mid Known)$.

For more details about the technique, the reader may refer to [5].

III. CASE STUDY

Real world intelligent systems, like social robots, need to analyze different types of human activities, assuming a natural HRI. A methodology that address jointly all human activity levels, mentioned in Table I, is not found in the literature. In this paper, we contribute with a new methodology (with roots in our previous works [25], [46], [47]) to deal with a rich ensemble of human motion-based activities. The different types of activities are categorized in a set of complexity levels: LMA as a body motion descriptor (Section II-A), human individual movement (Section V-A), human activity (H-OI and H-HI) (Section V-B), IB, and SR (Section V-C).

Fig. 1 shows a general structure and information processing flow of the proposed multilevel human activity analysis framework, and Fig. 3 presents the framework for our case study, in which two persons are involved, highlighting the different levels of feature extraction and estimation modules. The LMA method, used at the mid-level of analysis, parametrizes body motions with a set of descriptions and connects the LLFs level to higher layers of analysis. LMA is a system that provides proper body-motion-based features for the higher level

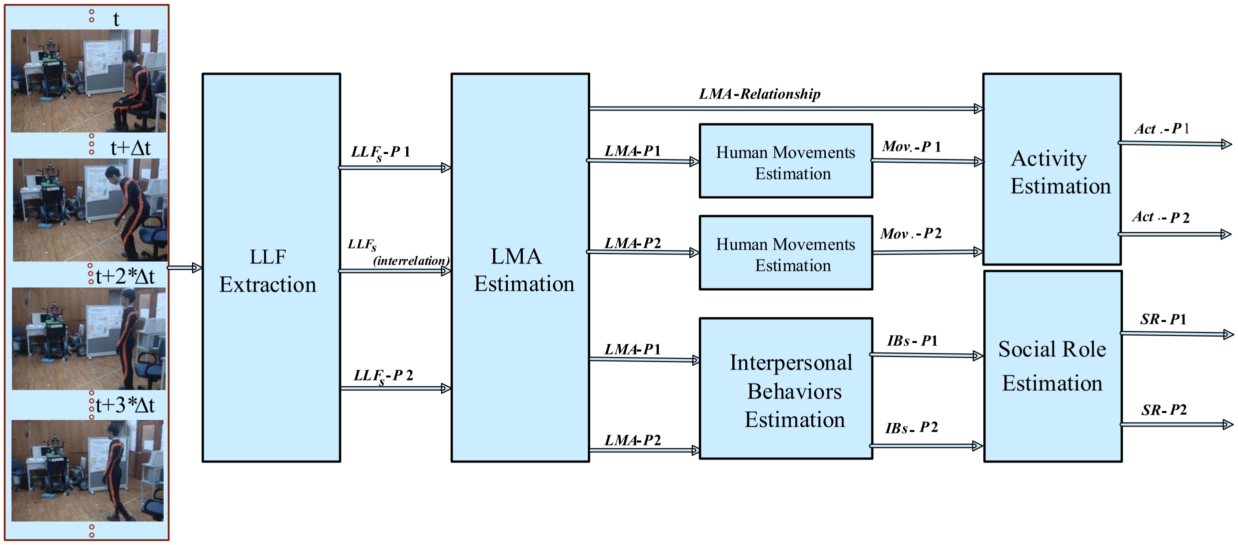


Fig. 3. Flow of information processing for the proposed framework for body motion-based human activity analysis ($P1$ and $P2$ denote the first and second person).

of analysis, which minimizes redundancy. The LLFs extraction methods are defined based on the LMA components. In Section IV, first, the LMA components are described, then, we explain the feature extraction methods for each of them. With respect to Fig. 3, each level of human activity is analyzed, modeled and implemented step-by-step. The type of movements and activities addressed in each of the mentioned levels are different. Thus, separate datasets for each of the mentioned levels of analysis, are recorded for learning/training and testing process.

A. Experimental Setup

In our experiments, an Xsens motion tracking suit (wearable sensor) has been used to collect skeleton tracking information of human body (see Fig. 4). The suit contains 17 inertial measurement units (IMUs) and comes with a specialized software which reconstructs motion from IMU measurements and provides 3-D trajectories of body parts motion. The output of the motion tracker is an XML file for each recording session. The XML file contains the 3-D position, velocity, acceleration, angular velocity, and angular acceleration of human body parts (head, hands, and feet) for a maximum of 120 frames/s. For each layer of our multilevel approach, specific experiments were performed with the aim of analyzing the performance of each level of analysis separately. A sliding window technique is used in the feature extraction process. The dataset used in our experiments (Fig. 4), which includes data regarding different human activities performed by different actors, is publicly available under the designation of 3-D-UC motion dataset [1].

IV. Body Motion-Based Feature Extraction

Human body motion analysis is a complex task given the diversity and dynamics of body movements. To analyze and understand human movements, we need not only to have data regarding the trajectory $\mathbf{p}^{\mathrm{bp}}(t)$ of each body part bp in time t, but also data regarding velocity $v^{\mathrm{bp}}(t) \equiv \dot{\mathbf{p}}^{\mathrm{bp}}(t)$, acceleration

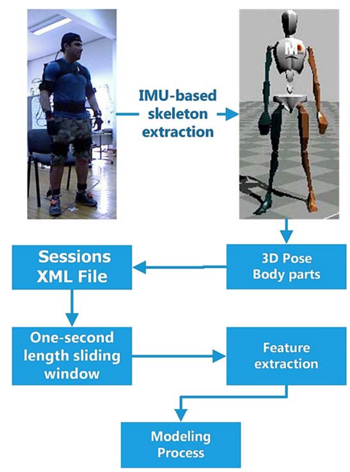


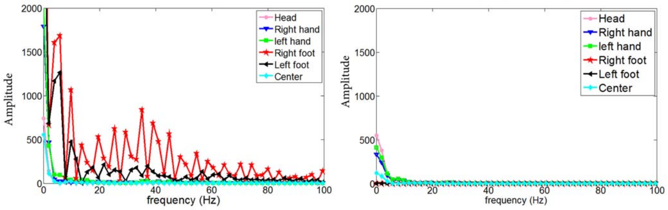
Fig. 4. Diagram of our motion capture setup, using the Xsens motion capture system. The diagram also shows the steps from capturing the motion to modeling it using our methodology for activity recognition.

$a^{\mathrm{bp}}(t) \equiv \ddot{\mathbf{p}}^{\mathrm{bp}}(t)$, frequency contents $F(a^{\mathrm{bp}}(t))$, and interrelation between body parts motions $R(\mathbf{p}_i^{\mathrm{bp}}(t), \mathbf{p}_j^{\mathrm{bp}}(t))$ (where i and j denote indexes of different body parts). Taking as an example the movements of punching and hand pointing, the corresponding hand's trajectories are similar but their timing and energy profile are significantly different.

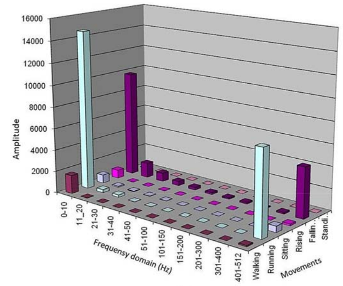
LMA is used to describe human movement for extracting proper features of human movements for various applications. The proper LLFs for estimating LMA components, can be divided into frequency and spatial-based features; frequency-based features are used for the *Effort* component while spatial features are used for the *Shape* and *Relationship* components [25], [43].

A. Frequency Domain

The *Effort* component of LMA is the component that is characterized in the frequency domain [25]. *Effort* deals with



(a)



(b)

Fig. 5. (a) Two PSD signals obtained from the 3-D acceleration signals of the five parts of the body for walking and rising movements. (b) Histogram for a single body part movement, showing the definition of some frequency sub-bands on the PSD signals for six different individual actions: *Walking, Running, Sitting, Rising, Falling down, and Standing*.

the dynamics of human motion, representing how a movement is performed by the person. For instance, the *Effort* distinguishes movements that are similar in space, like punching and pointing movements. In this case, the difference resides in the dynamics of the movements. The *Effort*, as described by the LMA, has different factors, with two bi-polar states for each of those factors. *Effort.time* is one of the factors, which is analyzed in the frequency domain, based on acceleration signals of body motions in 3-D space.

For a sequence of N_s acceleration samples $\{a^{\mathrm{bp}}[n]\}_{N_s}$ of a continuous acceleration signal $a(t)$, an estimate of its power spectral density (PSD) can be obtained by calculating its periodogram

$$\hat{P}_x\left(e^{j\omega}\right) = \frac{1}{N_s}\left|X\left(e^{j\omega}\right)\right|^2 \tag{1}$$

where $\omega = 2\pi f$ rad/s, and

$$X\left(e^{j\omega}\right) = \sum_{n=0}^{N_s-1} a^{\mathrm{bp}}[n]e^{-j\omega n} \tag{2}$$

is the Fourier transform of that data sequence.

In practice, a fast Fourier transform (FFT) algorithm is used to compute (2). We use an approach proposed in [26], which divides the frequency domain of the PSD signal into a set of frequency bands as shown in Fig. 5(b). To define the feature for a specific movement and body part we consider the peak values of the first four bands. This feature will be used to estimate the *Effort* component for each of the human movements of interest.

B. Spatial Domain

In this domain, we are looking for two kinds of features: 1) environment-dependent and 2) environment-independent features. Environment independent features are used for analyzing human movement, while the environment dependent ones are needed to explore human action and interaction in a scene. LMA defines components for the both categories. Regarding environment-independent features, *Shape* and *Space* LMA components are used, while for the environment-dependent ones the LMA *Relationship* component is used [43].

1) Shape Component: The *Shape* component characterizes the deformation of the human body in the frontal, horizontal, and sagittal planes. The qualities used to describe the deformation of human shape on each one of the three planes are

$$\mathrm{Plane}_{\mathrm{Frontal}} \in \{\mathrm{Sinking, Still, Rising}\}$$
$$\mathrm{Plane}_{\mathrm{Horizontal}} \in \{\mathrm{Enclosing, Still, Spreading}\}$$
$$\mathrm{Plane}_{\mathrm{Sagittal}} \in \{\mathrm{Retreating, Still, Advancing}\}.$$

The change in shape in the frontal plane is modeled by the following equation:

$$\Delta H = \sum_{t=2}^{N}\left(\left({}^{z}p_{(t)}^{\mathrm{head}} + {}^{z}p_{(t)}^{\mathrm{feet}}\right) - \left({}^{z}p_{(t-1)}^{\mathrm{head}} + {}^{z}p_{(t-1)}^{\mathrm{feet}}\right)\right) \tag{3}$$

where ${}^{z}p_{(t)}^{\mathrm{head}}$ and ${}^{z}p_{(t)}^{\mathrm{feet}}$ denote positions of head and feet (the lower foot) in the z-axis at time t, respectively. ΔH describes the variation in height during head movement, and N denotes the total number of samples contained in the sliding window. ΔH is used to estimate *Shape.frontal* S^v qualities.

2) Relationship Component: For context-based features, two parameters of the *Relationship* component were applied (for details see [43]): *Toward* and *Away* and *Contact* or *Touch.*

a) Toward and away relationship: A performer may gesture toward or away from a part of his body, another person, an object, or a part of the place [21]. The *Toward Relationship* has three states: 1) *Positive*; 2) *Still*; and 3) *Negative*. These states are used to represent the iteration status of a person with an object or another person in a scene (ΔD).

b) Contact or touch relationship: When a part of body is active in producing a touch or contact to another part, an object, or another person [21]. Some activities involve a contact between two entities (e.g., a body-part and an object), for example: grasping a glass, handshaking, pushing or kicking an object, etc. *Contact* qualities can be estimated using the 3-D distance between the two entities (body-part and/or objects). This relationship assumes two possible states: 1) *Connected* and 2) *Disconnected*, which can be inferred using a threshold on the difference of distance between two entities [43].

V. Human Activity Modeling

Human activity depends on the context, scenario or environment, where it is being performed. Human activity has different levels of complexity:

1) body parts motion such as gestures and sign language;
2) movements such as *Walking*, *Running*, and *Sitting*;
3) actions such as *Reaching*, *Following*, and *Spreading*;
4) H-OI such as *Grasping* and *Touching*;
5) H-HI such as *Hand Shaking* and *Punching*;
6) social behavior such as *Leading* and *Emphasizing*.

In this paper, we propose a bottom-up strategy through a multilevel framework to estimate each one of the mentioned types of activities, utilizing the well known advantages of Bayesian networks (BNs) [43], [45]. BP is used [5], [28] to model the features of interest and their dependencies on each level of the BN. The approach followed will be briefly explained in Section II-B.

A. Human Activity Analysis—Individual Case

Human movement understanding is an important key for analyzing human activity. The difference between a human motion (movement) and an action is that the action needs a context or previous knowledge, and there is a relation between human movement and the environmental parameters, while a movement is free of context and of the human-environment mentioned relation [6]. Since, the contextual information is not relevant in the individual case, the positions of the human body parts are the only input variables to our approach. Depending on the application, information relating to certain parts of the body is more important than information from other body-parts. For example, for gait recognition, leg motion is the relevant cue; for sign language recognition, hands are the relevant body part; while in body gestures, several body parts are involved, especially head and hands.

More than a couple of decades ago, a considerable amount of work focused on human motion analysis [2], [7], [33], [34], [39], [59]. Gavrila [15] presented different methodologies that categorized 2-D and 3-D approaches using or not using an explicit model of shape. His work pointed to a couple of directions of related research regarding tracking in 3-D and action recognition. LLFs can be extracted from 2-D (image domain) or 3-D data.

In general, specific applications have been addressed and so specific features have been considered in published research studies. For instance, different pieces of information from a sequence of images are used for detecting a set of outdoor activities in [30], a set of indoor human behaviors [4], and a set of human activities in a shopping space [37]. The mentioned studies tried to extract the features needed for their specific purposes, which limits the generalization of their model. Thus, we have been used LMA components as mid-level information in our framework, that can provide generic features for higher level analysis in different applications [9], [42], [63]. Based on the previous studies [26], [42], and after examining different existing data types (trajectory, velocity, acceleration, etc.), acceleration signals were selected as the data source of the LLFs extraction, as described in Section IV.

Next, our approach for human movement modeling and corresponding experimental results are presented.

Step 1 (Modeling): Since, it is supposed to use different components of LMA for the modeling process, a flexible modeling approach such as BN is most suitable for preparing a human motion framework to fuse the various features in various domains (frequency and spatial) into a single model. The three abstract levels (LLFs, LMA, and human movement levels) and the dependencies are modeled in a BN framework, as can be seen in Fig. 6. In LMA layer of the framework, two independent parallel BNs for *Effort* and *Shape* components are provided, which are fused in the next layer to estimate human movements.

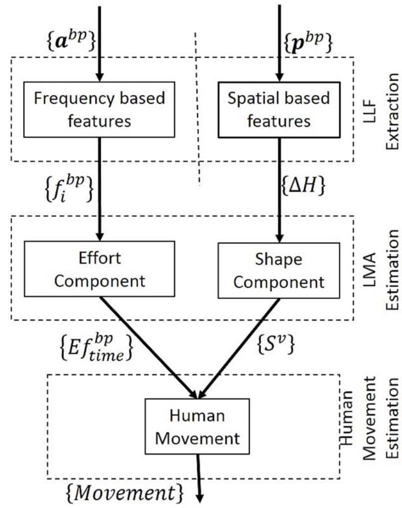


Fig. 6. Human movement analysis diagram. It presents the three layers: LLF, LMA, and human movement and their dependencies.

FFT-based features (f_i^{bp}), are discretized into four states {No, Low, Medium, and High}, as defined in [25]. *Effort-Time* subcomponent $\{Ef_{\text{time}}^{\text{bp}}\}$, for each body part, has two states: 1) *Sustained* and 2) *Sudden*. $\triangle H$ (3) is categorized into three states {Up, Still, and Down} [43].

The Bayesian model consists on estimating values for $P(\text{LMA}|\text{LLFs})$ (probability of the parameters of LMA given the observed features), and $P(M|\text{LMA})$ (probability of a movement given the LMA parameters) [25]. Fig. 7 presents the details of the Bayesian program corresponding to the proposed human movement model.

Step 2 (Experimental Results): Data regarding predefined movements was used to evaluate the proposed method (Table II). The dataset [1], which contains ten samples for each type of human movement (*Walking*, *Running*, *Sitting*, *Rising*, *Falling*, and *Standing*) in different durations, was used in the evaluation process. The data (body parts position and acceleration signals) was recorded using a motion tracker suit. The recorded data was segmented using a sliding window approach. The size of the window is one second, and there is a half second overlap between consecutive windows. Table II summarizes the results obtained using the proposed framework.

B. Human Activity Analysis—Context-Based

This section analyzes human activities by exploring the relations between a human movement and the environment. The proposed features related to human movements (trajectory, velocity, acceleration, etc.) have been used for human movement and activity analysis, but the interfeatures relationship is the factor for analyzing different types of human actions and interactions. The relationship can be categorized

Program / *Description* / *Specification*

Pertinent variables :

$$Pb \in \{Head, Left\,hand, Right\,hand, Left\,foot, Right\,foot\}$$

$$f_{i=1:4}^{p\in\{bp\}} \in \{No, Low, Medium, High\}$$

$$\triangle H \in \{Up, Still, Down\}$$

$$Ef_{time}^{p\in\{bp\}} \in \{Sustained, Sudden\}$$

$$S^{v} \in \{Sinking, Still, Rising\}$$

$$Movement \in \{Walk, Run, Sit, Rise, Fall, Stand\}$$

Decomposition :

$$\left(P\left(Movement \prod_{p\in\{bp\}} Ef_{time}^{p} \prod_{p\in\{bp\}}^{i=1:4} f_i^p \,\triangle H\, S^v\right) = \prod_{p\in\{bp\}}^{i=1:4} \left(P\left(f_i^p\right)\right) \prod_{p\in\{bp\}} \left(P\left(\prod^{i=1:4}\left(f_i^p\right)\middle| Ef_{time}^{p}\right)\right) P\left(\triangle H\right) P\left(\triangle H \mid S^v\right) P\left(\prod_{p\in\{bp\}}\left(Ef_{time}^{p}\right) S^v \mid Movement\right)\right)$$

Parametric forms :

$$P\left(f_{i=1:4}^{p\in\{bp\}}\right) : uniform$$

$$P\left(f_{i=1:4}^{p\in\{bp\}} \middle| Ef_{time}^{p\in\{bp\}}\right) : histogram(Matrix)$$

$$P\left(\triangle H\right) : uniform$$

$$P\left(\triangle H \mid S^v\right) : histogram(Matrix)$$

$$P\left(Ef_{time}^{p\in\{bp\}}\, S^v \mid Movement\right) : histogram(Matrix)$$

Iden.

Ques. : $P\left(Movement \middle| \prod_{p\in\{bp\}}^{i=1:4}\left(f_i^p\right) \triangle H\right)$

Fig. 7. Bayesian program for the human movement model.

TABLE II
CLASSIFICATION RESULT USING BODY-CENTERED-BASED FEATURES IN BOTH FREQUENCY AND SPATIAL DOMAINS(*Effort* AND *Shape*)

Movement types	%
Walking	100
Running	96.77
Sitting & Bending	97.6
Rising	100
Falling	95.7
Standing	100

in three levels: 1) between body parts; 2) human and environmental objects; and 3) human with human interaction. For instance, in a *Walking* action, there is a relation between legs and hands movements which is very important for performing the action. In another example, when clapping hands, there is a relation of proximity of two body parts; in this case the two hands. This notion also occurs between a body and an external object, such as *Reaching* to or *Sitting* on a chair, and also between two people, such as *Punching*, *Hand Shaking*, etc. In this paper, the mentioned relationships are analyzed for different human activities. LMA proposes a valid component for analyzing those relationships. Modeling those relationships, is a big challenge that our approach attempts to tackle.

1) Context-Based Analysis: Human activity analysis can be categorized as context-free-based and context-based. In context-free-based approaches the model is independent of scene parameters, and just relies on the features belonging to the person. However, in reality, context-based features play a very important role in analyzing human activities. For example, when a person is approaching a chair, the likelihood that the person intends to sit on the chair is high.

As Delaitre *et al.* [10] described since object detection is a widely studied topic in computer vision, analyzing the relation between human movements and the existent objects around, can produce valuable information for human daily activities.

The question we can pose is: which level of human movement information might be useful, and how can a general framework be defined for analyzing any possible H-OIs. In this context, both the lowest level information such as body parts motions and the higher ones such as human interactions can be useful. A complex model is needed to deal with all this information. Thus, hierarchical frameworks have been proposed aiming to deal with the inherent complexity of the models and provide integrated and scalable solutions [3], [46].

Gupta *et al.* [16] tackled the H-OI problem, based on 2-D images. They were focused more on the computer vision problems raised by the mentioned applications, and used human hand trajectory information to analyze H-OIs (reaching and manipulation). Their proposed Bayesian model cannot deal easily with other types of activities. Thus, we propose a hierarchical model to deal with the problem. There are a number of works using 3-D-based human movement analysis [8], [31], and also in 3-D virtual applications [12], achieving good performance, but they only focus on simple human movements classification. The LMA system is used in [43], to define proper human motion patterns and human-scene relations (*Relationship*) to analyze human activities. Next, we present our approach for context-dependent human activity modeling, and our experimental results when applying it to classify different activity types.

Step 1 (Modeling): The new level of analysis (human activity or interaction), is built on top of the previous model which was designed for human individual movement analysis. The Bayesian graphical model for the proposed system is illustrated in Fig. 8, which presents the dependencies between the different levels of information.

In the highest level, we estimate the probability of each human activity, given the movement states probabilities of both persons $p1$ and $p2$, and the relation between person $p1$ and the two defined objects ($p2$ which is other person and $o1$ which is a chair). Equation (4) presents the Bayesian rule which models these dependencies

$$P\big(I_{p1}|M_{p1}, M_{p2}, R_{p1-p2}, R_{p1-o1}\big) = \frac{P\big(I_{p1}\big)P\big(M_{p1}|I_{p1}\big)P\big(M_{p2}|I_{p1}\big)P\big(R_{p1-p2}|I_{p1}\big)P\big(R_{p1-o1}|I_{p1}\big)}{P\big(M_{p1}\big)P\big(M_{p2}\big)P\big(R_{p1-p2}\big)P\big(R_{p1-o1}\big)} \tag{4}$$

where I_{p1}, M_{p1}, and R_{p1-o1} denote person $p1$'s *Activity*, *Movement*, and *Relationship* with respect to object $o1$, variables, respectively. $P(M_{p1}|I_{p1})$ denotes the estimation of *Movement*'s states of person $p1$ probabilities given the probability of its I_{p1} states.

In (4), the variables which are considered as inputs: M_{p1}, M_{p2}, R_{p1-p2}, and R_{p1-o1} are estimated in the previous level of analysis. For instance, (5), which was proposed in the previous

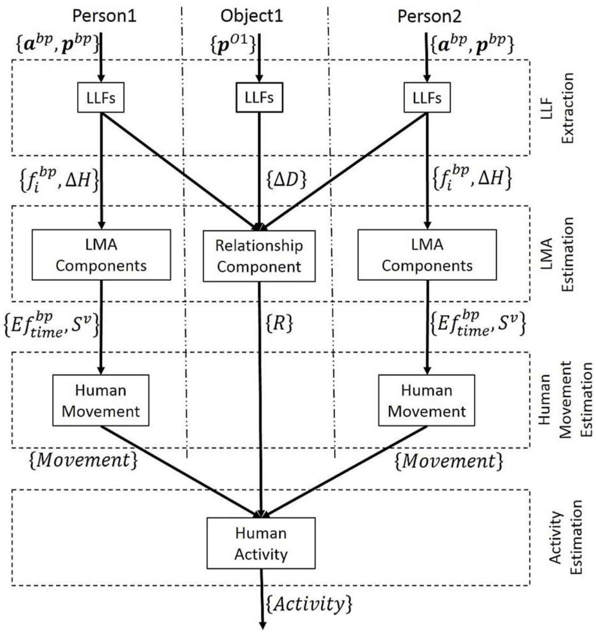


Fig. 8. Hierarchical framework for scene understanding. There are frequency-based and spatial-based features in LLF's level. LMA's level contains *Effort* and *Shape* components for people and the *Relationship* component. In the movement level, there are two movement classes belonging to person 1 and 2 given their corresponding *Effort* and *Shape* components, respectively. Finally, activity states are estimated given both person 1 and 2 movement classes and the existent *Relationship* states $\{R\}$ between people and object ($o1$) probabilities.

level of analysis, was used to model human movement

$$P\left(M_{p1}|Ef_{p1}^{\mathrm{bp}}S_{p1}^{v}\right) = \frac{P(M_{p1})P\left(Ef_{p1}^{\mathrm{bp}}|M_{p1}\right)P\left(S_{p1}^{v}|M_{p1}\right)}{P\left(Ef_{p1}^{\mathrm{bp}}\right)P\left(S_{p1}^{v}\right)} \tag{5}$$

where Ef_{p1}^{bp} and S_{p1}^{v} denote the *Effort* component of LMA for bp (body part) of person $p1$, and the *Shape* component of LMA for person $p1$ in frontal plane, respectively. bp is the index of body parts which are used (hands, feet, and head). Finally, $P(Ef_{p1}^{\mathrm{bp}})$ and $P(S_{p1}^{v})$ denote the probability of *Effort* and *Shape* components of LMA, respectively. Fig. 9 presents the details of the Bayesian program corresponding to the proposed human activity model.

Step 2 (Experimental Results): The motion tracker provides body parts positions, with respect to a global reference set of positions which is obtained in the sensors calibration step. Based on that global reference other objects of interest positions in the scene, are estimated (see Fig. 10). Ten different sequences (each sequence contains more than 1000 frames) are collected for each type of human movement, which performed within the context of different defined actions and interactions. The actions and interactions, which are listed in Table III, are defined in a scenario which contains two persons and a chair. For instance, a *Reaching* action can occur when a person is walking to reach the second person or the chair. However, a *Sitting* action can only occur when a person is going to sit on a chair. Thus, the activities that can be related to H-OI are *Reaching*, *Spreading*, *Sitting*, and *Standing,* and to H-HI are *Reaching*, *Spreading*, *Hand-Shaking*, and *Pushing*.

Program / *Description* / *Specification*

Pertinent variables :

$Pb \in \{Head, hands, feet\}$

$O \in \{p1-o1, p1-p2\}$

$\triangle D^t \in \{Positive, Still, Negative\}$

$\triangle D^c \in \{Positive, Negative\}$

$M_{p1,p2} \in \{Walk, Run, Sit, Rise, Fall, Stand\}$

$R^t_{(p1-p2),(p1-o1)} \in \{toward, still, away\}$

$R^c_{(p1-p2),(p1-o1)} \in \{connected, disconnected\}$

$I_{p1} \in \{Reach, Spread, Sit, Stand, Hshk, Push, Other\}$

Decomposition :

$$P\left(I_{p1} M_{p1} M_{p2} R^t_{(p1-o1)} R^t_{(p1-p2)} R^c_{(p1-o1)} R^c_{(p1-p2)}\right)$$
$$= P\left(I_{p1}\right) P\left(M_{p1} \mid I_{p1}\right) P\left(M_{p2} \mid I_{p1}\right)$$
$$P\left(R^t_{(p1-o1)}|I_{p1}\right) P\left(R^t_{(p1-p2)}|I_{p1}\right)$$
$$P\left(R^c_{(p1-o1)}|I_{p1}\right) P\left(R^c_{(p1-p2)}|I_{p1}\right)$$
$$P\left(M_{p1}\right) P\left(M_{p2}\right) P\left(R^t_{(p1-o1)}\right)$$
$$P\left(R^t_{(p1-p2)}\right) P\left(R^c_{(p1-o1)}\right) P\left(R^c_{(p1-p2)}\right)$$

Parametric forms :

$P\left(M_{p1,p2}\right) : BP$

$P\left(I_{p1}\right) : uniform$

$P\left(M_{p1,p2} \mid I_{p1}\right) : histogram(Matrix)$

$P\left(R^t_{(p1-o1),(p1-p2)}|I_{p1}\right) : histogram(Matrix)$

$P\left(R^c_{(p1-o1)}|I_{p1}\right) : histogram(Matrix)$

$P\left(R^t_{(p1-o1),(p1-p2)}\right) : uniform$

$P\left(R^c_{(p1-o1),(p1-p2)}\right) : uniform$

$$\left(P\left(R^t_O \mid \triangle D^t_O\right) = \frac{P\left(R^t_O\right)P\left(\triangle D^t_O|R^t_O\right)}{P\left(\triangle D^t_O\right)}\right)_O$$

$$\left(P\left(R^c_O \mid \triangle D^c_O\right) = \frac{P\left(R^c_O\right)P\left(\triangle D^c_O|R^c_O\right)}{P\left(\triangle D^c_O\right)}\right)_O$$

$\left(P\left(\triangle D^t_O\right)\right)_O : uniform$

$\left(P\left(\triangle D^c_O\right)\right)_O : uniform$

I.

Q. : $P\left(I_{p1} \mid M_{p1} M_{p2} R^t_{(p1-o1)} R^t_{(p1-p2)} R^c_{(p1-o1)} R^c_{(p1-p2)}\right)$

Fig. 9. Bayesian program for the human activity model.

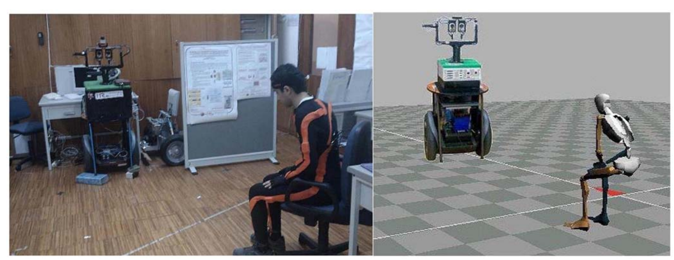


Fig. 10. Scene showing a person performing an activity (sitting) in the real world (left), and its 3-D representation provided by the motion tracker suit (right).

The classification results obtained can be seen in Table III. The overall performance proves that the context-based knowledge improves the accuracy of the model (from 92.22% to 96.80%) by reducing the false detections which were obtained in [43]. As it can be seen, between human-chair and H-HI there is no false detection. However, there are still some false detections between those similar context-based activities and especially between most of the classes with the *Other* class, which represents activities other than the ones listed in the

TABLE III
CLASSIFICATION RESULTS IN PERCENTAGE [46] (RCH:REACH, SPD:SPREAD, STD:STAND, HSK:HAND SHAKING, PSH:PUSH, AND OTH:OTHER)

	Rch	Spd	Sit	Std	Hsk	Psh	Oth
Rch	**97,78**	1,22	0,00	0,00	0,00	0,00	2,22
Spd	0,00	**95,74**	0,00	0,00	0,00	0,00	4,26
Sit	0,00	0,00	**96,67**	0,00	0,00	0,00	3,33
Std	0,00	0,00	2,70	**97,30**	0,00	0,00	0,00
Hsk	0,00	0,00	0,00	0,00	**97,83**	2,17	0,00
Psh	0,00	0,00	0,00	0,00	2,27	**97,73**	0,00
Oth	1,09	2,17	1,09	1,09	0,00	0,00	**94,57**

table. Most of those false detections happen in the boundary between two classes, mostly because of the use of sliding window-based segmentation approach. When a sliding window occurs in a boundary (transition of one activity to another one), the new class of activity is considered in the ground truth, however, the sliding window may contain more signal belonging to the previous activity, than new one. However, Santos *et al.* [51] proposed an adaptive sliding window approach to solve this problem.

C. Human Activity Analysis—Social Context-Based

Human behavior analysis in social context is a recent relevant field of study, which explores complex dynamic processes (SRs, relationships, etc.) in people's interaction and communication. There is a meaningful relation between nonverbal features (body motions, facial, and vocal expression) and social behaviors, which has been studied by psychologists for decades [57].

For the purpose of automatic social behavior analysis, studies have been conducted in both social and computer sciences, giving birth to a new interdisciplinary field social signal processing [56]. Several studies attempt to estimate social behaviors based on nonverbal features, such as: electroencephalogram (EEG) responses and eye gaze [52], prosodic features [35], prosodic and facial expressions [18], facial and verbal expressions [29], and visual motion and acoustic features [48].

Pentland [38] proposed a set of definitions for interpreting a number of IBs with respect to both psychology and artificial intelligence sciences. Some of the recent relevant works concerning social behavior estimation include: empathy behavior [29], frustration and delight smiles [18], valence and arousal emotions [52], approach-and-avoidance behavior [48], dominance patterns [22], natural, supporter, protagonist and attacker SRs [11], [62], mimicry expression [54], activation, valence and dominance emotional trends [32], and four IBs mentioned in [38] evaluated in [44] and extended to estimate an SR in [47].

Our approach to interpret complex human movement patterns in interpersonal activities reflects the multilevel framework we assumed, so two steps of analysis are proposed. The first one explores the richness of the LMA components in finding the existent dependencies between human body motions for each one of the defined IBs (*Interest*, *Indicator*, *Empathy*, and *Emphasis*). In the second one, we estimate "*Leading*" SRs by fusing the previously obtained four IBs.

TABLE IV
BRIEF DESCRIPTION OF IBS WITH THEIR STATES, AND THEIR RELEVANT LMA COMPONENTS, AS SHOWN IN [44]

IBs	IB def. [38]	States [44]	LMA [21]
Indicator	Most body movements	influenced influent	*Effort*
Empathy	Imitating and nodding	uncoordinated mimicry	*Space*
Interest	Energetic body motion	passive active	*Effort*
Emphasis	Jerky movements	consistent variable	*Effort, Space*

TABLE V
RELEVANT IBS STATES FOR THE LEADING ROLE, AS INTRODUCED IN [38]

IBs	State
Indicator	high level of influence
Empathy	no relevant information
Interest	high activity level
Emphasis	have consistence motion

Next, we present our approach for social context-dependent human movement modeling, and our experimental results when applying it to classify different IBs.

Step 1 (Modeling): IBs are analyzed by defining the dependencies between body-motion-based features of people during a conversation scenario [44]. However, SRs are a higher level of information than IBs [38]. To fill the gap between the nonverbal-based features (voice and body motion) and SRs, a set of IBs were proposed by Alex Pentland's group

$$\text{IB} \in \{\text{Indicator, Interest, Empathy, Emphasis}\}. \quad (6)$$

By combining these IBs, a number of SRs, like *Searching*, *Teaming*, *Listening*, and *Leading*, were defined [38]. In this paper, we estimate the *Leading* SR by analyzing the IBs obtained using the LMA components.

The proposed framework with three levels is presented in Fig. 11. LMA information for each person, in a face-to-face interaction context, is obtained in the first layer. Afterwards, IBs are estimated given the LMA features. Finally in the last layer, the SR is analyzed given the IBs. Table IV presents a brief description of the considered IBs, showing their respective states and related relevant LMA components, obtained in our previous work [44].

Based on Pentland's [38] definition about the SRs, the *Leading* role is described as a combination of attention, interest and great focus in thought and purpose. Thus, the IBs states, related with the *Leading* role, are described in Table V. As it can be seen, *Empathy* was not considered as an effective IB for the role.

By analyzing Tables IV and V, we conclude that the *Effort* component is the most important feature for the analysis of

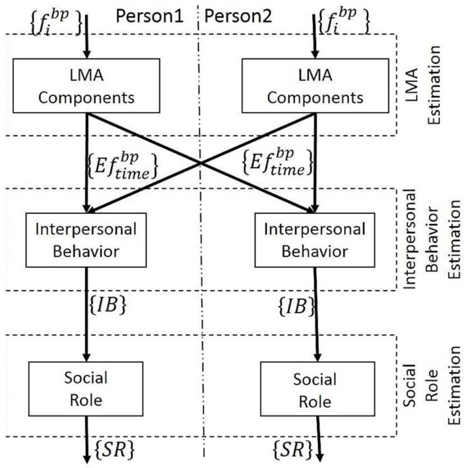


Fig. 11. Multilevel model used to estimate SRs from IBs.

the *Leading* role. This is the reason why we only considered the *Effort* component as an input feature to analyze the IBs.

In [44], the dependencies between LMA and Pentland's definitions were explored, and a Bayesian model for each one of the IBs was proposed.

SRs present different types of relationships between people in any community. In [38], several examples are proposed to illustrate how the SRs can be estimated by combining IBs. Fig. 11 presents a diagram of the proposed system. The estimation of the SR is performed by finding the Bayesian rule that maximizes (7) as a function of the previously estimated IBs

$$P\left(\mathrm{SR} \mid \prod_{l=1:k} \mathrm{IB}_l\right) \tag{7}$$

where l, k, IB_l, and SR denote the IB index (for the four IBs), the number of IBs, the lth IB, and the performed SR (*Leading*) variable. Fig. 12 presents the corresponding BP for the proposed *Leading* SR model.

We tested the model using H-HI scenarios. For the case of having more than a couple of persons, LMA components for each person will be estimated separately, and in parallel. *Interest* and *Emphasis* variables are analyzed individually and independently, but *Empathy* and *Indicator* variables need an interaction between two subjects to be estimated, which increases the complexity of the modeling process. To simplify the modeling process the "*influence model*" technique was used, as proposed in [11].

Step 2 (Experiments): In this step, the motion tracker suit is used to collect the 3-D positions of several body parts. Ten actors wore the suit and performed arbitrary conversations [1]. Table VI shows the list of the participants with respect to their nationality, level of education, and age range. All the participants are male. The output of the suit sensors was stored in an XML file in each trial. Each one of the ten actors performed three trials in different arbitrary scenarios, either in the role of leading or not-leading the conversation.

The correct classification rate for *Indicator*, *Interest*, *Empathy*, *Emphasis*, and *Leading* role were 86.5%, 96.5%, 72.5%, 97.5%, and 90.5%, respectively. The results proved

$$
\text{Program}\begin{cases}
\text{Description}\begin{cases}
\text{Specification}\begin{cases}
\textbf{Pertinent variables}: \\
bp \in \{Head, LeftHand, RightHand\} \\
i \in \{first\,person, second\,person\} \\
S_i^{T\in\{t,t-1\}} \in \{Sinking, Still, Rising\} \\
Ef_i^{p\in\{bp\}} \in \{Sustained, Sudden\} \\
Ind \in \{Influenced, Influent\} \\
Int \in \{Passive, Active\} \\
Emp \in \{Mimicry, Uncoordinated\} \\
Emf \in \{Consistent, Variable\} \\
IB \in \{Indicator, Interest, Empathy, Emphasis\} \\
SR \in \{Leading, No-Leading\} \\
\textbf{Decomposition}: \\
\left(P\left(SR_i \prod_{B\in\{IB\}} (B_i)\ Ef_i^{p\in\{bp\}}\ S_i^{T\in\{t,t-1\}}\right)\right) = \\
P(SR_i) \prod_{B\in\{IB\}} \\
\left(P(B_i \mid SR_i)\, P\left(Ef_i^{p\in\{bp\}} \mid B_i\right) P\left(S_i^{T\in\{t,t-1\}} \mid B_i\right)\right)^{i=1:2} \\
\textbf{Parametric forms}: \\
P(SR_i) : uniform \\
P\left(Ef_i^{p\in\{bp\}} \mid B_i\right)_{B\in\{IB\}}^{i=1:2} : histogram(Matrix) \\
P\left(S_i^{T\in\{t,t-1\}} \mid B_i\right)_{B\in\{IB\}}^{i=1:2} : histogram(Matrix) \\
\left(B_{i=1,2}^{B\in\{IB\}}\right) : BP \\
P(B_i \mid SR_i)_{B\in\{IB\}}^{i=1:2} : histogram(Matrix)
\end{cases} \\
\mathbf{I}:
\end{cases} \\
\mathbf{Q}: \ P\left(SR_i \,\middle|\, \prod_{B\in\{IB\}} (B_i)\right)^{i=1:2}
\end{cases}
$$

Fig. 12. BP for the *Leading* role model.

TABLE VI
LIST OF PARTICIPANTS IN OUR DATASET (NP: NUMBER OF PEOPLE)

NP	Nationality	Level of education	Age range
1	Spain	Post-doc	30-31
1	Brazil	PhD	35-36
1	Iran	PhD	37-38
2	Portugal	PhD	30-37
3	Portugal	Master	22-28
2	Portugal	Bachelor	19-22

that by using frequency-based features in LMA space, the accuracy of IBs recognition module improves, making the system more precise than the previous approach [44] (from about 77% to 88%). More details about the SR analysis can be found in [47].

VI. MULTILEVEL FRAMEWORK

In this section, we present how the different levels of analysis, described in Section V, can be interconnected. This process establishes a multilevel framework to deal with these levels. In previous sections, the modeling and experiment steps for each level, were described. Since, each level is built on top of the previous ones, we propose a hierarchical framework that consists of all of the previously mentioned levels, as can be seen in Fig. 13. It should be noted that, when a new level of analysis is added on top of a previous one, a learning process is needed just for this new level. The proposed multilevel

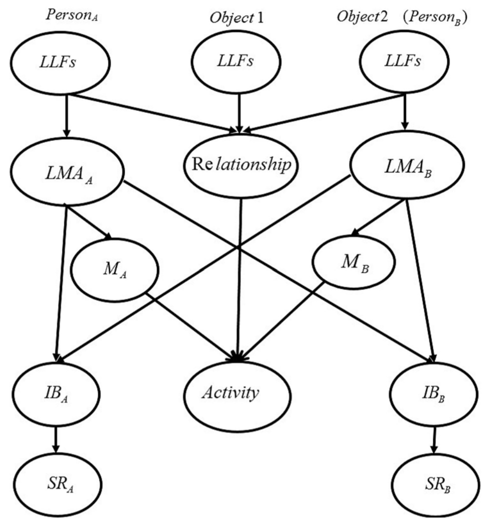


Fig. 13. Proposed multilevel framework for analyzing human movement, action and behavior in different levels, for human–human and human–object interaction-based scenarios.

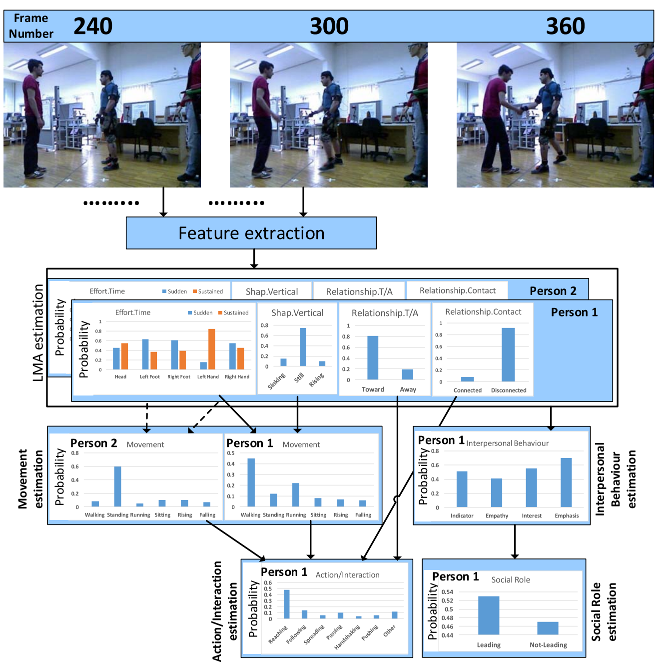


Fig. 14. Diagram showing the different levels of the estimation process for an H-HI scenario, where a user is walking to reach another one to perform a handshake [1].

framework analyzes all those different levels of human activities, with each level of analysis being represented by a set of semantic descriptions which can become useful for a variety applications.

A. Sample Scenario

To showcase a sample application of the proposed framework, we select an example of an activity, where a person walks toward another one, who is in front of him, to perform a handshake. Fig. 14 illustrates this scenario, representing all considered levels of the estimation process. After extracting

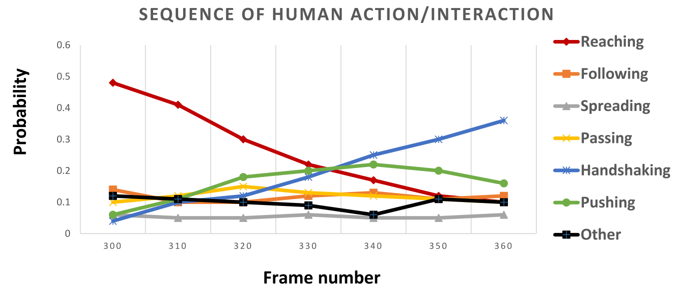


Fig. 15. Sequence of probabilistic results for a human action/interaction analysis.

a set of features, the first level of analysis is the estimation of the LMA parameters probabilities for both persons [*Effort.time* of body parts, *Shape.frontal*, *Relationship* parameters (*Away/Toward* and *Connected/Disconnected*)]. Each person's LMA parameters are used to estimate individual movement probabilities. However, to recognize the IB states for each person, the LMA parameters probabilities of both persons are considered. Then, we use the movement probabilities of both persons together with the LMA.*Relationship* parameters to estimate the action/interaction states probabilities. Finally, the SR states are estimated by using the IB states probabilities. Fig. 15 shows a sequence of estimations for the human action/interaction level of analysis for the mentioned scenario.

Table VII presents the information obtained in several consecutive frames, for all levels of the estimation steps represented in Fig. 14. Looking into the results in Table VII, at frame number 300, it can be seen that the *Effort.time* of the feet and the right hand are in the *sudden*'s state, however, the left hand and the head are in the *sustained*'s state, while *Shape.frontal* of the person is in *still*'s state. These states of the LMA parameters mean that *walking*'s class has the highest probability in the movement level. Since, the *Relationship.T/A* is in *Toward*'s state, and the *Relationship.contact* is in *disconnected*'s state, it can be inferred that the person is walking to reach the other person, thus, the *Reaching* class has the highest probability in the action/interaction level. The obtained IB states cause the probability of the *Leading* state to be slightly higher than the opposite one, probably because of having high probability in *consistent*'s state from IB.*Emphasis*.

At frame number 360, the *Effort.time* of the feet and left hand are in *sustained* state, but the right hand in *sudden*. The *Shape.frontal* of the person has remained in the *Still*'s, thus, the *Standing* is estimated as the one with the highest probability in the movement level. The *Relationship.T/A* of the person has remained slightly in *Toward*'s state, however, the *Relationship.contact* is in *connected* state, thus, the system detected the *Hand Shaking* with the highest probability in the action/interaction level. In the IB level, there is high probability in *Mimicry*'s state in the IB.*Empathy*, which is understandable because of the hand shaking activity. The IB.*Indicator* and the IB.*Interest* are estimated in the same state as the previous time, but with higher probability, however, the IB.*Emphasis* remains in the same condition. These changes of the IB states result in estimating the *Leading* of the SR as having higher probability than the state in the previous time.

TABLE VII
DIFFERENT LEVELS OF BODY-MOTION-BASED INFORMATION FOR A PERSON WALKING TO REACH ANOTHER PERSON TO PERFORM A HANDSHAKE. THE PARAMETERS VALUES OF TWO CONSECUTIVE POSITIONS OF A SLIDING WINDOW ARE SHOWN

Frame	Level	States and their probability
300		
	LMA.Effort.Time-Head	Sudden:45%, Sustained:**55%**
	LMA.Effort.Time-LFoot	Sudden:**63%**, Sustained:37%
	LMA.Effort.Time-RFoot	Sudden:**61%**, Sustained:39%
	LMA.Effort.Time-LHand	Sudden:15%, Sustained:**85%**
	LMA.Effort.Time-RHand	Sudden:**55%**, Sustained:45%
	LMA.Shape.Frontal	Sinking:15%,Still:**75%**,Rising:10%
	LMA.Relationship.T/A	Toward:**81%**, Away:19%
	LMA.Relationship.Contact	Connected:8%, Disconnected:**92%**
	Movement	Walking:**45%**, Standing:12%, Running:22%,Sitting:8%, Rising:7%,Falling:6%
	Action and Interaction	Reaching:**48%**, Following:14%, Spreading:6%,Passing:10%, Handshaking:4%,Pushing:6%,Other:12%
	IB.Indicator	Influnent:**51%**, Influenced:49%
	IB.Empathy	Mimicry: 41%, Uncoordinated:**59%**
	IB.Interest	Active:**55%**, Passive:45%
	IB.Emphasis	Consistent:**70%**, Variable:30%
	SR.Leading	Leader:**53%**, Non-Leader:47%
360		
	LMA.Effort.Time-Head	Sudden:31%, Sustained:**69%**
	LMA.Effort.Time-LFoot	Sudden:29%, Sustained:**71%**
	LMA.Effort.Time-RFoot	Sudden:23%, Sustained:**77%**
	LMA.Effort.Time-LHand	Sudden:**25%**, Sustained:**75%**
	LMA.Effort.Time-RHand	Sudden:**55%**, Sustained:45%
	LMA.Shape.Frontal	Sinking:25%,Still:**67%**,Rising:18%
	LMA.Relationship.T/A	Toward:**51%**, Away:49%
	LMA.Relationship.Contact	Connected:**67%**, Disconnected:33%
	Movement	Walking:18%, Standing:**52%**, Running:9%,Sitting:7%, Rising:7%,Falling:5%
	Action and Interaction	Reaching:10%, Following:12%, Spreading:6%,Passing:10%, Handshaking:**36%**,Pushing:16%,other:10%
	IB.Indicator	Influnent:**60%**, Influenced:40%
	IB.Empathy	Mimicry: **72%**, Uncoordinated:28%
	IB.Interest	Active:**74%**, Passive:26%
	IB.Emphasis	Consistent:**66%**, Variable:44%
	SR.Leading	Leader:**68%**, Non-Leader:32%

Normally, during walking, all limbs are in *sudden* state for their *Effort.time* parameter. In the first step of the previous example (frame number 300) the left hand was in *sustained* state while in walking movement, however, the system could still estimate the movement correctly. This is one of the advantages of the BN-based multilevel framework which can deal with uncertainty, and imperfect data from lower levels.

VII. CONCLUSION

This paper presents a general framework for the analysis of human activities, taking into account possible interactions with the surrounding environment. Human movement dynamics and their relations with other objects in a scene, need to be explored by intelligent systems like robots, in real world application scenarios. For understanding a human behavior during their interaction with another person, object or scene, several levels of analysis are involved. Thus, a multilevel framework is proposed to provide an understanding of the probabilistic causes that lead to an effect. This understanding could also be beneficial for the maintenance and extension of the probabilistic models. An intelligent system, like a robot, can link to each of those levels of information for performing an appropriate interaction with its environment.

The contribution of this paper is that it combines different levels of human activity into a multilevel framework, where the interactions between these levels are modeled. For this purpose, a mid-level human motion descriptor, the LMA, is used and formulated to fill the gap between LLFs and higher level human activities. Since, the level of the LMA parameters is close to the LLFs, thus, an accurate estimation of the LMA parameters can be obtained. Depending on these parameters, the relevant LLFs of different human motions can be defined, which reduces data redundancy. A BP approach is used to model the dependencies in different level of analysis.

A short review about the relevant issues involved in human activity analysis was conducted to explore how the body motion-based human activities, in different levels, can be modeled using a multilevel framework. The proposed framework takes advantage of a hierarchical approach, dividing a complex problem into simpler ones, making the system more feasible. We believe that semantically defined multilevel analyses, like the one we propose, will be suitable for helping intelligent systems in the decision making process of future social/assistive robots, in several applications.

ACKNOWLEDGMENT

The authors would like to thank their laboratory colleagues for their support.

REFERENCES

[1] (Jan. 03, 2014). *University of Coimbra 3-D Motion Database*. [Online]. Available: http://mrl.isr.uc.pt/experimentaldata/public/uc-3d/

[2] J. K. Aggarwal, Q. Cai, W. Liao, and B. Sabata, "Articulated and elastic non-rigid motion: A review," in *Proc. IEEE Workshop Motion Non-Rigid Articulated Objects*, Austin, TX, USA, 1994, pp. 2–14.

[3] J. K. Aggarwal and M. S. Ryoo, "Human activity analysis: A review," *ACM Comput. Surveys*, vol. 43, no. 3, Apr. 2011, Art. no. 16.

[4] D. Arsic, W. Bjorn, B. Schuller, and G. Rigoll, "Video based online behavior detection using probabilistic multi stream fusion," in *Proc. IEEE Int. Conf. Image Process.*, vol. 2. Genoa, Italy, 2005, pp. 606–609.

[5] P. Bessière, E. Mazer, J.-M. Ahuactzin, and K. Mekhnacha, *Bayesian Programming*. London, U.K.: Chapman & Hall, 2013.

[6] A. F. Bobick, "Movement, activity and action: The role of knowledge in the perception of motion," *Philosoph. Trans. Roy. Soc. B Biol. Sci.*, vol. 352, no. 1358, pp. 1257–1265, 1997.

[7] C. Cédras and M. Shah, "Motion-based recognition a survey," *Image Vis. Comput.*, vol. 13, no. 2, pp. 129–155, 1995.

[8] M. Chen, S. Gonzalez, A. Vasilakos, H. Cao, and V. C. M. Leung, "Body area networks: A survey," *Mobile Netw. Appl.*, vol. 16, no. 2, pp. 171–193, 2011.

[9] D. Chi, M. Costa, L. Zhao, and N. Badler, "The emote model for effort and shape," in *Proc. 27th Annu. Conf. Comp. Graph. Interact. Tech.*, New Orleans, LA, USA, 2000, pp. 173–182.

[10] V. Delaitre *et al.*, "Scene semantics from long-term observation of people," in *Proc. 12th Eur. Conf. Comput. Vis. (ECCV)*, Florence, Italy, 2012, pp. 284–298.

[11] W. Dong *et al.*, "Using the influence model to recognize functional roles in meetings," in *Proc. 9th Int. Conf. Multimodal Interfaces*, Nagoya, Japan, 2007, pp. 271–278.

[12] M. S. El-Nasr, A. Vasilakos, C. Rao, and J. A. Zupko, "Dynamic intelligent lighting for directing visual attention in interactive 3-D scenes," *IEEE Trans. Comput. Intell. AI in Games*, vol. 1, no. 2, pp. 145–153, Jun. 2009.

[13] D. R. Faria, C. Premebida, and U. Nunes, "A probabilistic approach for human everyday activities recognition using body motion from RGB-D images," in *Proc. IEEE RO-MAN*, Edinburgh, U.K., 2014, pp. 732–737.

[14] A. Foroud and I. Q. Whishaw, "Changes in the kinematic structure and non-kinematic features of movements during skilled reaching after stroke: A Laban movement analysis in two case studies," *J. Neurosci. Methods*, vol. 158, no. 1, pp. 137–149, 2006.

[15] D. M. Gavrila, "The visual analysis of human movement: A survey," *Comput. Vis. Image Understand.*, vol. 73, no. 1, pp. 82–98, 1999.

[16] A. Gupta, A. Kembhavi, and L. S. Davis, "Observing human-object interactions: Using spatial and functional compatibility for recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 10, pp. 1775–1789, Oct. 2009.

[17] Y.-J. Hong, I.-J. Kim, S. C. Ahn, and H.-G. Kim, "Activity recognition using wearable sensors for elder care," in *Proc. 2nd Int. Conf. Future Gener. Commun. Netw.*, vol. 2. Hainan, China, Dec. 2008, pp. 302–305.

[18] M. E. Hoque, D. McDuff, and R. W. Picard, "Exploring temporal patterns in classifying frustrated and delighted smiles," *IEEE Trans. Affect. Comput.*, vol. 3, no. 3, pp. 323–334, Jul. 2012.

[19] F. Hülsmann, C. Frank, T. Schack, S. Kopp, and M. Botsch, "Multi-level analysis of motor actions as a basis for effective coaching in virtual reality," in *Proc. 10th Int. Symp. Comput. Sci. Sports (ISCSS)*, 2016, pp. 211–214.

[20] Y.-X. Hung, C.-Y. Chiang, S. J. Hsu, and C.-T. Chan, "Abnormality detection for improving elder's daily life independent," in *Proc. Aging Friendly Technol. Health Independence*, Seoul, South Korea, 2010, pp. 186–194.

[21] A. Hutchinson, *Labanotation*. London, U.K.: Oxford Univ. Press, 1974.

[22] D. B. Jayagopi, H. Hung, Y. Chuohao, and G.-P. Daniel, "Modeling dominance in group conversations using nonverbal activity cues," *IEEE Audio, Speech, Language Process.*, vol. 17, no. 3, pp. 501–513, Mar. 2009.

[23] T. Kadir, R. Bowden, E.-J. Ong, and A. Zisserman, "Minimal training, large lexicon, unconstrained sign language recognition," in *Proc. Brit. Mach. Vis. Conf.*, London, U.K., 2004, pp. 939–948.

[24] R. Kelley *et al.*, "Understanding human intentions via hidden Markov models in autonomous mobile robots," in *Proc. 3rd ACM/IEEE Int. Conf. Human Robot Interact. (HRI)*, Amsterdam, The Netherlands, 2008, pp. 367–374.

[25] K. Khoshhal Roudposhti, H. Aliakbarpour, J. Quintas, P. Drews, and J. Dias, "Probabilistic LMA-based classification of human behaviour understanding using power spectrum technique," in *Proc. 13th Int. Conf. Inf. Fusion*, Jul. 2010, pp. 1–7.

[26] K. Khoshhal Roudposhti, H. Aliakbarpour, J. Quintas, M. Hofmann, and J. Dias, "Probabilistic LMA-based human motion analysis by conjugating frequency and spatial based features," in *Proc. 12th Int. Workshop Image Anal. Multimedia Interact. Services*, Delft, The Netherlands, Apr. 2011, pp. 1–4.

[27] H. S. Koppula, R. Gupta, and A. Saxena, "Learning human activities and object affordances from RGB-D videos," *Int. J. Robot. Res.*, vol. 32, no. 8, pp. 951–970, Jul. 2013.

[28] O. Lebeltel, P. Bessière, J. Diard, and E. Mazer, "Bayesian robot programming," *Auton. Robots*, vol. 16, no. 1, pp. 49–79, Jan. 2004.

[29] I. Leite *et al.*, "The influence of empathy in human-robot relations," *Int. J. Human Comput. Stud.*, vol. 71, no. 3, pp. 250–260, 2013.

[30] M. Leo, T. Dorazio, and P. Spagnolo, "Human activity recognition for automatic visual surveillance of wide areas," in *Proc. 2nd Int. Workshop Video Surveillance Sensor Netw.*, New York, NY, USA, 2004, pp. 124–130.

[31] A. Mannini and A. M. Sabatini, "Machine learning methods for classifying human physical activity from on-body accelerometers," *Sensors*, vol. 10, no. 2, pp. 1154–1175, 2010.

[32] A. Metallinou, A. Katsamanis, and S. Narayanan, "Tracking continuous emotional trends of participants during affective dyadic interactions using body language and speech information," *Image Vis. Comput.*, vol. 31, no. 2, pp. 137–152, Feb. 2013.

[33] T. B. Moeslund, A. Hilton, and V. Krüger, "A survey of advances in vision-based human motion capture and analysis," *Comput. Vis. Image Understand.*, vol. 104, nos. 2–3, pp. 90–126, 2006.

[34] T. B. Moeslund and E. Granum, "A survey of computer vision-based human motion capture," *Comput. Vis. Image Understand.*, vol. 81, no. 3, pp. 231–268, 2001.

[35] G. Mohammadi and A. Vinciarelli, "Automatic personality perception: Prediction of trait attribution based on prosodic features," *IEEE Trans. Affect. Comput.*, vol. 3, no. 3, pp. 273–284, Jul./Sep. 2012.

[36] T. Nakata, T. Mori, and T. Sato, "Analysis of impression of robot bodily expression," *J. Robot. Mechatronics*, vol. 14, no. 1, pp. 27–36, 2002.

[37] J. C. Nascimento, M. A. T. Figueiredo, and J. S. Marques "Segmentation and classification of human activities," in *Proc. Int. Workshop Human Act. Recognit. Model.*, Oxford, U.K., 2005, pp. 79–86.

[38] A. Pentland, *Honest Signals: How They Shape Our World*. Cambridge, MA, USA: MIT Press, 2008.

[39] R. Poppe, "Vision-based human motion analysis: An overview," *Comput. Vis. Image Understand.*, vol. 108, nos. 1–2, pp. 4–18, 2007.

[40] J. Rett and J. Dias, "Human robot interaction based on Bayesian analysis of human movements," in *Proc. EPIA*, Guimarães, Portugal, 2007, pp. 530–541.

[41] J. Rett, "Robot-human interface using Laban Movement analysis inside a Bayesian framework," Ph.D. dissertation, Dept. Elect. Comput. Eng., Univ. Coimbra, Coimbra, Portugal, 2008.

[42] J. Rett, J. Dias, and J.-M. Ahuactzin, "Laban movement analysis using a Bayesian model and perspective projections," in *Brain, Vision and AI*, Rijeka, Croatia: InTech, 2008, pp. 183–192.

[43] K. Khoshhal Roudposhti and J. Dias, "Probabilistic human interaction understanding: Exploring relationship between human body motion and the environmental context," *Pattern Recognit. Lett.*, vol. 34, no. 7, pp. 820–830, 2013.

[44] K. Khoshhal Roudposhti, L. Santos, H. Aliakbarpour, and J. Dias, "Parameterizing interpersonal behaviour with Laban movement analysis—A Bayesian approach," in *Proc. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Providence, RI, USA, Jun. 2012, pp. 7–13.

[45] K. Khoshhal Roudposhti, "Probabilistic-based human behaviour analysis using hierarchical framework," Ph.D. dissertation, Dept. Elect. Comput. Eng., Univ. Coimbra, Coimbra, Portugal, 2014.

[46] K. Khoshhal Roudposhti and J. Dias, "Improved semantic-based human interaction understanding using context-based knowledge," in *Proc. IEEE Int. Conf. Syst. Man Cybern (SMC)*, Santa Monica, CA, USA, Oct. 2013, pp. 2899–2904.

[47] K. K. Rousposhti, U. Nunes, and J. Dias, "Probabilistic social behavior analysis by exploring body motion-based patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 8, pp. 1679–1691, Aug. 2016.

[48] V. Rozgic *et al.*, "Estimation of ordinal approach-avoidance labels in dyadic interactions: Ordinal logistic regression approach," in *Proc. ICASSP*, Prague, Czech Republic, 2011, pp. 2368–2371.

[49] M. S. Ryoo and J. K. Aggarwal, "Semantic representation and recognition of continued and recursive human activities," *Int. J. Comput. Vis.*, vol. 82, no. 1, pp. 1–24, Apr. 2009.

[50] M. S. Ryoo and J. K. Aggarwal, "Hierarchical recognition of human activities interacting with objects," in *Proc. Comput. Vis. Pattern Recognit. Conf.*, Minneapolis, MN, USA, Jun. 2007, pp. 1–8.

[51] L. Santos, K. Khoshhal Roudposhti, and J. Dias, "Trajectory-based human action segmentation," *Pattern Recognit.*, vol. 48, no. 2, pp. 568–579, 2015.

[52] M. Soleymani, M. Pantic, and T. Pun, "Multimodal emotion recognition in response to videos," *IEEE Trans. Affect. Comput.*, vol. 3, no. 2, pp. 211–223, Apr./Jun. 2012.

[53] T. Starner and A. Pentland, "Visual recognition of American sign language using hidden Markov models," in *Proc. Int. Workshop Autom. Face Gesture Recognit.*, Zürich, Switzerland, 1995, pp. 189–194.

[54] X. Sun, A. Nijholt, K. P. Truong, and M. Pantic, "Automatic visual mimicry expression analysis in interpersonal interaction," in *Proc. IEEE Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Aug. 2011, pp. 40–46.

[55] D. Swaminathan *et al.*, "A dynamic Bayesian approach to computational Laban shape quality analysis," *Adv. Human Comput. Interact.*, vol. 2009, Jan. 2009, Art. no 2.

[56] A. Vinciarelli *et al.*, "Bridging the gap between social animal and unsocial machine: A survey of social signal processing," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 69–87, Jan. 2012.

[57] A. Vinciarelli, H. Salamin, and M. Pantic, "Social signal processing: Understanding social interactions through nonverbal behavior analysis," in *Proc. CVPR Workshops*, Miami, FL, USA, Jun. 2009, pp. 42–49.

[58] J. Wang, Z. Liu, Y. Wu, and J. Yuan, "Learning actionlet ensemble for 3D human action recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 5, pp. 914–927, May 2014.

[59] L. Wang, W. Hu, and T. Tan, "Recent developments in human motion analysis," *Pattern Recognit.*, vol. 36, no. 3, pp. 585–601, 2003.

[60] Z. Wang *et al.*, "Probabilistic movement modeling for intention inference in human-robot interaction," *Int. J. Robot. Res.*, vol. 32, no. 7, pp. 841–858, Jun. 2013.

[61] A. D. Wilson and A. F. Bobick, "Parametric hidden Markov models for gesture recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 9, pp. 884–900, Sep. 1999.

[62] M. Zancanaro, B. Lepri, and F. Pianesi, "Automatic detection of group functional roles in face to face interactions," in *Proc. 8th ICMI*, Banff, AB, Canada, 2006, pp. 28–34.

[63] L. Zhao, "Synthesis and acquisition of Laban movement analysis qualitative parameters for communicative gestures," Ph.D. dissertation, Dept. Comput. Inf. Sci. Tech., Univ. Pennsylvania, Philadelphia, PA, USA, 2001.

Kamrad Khoshhal Roudposhti received the M.Sc. degree in artificial intelligence and robotic from the University of Azad (Science and Research Branch), Tehran, Iran, and the Ph.D. degree in automation and robotics from the Electrical and Computer Engineering, University of Coimbra, Coimbra, Portugal, in 2014.

He is currently a Post-Doctoral Research Associate with the Computer Science, Texas State University, San Marcos, TX, USA. His current research interests include body motion analysis for human activity/behavior understanding, social signal processing, biosignal processing, and HRI.

Jorge Dias received the Ph.D. degree in electrical engineering and the Habilitation degree from the University of Coimbra (UC), Coimbra, Portugal, specialization in control and instrumentation.

He has been an Associate Professor with the UC and the Institute of Systems and Robotics, UC. He does research in the area of Computer Vision and Robotics and has contributions on the field since 1984. He has coordinated the Mobile Robotics Laboratory with the Instituto of Systems and Robotics and the Laboratory of Systems and Automation with the Instituto Pedro Nunes (IPN), Coimbra. IPN is a technology transfer institute from the University of Coimbra. He was the Vice-President with IPN from 2008 to 2011. Since 2011, he has been supporting the set-up of the Robotics Institute with Khalifa University, Abu Dhabi, UAE.

Paulo Peixoto received the Ph.D. degree in electrical engineering from the University of Coimbra, Coimbra, Portugal, in 2003.

He is currently an Assistant Professor with the Department of Electrical and Computer Engineering, University of Coimbra and a Researcher with the Institute of Systems and Robotics, University of Coimbra. He has been involved/responsible for several funded projects, at both National and International levels, in the areas of computer vision, visual surveillance, and intelligent vehicles. His current research interests include computer vision applied to intelligent transportation systems, pattern recognition, and human–computer interaction.

Vangelis Metsis received the Bachelor of Science degree in computer science from the Department of Informatics, Athens University of Economics and Business, Athens, Greece, in 2005, and the Doctoral degree from the Department of Computer Science and Engineering, University of Texas at Arlington, Arlington, TX, USA, in 2011.

He is an Assistant Professor with the Department of Computer Science, Texas State University, San Marcos, TX, USA. His current research interests include machine learning, data mining, and computer vision with special focus on applications of smart health & wellbeing, and pervasive computing.

Dr. Metsis was a recipient of the Federal Support. He has a strong publication record in related research areas, such as human activity and behavior analysis, human–computer interaction and motion tracking using sensors, sleep monitoring, and human biosignal analysis.

Urbano Nunes received the Ph.D. degree in electrical engineering from the University of Coimbra, Coimbra, Portugal, in 1995.

He is a Full Professor with the Electrical and Computer Engineering Department, Coimbra University. He is the Coordinator of the Automation and Robotics for Human Life Group, Institute for Systems and Robotics. He has been involved with/responsible for several funded projects at both national and international levels in the areas of human-centered mobile robotics and intelligent vehicles.

Prof. Nunes is an Associate Editor of the IEEE TRANSACTIONS ON INTELLIGENT VEHICLES. He was the General Chair of the 2012 IEEE International Conference on Intelligent Robots and Systems.